



UNIVERSIDAD DE VALLADOLID

FACULTAD DE MEDICINA

ESCUELA DE INGENIERÍAS INDUSTRIALES

TRABAJO DE FIN DE GRADO

GRADO EN INGENIERÍA BIOMÉDICA

**ESTUDIO DE LAS SEÑALES DE VARIABILIDAD DEL  
RITMO CARDIACO Y SATURACIÓN DE OXÍGENO EN  
SANGRE EN PACIENTES CON COVID-19  
PERSISTENTE**

Autora:

**D.ª María Fernández Vaquerizo**

Tutores:

**D. Daniel Álvarez González**

**D. Félix del Campo Matía**

Valladolid, 05 de julio de 2023



---

**TÍTULO:** **Estudio de las Señales de Variabilidad del Ritmo Cardíaco y Saturación de Oxígeno en Sangre en Pacientes con COVID-19 Persistente**

**AUTORA:** **D<sup>a</sup> María Fernández Vaquerizo**

**TUTORES:** **D. Daniel Álvarez González  
D. Félix del Campo Matía**

**DEPARTAMENTO:** **Departamento de Teoría de la Señal y Comunicaciones e Ingeniería Telemática  
Departamento de Medicina, Dermatología y Toxicología**

---

**TRIBUNAL**

---

**PRESIDENTE:** **D. Roberto Hornero Sánchez**

**SECRETARIO:** **D. Félix del Campo Matía**

**VOCAL:** **D. Daniel Álvarez González**

**SUPLENTE 1:** **D. Jesús Poza Crespo**

**SUPLENTE 2:** **D. Carlos Gómez Peña**

---

**FECHA:** **05 de julio de 2023**

**CALIFICACIÓN:**

---



*A Sergio y Jimena,  
mi mayor motivo de orgullo.*



# Agradecimientos

En primer lugar, me gustaría mostrar mi agradecimiento a mis tutores, el Dr. Daniel Álvarez González y el Dr. Félix del Campo Matía, por el apoyo y la confianza que me han brindado en todo momento. Ha sido una verdadera suerte poder aprender de ellos y gracias a esto, haber podido terminar el presente Trabajo de Fin de Grado de manera satisfactoria.

Deseo también expresar mi agradecimiento a Fernando Moreno, por su paciencia y disposición a ayudarme siempre que lo he necesitado. También me gustaría agradecer al personal sanitario del Servicio de Neumología del Hospital Universitario Río Hortega, en especial a las enfermeras y auxiliares de enfermería, por haber hecho mi estancia allí mucho más fácil.

Gracias también a los amigos que me ha regalado la carrera, estoy convencida de que sin ellos habría resultado mucho más difícil llegar hasta aquí. Me gustaría agradecer personalmente a Leyre, Ana, Álvaro y Marcos, por haber sido el mejor equipo con el que se podía contar.

Finalmente, deseo expresar mi más profundo agradecimiento a mi familia. A mi madre y a mi padre, por haber sido los dos pilares que me han impulsado siempre hacia delante. A Sergio, Alicia y Jimena, por el apoyo fundamental que me habéis aportado en cada momento de estos cuatro años. Y a Raúl, por la confianza que siempre ha puesto en mí.





# Resumen

**Antecedentes.** El síndrome post – COVID o COVID persistente (SPC-CP) hace referencia a los síntomas que continúan o se desarrollan después de una infección por coronavirus SARS-CoV-2, conocido por COVID-19, y que no se pueden explicar por un diagnóstico alternativo, incluyendo este término a los síntomas que continúan más allá de las 12 semanas posteriores a la infección, exhibiendo las personas que lo padecen un deterioro en la estructura y en la función de múltiples órganos. El SPC-CP es un gran desafío para la comunidad médica ya que su estudio y comprensión todavía es insuficiente. Entre el 10% y el 35% de los pacientes que han sido diagnosticados de COVID-19 refieren la persistencia de al menos un síntoma después de tres semanas de seguimiento a partir del diagnóstico, destacando como dos de los síntomas más frecuentes la fatiga y la disnea. Esta incidencia se incrementa hasta el 50-84% en caso de que los pacientes hayan requerido hospitalización durante el periodo de infección.

**Hipótesis y Objetivos.** El presente trabajo se ha desarrollado bajo la hipótesis principal de que el análisis de la actividad cardiaca, actividad respiratoria y acoplamiento cardiorrespiratorio podría aportar información relevante y complementaria útil en la gestión de pacientes con síndrome post-COVID. El objetivo general del estudio consistió en analizar y caracterizar los patrones cardiorrespiratorios de pacientes con y sin diagnóstico de SPC-CP para identificar los índices autonómicos, oximétricos y de acoplamiento con mayor nivel de asociación con la persistencia de los síntomas y poder diferenciar a estos pacientes de otros sujetos sin diagnóstico de SPC-CP.

**Material y Métodos.** Se reclutaron pacientes que acudían de forma consecutiva al Servicio de Neumología del Hospital Río Hortega de Valladolid, que fueron agrupados en las siguientes categorías: (i) sujetos de control sin diagnóstico previo de COVID-19; (ii) pacientes con diagnóstico confirmado de COVID-19; (iii) pacientes con diagnóstico de SPC-CP. A todos los participantes se les registró la señal de electrocardiograma y oximetría en reposo durante un mínimo de 7 minutos. La metodología llevada a cabo en este estudio se dividió en tres fases: extracción, selección y clasificación de características. En la primera fase se aplicaron métodos de análisis en el dominio temporal, espectral y no lineal para caracterizar las señales de HRV derivada del ECG, SpO<sub>2</sub> y el acoplamiento cardiorrespiratorio, incorporándose además variables de la historia clínica de cada paciente. En la segunda fase se seleccionó un subconjunto óptimo de características con la mayor cantidad de información complementaria para realizar el diagnóstico de SPC-CP, empleando para ello algoritmos genéticos. Por último, la clasificación se realizó mediante una red neuronal perceptrón multicapa (MLP), entrenada para identificar a los 3 grupos de pacientes bajo estudio (controles, COVID, SPC-CP).

**Resultados.** Un total de 105 pacientes cumplieron los criterios de inclusión y se incorporaron al estudio. De ellos, 83 sujetos pasaron finalmente a la etapa de modelado de datos: (i) 25 controles, (ii) 25 con diagnóstico previo de COVID sin persistencia de síntomas y (iii) 33 con diagnóstico confirmado de SPC-CP. Las siguientes características fueron seleccionadas automáticamente mediante el algoritmo genético: presencia de apnea obstructiva del sueño, potencia total en la banda de altas frecuencias de la señal de variabilidad del ritmo cardiaco, magnitud de la coherencia espectral (acoplamiento cardiorrespiratorio) y la puntuación del test CAT de calidad de vida. La red MLP alcanzó una precisión del 75%, un índice Kappa de 0.621 y un Macro F1-score del 76.52% en la identificación de las 3 clases de pacientes en una población independiente de test. Alternativamente, se alcanzó una precisión del 90% (92% Se y 88% Sp) en la diferenciación entre sujetos con y sin SPC-CP bajo un enfoque binario.

**Conclusiones.** El análisis automático de la actividad cardiaca, respiratoria y de la interacción entre ambas ha demostrado aportar información complementaria relevante en la caracterización del SPC-CP. Los resultados obtenidos sugieren que el empleo de esta información como entrada a una red neuronal podría resultar útil en la identificación y gestión de pacientes con SPC-CP.

## **Palabras clave**

COVID-19, síndrome post-COVID, COVID persistente, variabilidad del ritmo cardiaco, saturación de oxígeno en sangre, acoplamiento cardiorrespiratorio, procesamiento de señales biomédicas, aprendizaje computacional, algoritmos genéticos, redes neuronales.

# Abstract

**Background.** Post-COVID syndrome refers to symptoms that continue or develop after infection with SARS-CoV-2 coronavirus, known as COVID-19, and cannot be explained by any alternative diagnosis, including symptoms that continue beyond 12 weeks post-infection, with sufferers exhibiting impaired structure and function of multiple organs. SPC-CP is a major challenge for the medical community as it is still poorly studied or understood. Between 10% and 35% of patients diagnosed with COVID-19 report persistence of at least one symptom after three weeks of follow-up from diagnosis, with fatigue and dyspnoea being two of the most frequent symptoms. This incidence increases to 50-84% if patients have required hospitalisation during the infection period.

**Hypothesis and Objectives.** The present project has been developed under the main hypothesis that the analysis of cardiac activity, respiratory activity and cardiorespiratory coupling could provide relevant and complementary information useful in the management of patients with post-COVID syndrome. The overall aim of the study was to analyse and characterise the cardiorespiratory patterns of patients with and without a diagnosis of SPC-CP in order to identify the autonomic, oximetric and coupling indices with the highest level of association with symptom persistence and to be able to differentiate these patients from other subjects without a diagnosis of PC-PCS.

**Material and Methods.** Patients attending consecutively to the Pneumology Department of the Río Hortega Hospital in Valladolid were recruited and grouped into the following categories: (i) control subjects with no previous diagnosis of COVID-19; (ii) patients with a confirmed diagnosis of COVID-19; (iii) patients with a diagnosis of SPC-CP. All participants had their resting electrocardiogram and oximetry signal recorded for a minimum of 7 minutes. The methodology carried out in this study was divided into three phases: feature extraction, feature selection and feature classification. In the first phase, time-domain, spectral and non-linear analysis methods were applied to characterise the ECG-derived HRV, SpO<sub>2</sub> and cardiorespiratory coupling signals, also incorporating variables from each patient's clinical history. In the second phase, an optimal subset of features with the most complementary information was selected to make the diagnosis of SPC-CP using genetic algorithms. Finally, classification was performed using a multilayer perceptron neural network, trained to identify the 3 groups of patients under study (controls, COVID, SPC-CP).

**Results.** A total of 105 patients met the inclusion criteria and joined the study. Among these, 83 subjects finally passed to the data modelling stage: (i) 25 controls, (ii) 25 with a previous diagnosis of COVID without persistence of symptoms and (iii) 33 with a confirmed diagnosis of

SPC-CP. The following features were automatically selected using the genetic algorithm: presence of obstructive sleep apnoea, total power in the high frequency band of the heart rate variability signal, magnitude of spectral coherence (cardiorespiratory coupling) and CAT quality of life test score. The MLP network achieved an accuracy of 75%, a Kappa index of 0.621 and a Macro F1-score of 76.52% in identifying the 3 classes of patients in an independent test population. Alternatively, an accuracy of 90% (88% Se and 92% Sp) was achieved in differentiating between subjects with and without SPC-CP under a binary approach.

**Conclusions.** The automatic analysis of cardiac and respiratory activity and the interaction between them has been shown to provide relevant complementary information in the characterisation of SPC-CP. The results obtained suggest that the use of this information as input to a neural network could be useful in the identification and management of patients with SPC-CP.

## **Keywords**

COVID-19, post-COVID syndrome, long COVID, heart rate variability, blood oxygen saturation, cardiorespiratory coupling, biomedical signal processing, computational learning, genetic algorithms, neural networks.

# ÍNDICE GENERAL

<b>CAPÍTULO 1. INTRODUCCIÓN .....</b>	<b>1</b>
1.1. Infección por virus SARS-CoV-2 y COVID persistente .....	1
1.2. Análisis de datos clínicos y técnicas de inteligencia artificial.....	3
1.3. Estructura.....	3
<b>CAPÍTULO 2. SÍNDROME POST – COVID.....</b>	<b>5</b>
2.1. Síndrome post – COVID .....	5
2.1.1. Magnitud del problema .....	6
2.1.2. Mecanismos que contribuyen a la fisiopatología del síndrome post – COVID .....	7
2.1.3. Manifestaciones clínicas .....	7
2.1.4. Tratamiento y seguimiento de la enfermedad.....	10
2.2. Estado de la técnica en el procesado de señales biomédicas dentro del contexto del síndrome post-COVID.....	12
2.2.1. Señal de variabilidad de la frecuencia cardiaca en pacientes con síndrome post-COVID .....	13
2.2.2. Señal de saturación de oxígeno en sangre en pacientes con síndrome post-COVID 13	
2.2.3. Señal de respiración derivada del ECG en pacientes con síndrome post-COVID: cuantificación del acoplamiento cardiorrespiratorio.....	14
<b>CAPÍTULO 3. HIPÓTESIS Y OBJETIVOS.....</b>	<b>15</b>
3.1. Hipótesis .....	15
3.2. Objetivos.....	15
<b>CAPÍTULO 4. SUJETOS Y SEÑALES .....</b>	<b>17</b>
4.1. Diseño del estudio. Criterios de inclusión y exclusión.....	17
4.2. Protocolo de adquisición de datos .....	18
4.2.1. Variables clínicas .....	18
4.2.2. Señal de variabilidad de la frecuencia cardiaca.....	19
4.2.3. Señal de saturación de oxígeno en sangre.....	20

4.2.4. Señal de respiración derivada del ECG: cuantificación del acoplamiento cardiorrespiratorio .....	21
--	----

**CAPÍTULO 5. METODOLOGÍA.....23**

5.1. Limpieza de datos y preprocesado de señales biomédicas .....	23
5.1.1. Detección de outliers.....	23
5.1.2. Preprocesado de la señal de variabilidad de frecuencia cardíaca.....	23
5.1.3. Preprocesado de la señal de saturación de oxígeno en sangre .....	26
5.2. Estimación de la densidad espectral de potencial .....	26
5.3. Análisis no lineal.....	29
5.3.1. Entropía muestral .....	29
5.3.2. Medida de la tendencia central.....	30
5.3.3. Complejidad de Lempel-Ziv.....	31
5.4. Cuantificación del sincronismo entre señales biomédicas .....	32
5.4.1. La secuencia de correlación .....	32
5.4.2. Magnitud de la coherencia espectral .....	33
5.4.3. Información mutua .....	33
5.5. Caracterización de la señal de variabilidad de la frecuencia cardíaca .....	34
5.5.1. Análisis temporal de la señal HRV .....	34
5.5.2. Análisis espectral de la señal HRV .....	35
5.5.3. Análisis no lineal de la señal HRV .....	36
5.6. Caracterización de la señal de saturación de oxígeno en sangre.....	38
5.6.1. Índices de oximetría clásicos.....	38
5.6.2. Estimación de la densidad espectral de potencial .....	38
5.6.3. Análisis no lineal .....	39
5.7. Caracterización del acoplamiento cardiorrespiratorio.....	41
5.8. Selección de características. Algoritmos genéticos.....	41
5.9. Clasificación de características. Red neuronal perceptrón multicapa.....	45
5.10. Análisis estadístico .....	48
5.10.1. Test no paramétrico de Kruskal Wallis .....	48
5.10.2. Medidas de la capacidad predictiva de un test diagnóstico .....	49

**CAPÍTULO 6. RESULTADOS .....55**

6.1. Población bajo estudio .....	55
6.2. Análisis descriptivo de la población bajo estudio .....	56
6.3. Caracterización de la señal de variabilidad de la frecuencia cardíaca .....	60

6.4. Caracterización de la señal de saturación de oxígeno en sangre.....	61
6.5. Caracterización del acoplamiento cardiorrespiratorio.....	63
6.6. Selección de características .....	64
6.7. Red neuronal perceptrón multicapa .....	66
6.7.1. Elección de los hiperparámetros óptimos .....	66
6.7.2. Evaluación de la capacidad diagnóstica del clasificador .....	67
<b>CAPÍTULO 7. DISCUSIÓN.....</b>	<b>71</b>
7.1. Extracción de características y análisis estadístico .....	71
7.2. Selección y clasificación de características .....	73
7.3. Comparación con otros estudios.....	74
7.4. Limitaciones .....	76
<b>CAPÍTULO 8. CONCLUSIONES.....</b>	<b>79</b>
8.1. Contribuciones .....	80
8.2. Principales conclusiones del estudio .....	80
8.3. Líneas futuras de investigación .....	81
<b>ANEXOS.....</b>	<b>83</b>
ANEXO 1. Código para el preprocesado y extracción de variables de la señal de saturación de oxígeno en sangre.....	83
ANEXO 2. Código para el preprocesado y extracción de variables de la señal de variabilidad de la frecuencia cardiaca .....	92
ANEXO 3. Código para el preprocesado y extracción de variables de la señal de EDR para la cuantificación del acoplamiento cardiorrespiratorio.....	98
ANEXO 4. Código para la selección de variables mediante algoritmos genéticos.....	104
ANEXO 5. Código para la clasificación multiclase mediante una red neuronal multicapa....	108
<b>BIBLIOGRAFÍA.....</b>	<b>119</b>





# ÍNDICE DE FIGURAS

<b>Figura 1.</b> Dispositivo Biosignalsplux (Plux Wireless Biosignals S.A., Lisboa, Portugal).....	20
<b>Figura 2.</b> Pulsioxímetro de muñeca NONIN WristOx2 3150 (Nonin Medical Inc., MN, USA).....	21
<b>Figura 3.</b> Ejemplo 1 de un registro de ECG ampliado marcando su envolvente superior en naranja y su envolvente inferior en amarillo.....	22
<b>Figura 4.</b> Ejemplo 2 de un registro de ECG ampliado marcando su envolvente superior en naranja y su envolvente inferior en amarillo.....	22
<b>Figura 5.</b> Registro HRV preprocesado para eliminar las muestras con valores menores a 0.33.	24
<b>Figura 6.</b> Registro HRV preprocesado para eliminar las muestras con valores mayores a 1.50.	24
<b>Figura 7.</b> Registro de HRV preprocesado con muestras fisiológicamente no válidas.....	25
<b>Figura 8.</b> Preprocesado de una señal de oximetría. A la izquierda, un registro original con artefactos. A la derecha, un registro preprocesado sin artefactos. ....	26
<b>Figura 9.</b> Secuencia temporal original segmentada [51]. ....	27
<b>Figura 10.</b> PSD de un registro de HRV calculada mediante el método de Welch.....	35
<b>Figura 11.</b> Comprobación de los valores óptimos de m y r para la SampEn para la señal de HRV. ....	37
<b>Figura 12.</b> Comprobación del valor óptimo del radio para la CTM para la señal de HRV. ....	37
<b>Figura 13.</b> PSD de un registro de la SpO <sub>2</sub> calculada mediante el método de Welch. ....	39
<b>Figura 14.</b> Comprobación de los valores óptimos de m y r para la SampEn para la señal de SpO <sub>2</sub> .....	40
<b>Figura 15.</b> Comprobación del valor óptimo de r para la CTM para la señal de SpO <sub>2</sub> .....	40
<b>Figura 16.</b> Esquema de la partición de la población bajo estudio en conjuntos de entrenamiento y test para la selección de características.....	42

<b>Figura 17.</b> Diagrama de flujo de un algoritmo genético [70].	45
<b>Figura 18.</b> Ejemplo de una MLP genérica $N_h$ nodos en una única capa oculta y $N_o$ salidas [16].	46
<b>Figura 19.</b> Diagrama de flujo de pacientes que forman parte de la población bajo estudio.	56
<b>Figura 20.</b> Diagrama de barras para la variable "Duración de los síntomas".	58
<b>Figura 21.</b> Diagrama de cajas para la variable "TAD".	59
<b>Figura 22.</b> Diagrama de barras de la variable "mMRC".	59
<b>Figura 23.</b> Diagrama de cajas para la variable "CAT".	60
<b>Figura 24.</b> Diagrama de cajas para la variable "CTM ( $SpO_2$ )".	63
<b>Figura 25.</b> Diagrama de cajas para la variable "MCC (EDR mediante envolvente inferior)".	65
<b>Figura 26.</b> Diagrama de barras para la variable "AOS".	65
<b>Figura 27.</b> Diagrama de cajas para la variable "MSC".	65
<b>Figura 28.</b> Diagrama de cajas para la variable "PHF".	65
<b>Figura 29.</b> Diagrama de cajas para la variable "CAT".	65
<b>Figura 30.</b> Gráfica del mejor ajuste obtenido con la función fitness.	66
<b>Figura 31.</b> Elección de los hiperparámetros óptimos ( $N_h = 5$ y $\alpha = 1$ ).	67
<b>Figura 32.</b> Matriz de confusión multiclase para evaluar la eficacia del clasificador sobre el conjunto test.	67
<b>Figura 33.</b> Matriz de confusión de clasificación binaria (No COVID vs. (COVID + SPC-CP)).	68
<b>Figura 34.</b> Matriz de confusión de clasificación binaria (No COVID + COVID) vs. SPC-CP).	69

# ÍNDICE DE TABLAS

<b>Tabla 1.</b> Resumen de la presentación clínica más frecuente del SPC-CP [20], [23], [25].	11
<b>Tabla 2.</b> Distribución de una matriz de confusión de dos clases.	52
<b>Tabla 3.1.</b> Mediana, rango intercuartil y p - valor de las características clínicas para los tres grupos bajo estudio.	56
<b>Tabla 3.2.</b> Mediana, rango intercuartil y p - valor de las características clínicas para los tres grupos bajo estudio.	57
<b>Tabla 4.</b> Análisis post-hoc de la variable "Duración de los síntomas".	58
<b>Tabla 5.</b> Análisis <i>post-hoc</i> de la variable "TAD".	58
<b>Tabla 6.</b> Análisis <i>post-hoc</i> de la variable "mMRC".	59
<b>Tabla 7.</b> Análisis <i>post-hoc</i> de la variable "CAT".	60
<b>Tabla 8.1.</b> Mediana, rango intercuartil y p-valor de las características de la señal de HRV para los tres grupos bajo estudio.	60
<b>Tabla 8.2.</b> Mediana, rango intercuartil y p-valor de las características de la señal de HRV para los tres grupos bajo estudio.	61
<b>Tabla 9.1.</b> Mediana, rango intercuartil y p - valor de las características de la señal de SpO <sub>2</sub> para los tres grupos bajo estudio.	61
<b>Tabla 9.2.</b> Mediana, rango intercuartil y p - valor de las características de la señal de SpO <sub>2</sub> para los tres grupos bajo estudio.	62
<b>Tabla 10.</b> Análisis post-hoc de la variable "CTM".	62
<b>Tabla 11.</b> Mediana, rango intercuartil y p - valor de las características del CRC (EDR mediante la envoltante superior) para los tres grupos bajo estudio.	63
<b>Tabla 12.</b> Mediana, rango intercuartil y p - valor de las características del CRC (EDR mediante la envoltante inferior) para los tres grupos bajo estudio.	64
<b>Tabla 13.</b> Análisis <i>post-hoc</i> de la variable "MCC (EDR mediante envoltante inferior)".	64

<b>Tabla 14.</b> Métricas utilizadas para evaluar la eficacia del clasificador. ....	68
<b>Tabla 15.</b> Métricas utilizadas para la evaluación del clasificador agrupando clases (No COVID vs. COVID + SPC-CP). ....	69
<b>Tabla 16.</b> Métricas utilizadas para la evaluación del clasificador agrupando clases (No COVID + COVID) vs. SPC-CP). ....	69
<b>Tabla 17.</b> Comparación de las variables significativas obtenidas en diferentes estudios. ....	75

# Glosario de siglas y acrónimos

<b>SPC – PC</b>	Síndrome post – COVID / COVID persistente
<b>S</b>	Espiga
<b>E</b>	Envoltura
<b>M</b>	Membrana
<b>N</b>	Nucleocápside
<b>HRV</b>	Variabilidad de la frecuencia cardiaca
<b>SpO<sub>2</sub></b>	Saturación de oxígeno en sangre
<b>ECG</b>	Electrocardiograma
<b>EDR</b>	Respiración derivada del ECG
<b>CRC</b>	Acoplamiento cardiorrespiratorio
<b>GAs</b>	Algoritmo(s) genético(s)
<b>MLP</b>	Red neuronal perceptrón multicapa
<b>ACE2</b>	Encima transformadora de angiotensina
<b>NICE</b>	<i>National Institute for Health and Care Excellence</i>
<b>INSS</b>	Instituto Nacional de la Seguridad Social
<b>IMC</b>	Índice de Masa Corporal
<b>RAAS</b>	Sistema renina angiotensina aldosterona
<b>POTS</b>	Síndrome de taquicardia postural ortostática
<b>UCI</b>	Unidad de cuidados intensivos
<b>SNA</b>	Sistema nervioso autónomo
<b>AVNN</b>	<i>Average NN intervals</i>
<b>SDNN</b>	<i>Standard derivation of NN intervals</i>
<b>RMDSS</b>	<i>Root mean square of the differences of successive NN intervals</i>
<b>VLF</b>	<i>Very low frequency</i>
<b>LF</b>	<i>Low frequency</i>
<b>HF</b>	<i>High frequency</i>
<b>P<sub>VLF</sub></b>	Potencia total en la banda VLF

<b>P<sub>LF</sub></b>	Potencia total en la banda LF
<b>P<sub>HF</sub></b>	Potencia total en la banda HF
<b>P<sub>LF/HF</sub></b>	Relación entre la P <sub>LF</sub> y la P <sub>HF</sub>
<b>RSA</b>	<i>Respiratory sinus arrhythmia</i>
<b>PCR</b>	Reacción en cadena de la polimerasa
<b>EPOC</b>	Enfermedad pulmonar obstructiva crónica
<b>CELM</b>	Comité de Ética de Investigación con Medicamentos
<b>TAS</b>	Tensión arterial sistólica
<b>TAD</b>	Tensión arterial diastólica
<b>HTA</b>	Hipertensión arterial
<b>DM</b>	Diabetes mellitus
<b>AOS</b>	Apnea obstructiva del sueño
<b>mMRC</b>	Escala de disnea modificada del <i>Medical Research Council</i>
<b>CAT</b>	<i>COPD Assessment Test</i> (test de calidad de vida)
<b>DFT</b>	<i>Discrete fourier transform</i>
<b>FFT</b>	<i>Fast fourier transform</i>
<b>STFT</b>	<i>Short time fourier transform</i>
<b>PSD</b>	<i>Power spectral density</i>
<b>PT</b>	Potencia total
<b>MF</b>	Frecuencia mediana
<b>SE</b>	Entropía espectral
<b>SampEn</b>	Entropía muestral
<b>CTM</b>	Medida de la tendencia central
<b>LZC</b>	Complejidad de Lempel – Ziv
<b>MCC</b>	Secuencia de correlación cruzada
<b>MSC</b>	Coherencia espectral
<b>MI</b>	Información mutua
<b>CT95</b>	<i>Cumulative time below 95%</i>
<b>CT90</b>	<i>Cumulative time below 90%</i>
<b>ODI</b>	<i>Oxygen desaturation index</i>
<b>P<sub>c</sub></b>	Probabilidad de cruce
<b>P<sub>m</sub></b>	Probabilidad de mutación
<b>Nh</b>	Número de nodos en la capa oculta
<b>α</b>	Tasa de aprendizaje
<b>IQR</b>	Rango intercuartil

<b>KW</b>	Kruskal Wallis
<b>Acc</b>	Índice de precisión ( <i>accuracy</i> )
<b>K</b>	Índice Kappa
<b>MF1S</b>	Índice Macro F1 Score
<b>TP</b>	Verdaderos positivos
<b>FN</b>	Falsos negativos
<b>TN</b>	Verdaderos negativos
<b>FP</b>	Falsos positivos
<b>Se</b>	Sensibilidad
<b>Sp</b>	Especificidad
<b>PPV</b>	Valor predictivo positivo
<b>NPV</b>	Valor predictivo negativo
<b>LR+</b>	Razón de verosimilitud positiva
<b>LR-</b>	Razón de verosimilitud negativa
<b>PBT</b>	Temperatura corporal máxima





# CAPÍTULO 1. INTRODUCCIÓN

## 1.1. Infección por virus SARS-CoV-2 y COVID persistente

Los coronavirus fueron identificados por primera vez en humanos en 1965 durante la búsqueda del agente responsable del resfriado común. Su nombre se deriva de su apariencia bajo el microscopio electrónico, ya que exhibe una forma esférica con proyecciones en forma de espigas que les proporcionan una apariencia similar a una corona. Tienen una amplia distribución en la naturaleza y se encuentran muy dispersos tanto en humanos como en animales. Pertenecen al orden Nidovirales, familia Coronaviridae, y subfamilia Orthocoronavirinae. Se distinguen cuatro géneros, incluyendo los alphacoronavirus y betacoronavirus. En los seres humanos, se han identificado siete especies de coronavirus que causan infecciones, dos de ellas pertenecientes al género alphacoronavirus y las demás al género betacoronavirus. Los análisis filogenéticos han clasificado al SARS-CoV2 en el subgénero de los sarbecovirus, dentro del género betacoronavirus de la familia Coronaviridae [1].

Los coronavirus presentan su impacto principal en las vías respiratorias, provocan infecciones en el tracto respiratorio superior, como el resfriado común. Por lo general, estas infecciones son leves. Su afectación en el aparato respiratorio se debe a su afinidad por las células epiteliales ciliadas respiratorias, aunque también pueden causar trastornos en el tracto gastrointestinal. Estos virus muestran una prevalencia estacional, siendo más comunes durante el invierno, aprovechando las bajas temperaturas y la cercanía entre las personas para propagarse. Además, presentan un carácter endémico, es decir, están establecidos de manera habitual en ciertas poblaciones [2].

Los coronavirus tienen una estructura compuesta por una cápside de proteínas, una envoltura lipídica que contiene espículas de glicoproteínas en su superficie, y un ácido ribonucleico (RNA) monocatenario de polaridad positiva. Estos virus se caracterizan por su gran tamaño y su capacidad de recombinarse con otros virus de RNA, lo que les permite cambiar de huésped [3]. El genoma del virus codifica cuatro proteínas estructurales que desempeñan un papel importante tanto en la estructura del virus como en su replicación. Estas proteínas se denominan S (espiga), E (envoltura), M (membrana) y N (nucleocápside). La proteína S es crucial para la unión del virus a la célula que infectará, lo que determina el tropismo viral y su capacidad para acceder al retículo endoplásmico de la célula infectada. Esta proteína es responsable de la apariencia distintiva en forma de corona del virus. La proteína E se encuentra en cantidades pequeñas, pero desempeña un papel importante en la maduración y liberación de las partículas

virales de la célula hospedadora. La proteína M es la más abundante y está asociada a la envoltura viral, definiendo su forma. La proteína N se asocia con el RNA viral y es responsable de su empaquetamiento, así como de regular la síntesis de RNA viral [4].

Aunque los coronavirus generalmente tienen baja patogenicidad, ha habido dos ocasiones en las que han mostrado un carácter epidémico antes de la pandemia producida en 2019. La primera ocurrió en China en 2002, dando origen a la "enfermedad aguda grave por coronavirus (SARS-CoV)" transmitida por gatos, con una tasa de mortalidad aproximada del 9.6%. La segunda epidemia tuvo lugar en Arabia Saudí en 2012, denominada MERS-CoV (Middle East respiratory syndrome coronavirus), debido a su aparición en Oriente Medio. Inicialmente se atribuyó su origen a los murciélagos, pero afectó a los camellos dromedarios. Aunque la transmisión entre humanos fue limitada, la tasa de mortalidad fue significativa, cercana al 35%. La enfermedad afectó a varios países, pero todos los casos estaban relacionados con estancias en la península arábiga. En ambos casos, se asoció con una enfermedad respiratoria grave y una alta mortalidad [5].

El virus coronavirus de tipo 2 asociado al síndrome respiratorio agudo grave, conocido como SARS-CoV-2, es el agente causal del COVID-19, síndrome descubierto en China en diciembre del año 2019 [6]. Esta enfermedad por COVID-19 se ha propagado alrededor del mundo desde su foco inicial. Las cifras ascendían a más de 183 millones de casos confirmados y registrados en todo el mundo en julio del 2021, con más de 3,97 millones de muertes reportadas por la Organización Mundial de la Salud. Esta enfermedad sigue siendo uno de los desafíos más importantes para la comunidad médica, tanto porque su estudio y comprensión todavía es insuficiente, como porque representa una gran parte de la carga asistencial. Entre el 10% y el 35% de los pacientes que se han recuperado de COVID-19 refieren la persistencia de al menos un síntoma después de tres semanas de seguimiento a partir del diagnóstico, destacando como dos de los síntomas más frecuentes la fatiga y la disnea [7]. Esta incidencia se incrementa hasta el 50 – 84% en caso de que los pacientes hayan requerido hospitalización durante el periodo de infección. La afectación de la enfermedad se ha reportado en diferentes áreas, incluyendo la neurocognitiva (confusión, pérdida de atención), autonómica (taquicardia, palpitaciones), gastrointestinal (dolor abdominal, vómitos), musculoesquelética (mialgias), psicológica (ansiedad, depresión) y, por su puesto, respiratoria (fatiga, disnea, tos, dolor de garganta) [8]. Esta situación clínica, que incluye muchos otros síntomas, se intenta englobar en el denominado SPC-CP, aunque a día de hoy no existe una definición estándar ni consenso en cuanto a la temporalidad de los síntomas [8], [9], [10]. Los datos sobre las incidencias conservadoras de este trastorno estiman que unas 450.000 – 900.000 personas en España pueden estar afectadas, tanto respecto a su salud como a su calidad de vida, debido a la persistencia de los síntomas de COVID-19 [8].

Diferentes estudios han demostrado que el SARS-CoV-2 afecta al sistema autónomo [8], [11], [12]. Una forma de determinar esta afectación es mediante la variabilidad de la frecuencia (*heart rate variability*, HRV). El análisis de esta señal ha servido para predecir el diagnóstico de COVID-19, la supervivencia y los ingresos a la unidad cuidados intensivos en la primera semana de hospitalización [13], [14]. En otros estudios se ha observado que los pacientes con síndrome post-COVID exhiben una desregulación de la HRV que puede explicar los síntomas que los

aquejan [11], [12]. Sin embargo, ninguno de ellos ha tenido en cuenta el incluir como un grupo control a pacientes que hayan tenido COVID-19 sin síntomas a largo plazo.

## 1.2. Análisis de datos clínicos y técnicas de inteligencia artificial

En el presente trabajo se realiza un estudio sobre la señal HRV, saturación de oxígeno en sangre (SpO<sub>2</sub>) y respiración derivada del electrocardiograma (ECG) (*ECG-derived respiration*, EDR). Las dos últimas señales serán utilizadas para evaluar para la interacción entre la actividad cardiaca y la respiratoria mediante el acoplamiento cardiorrespiratorio (*cardiorespiratory coupling*, CRC). La elección de estas señales es debida a que los síntomas más frecuentes entre los pacientes con SPC-CP son los respiratorios (disnea) y los cardiacos (arritmias).

A la hora de trabajar con los registros se plantean tres etapas: extracción de características, selección de variables y reconocimiento automático de patrones. En la fase de extracción referente a la señal de HRV se aplicaron técnicas temporales, espectrales y no lineales. En el caso de la señal de SpO<sub>2</sub>, se aplicaron técnicas espectrales y no lineales, además de extraer los índices clásicos de esta señal en el dominio del tiempo. Por último, para el estudio del acoplamiento también se aplicaron técnicas temporales, espectrales y no lineales, siempre con el objetivo de obtener información complementaria a partir de diferentes enfoques de procesado.

Una vez extraídas las características, tanto las procesadas como las variables clínicas, se implementa un procedimiento para la selección de variables, basado en algoritmos genéticos (GAs), siendo un método de optimización estocástico perteneciente a la computación evolutiva [15]. Finalmente, el algoritmo de clasificación multiclase se plantea mediante una red neuronal perceptrón multicapa (MLP), la cual es una red neuronal artificial en la que se construye un modelo inspirado en cómo el cerebro humano almacena y estructura la información mediante la interconexión de sus neuronas (nodos de la red) [16]. Mediante este método de reconocimiento automático de patrones se busca identificar los 3 grupos de pacientes bajo estudio.

## 1.3. Estructura

La memoria del presente Trabajo de Fin de Grado se ha estructurado siguiendo los pasos habituales establecidos en una investigación que conlleve relación con las señales biomédicas.

En este primer capítulo se presenta el problema bajo estudio, recalcando su relevancia y prevalencia en los últimos años desde su aparición. El considerable porcentaje de población que padece este síndrome y las numerosas incógnitas que todavía no se han resuelto son algunos de los factores que motivan este estudio.

En el segundo, se desarrolla el estado del arte del COVID-19 y del SPC-CP, haciendo énfasis en los conceptos y síntomas claves para entender mejor la enfermedad. También se explicarán las tres señales a analizar y la potencial relación de estas con el SPC-CP.

En el tercer capítulo se plantea las hipótesis del trabajo teniendo en cuenta que entre los síntomas más frecuentes del denominado SPC-CP están los respiratorios (disnea) y cardiovasculares (arritmias). Se pretende corroborar estas hipótesis mediante los objetivos planteados, entre los que destacan identificar las variables clínicas y/o derivadas del procesado de señal que más se asocian al padecimiento de SPC-CP, así como desarrollar y validar un modelo predictivo de padecimiento de SPC-CP.

Los criterios de inclusión y exclusión de la población bajo estudio, así como los grupos particulares de pacientes que se han analizado en este TFG, de describirán en el cuarto capítulo. También se detallarán las señales biomédicas registradas y los dispositivos utilizados para ello.

En el quinto se describen los métodos de procesado de señal aplicados en este estudio, profundizando en las técnicas temporales, espectrales y no lineales. También se muestra el proceso de selección de características, el algoritmo de red neuronal utilizado y los test estadísticos. Finalmente, se describen las métricas de rendimiento diagnóstico empleadas para evaluar las predicciones obtenidas.

En el sexto capítulo se exponen los resultados obtenidos comenzando por un análisis descriptivo de la población bajo estudio a través de las variables extraídas. A su vez se muestran los resultados del procedimiento de selección de variables, entrenamiento, validación y test del clasificador multiclase desarrollado.

La discusión de los resultados alcanzados se relata en el séptimo capítulo, buscando justificar las razones por las cuales se han obtenido esos datos. Igualmente, también se comparan los resultados de este TFG con los de otros estudios similares.

Finalmente, las conclusiones a las que se ha llegado se presentan en el octavo capítulo, destacando los hallazgos realizados, las limitaciones que podrían condicionar la generalización de los mismos y las futuras líneas de investigación que pueden seguirse a raíz del presente estudio.

## CAPÍTULO 2. SÍNDROME POST – COVID

En este primer capítulo, con el objetivo de comprender la enfermedad, se describirán en primer lugar los antecedentes del SPC-CP para poder entender mejor el origen y la importancia del estudio de esta enfermedad debido a la gran magnitud que presenta el problema. También se abordará la fisiopatología, la sintomatología y el seguimiento y tratamiento actual de la enfermedad.

Por último, se realizará una revisión de las tres señales que se tratan, las cuales son la SpO<sub>2</sub>, la HRV y la EDR, estas últimas derivadas ambas del registro de ECG, junto con sus actuales evidencias en relación con esta condición.

### 2.1. Síndrome post – COVID

El virus coronavirus de tipo 2 asociado al síndrome respiratorio agudo grave, conocido como SARS-CoV-2, es el agente causal de la COVID-19, síndrome descubierto en China en diciembre del año 2019 [6]. Esta enfermedad por COVID-19 se ha propagado alrededor del mundo desde su foco inicial. Las cifras ascendían a más de 183 millones de casos confirmados y registrados en todo el mundo en julio del 2021, con más de 3,97 millones de muertes reportadas por la Organización Mundial de la Salud. El espectro clínico de la COVID-19 puede llegar a variar desde una infección asintomática hasta una enfermedad mortal. El virus SARS-CoV-2 ingresa en las células a través del receptor de la enzima transformadora de angiotensina 2 (ACE2) y, una vez dentro, el virus se replica y madura, provocando una respuesta inflamatoria que involucra la activación e infiltración de células inmunitarias por varias citoquinas en algunos pacientes. El receptor ACE2 está presente en numerosos tipos celulares en todo el organismo, incluyéndose la mucosa oral y nasal, pulmones, corazón, tracto gastrointestinal, hígado, riñones, bazo, cerebro y las células endoteliales arteriales y venosas. Por tanto, es importante recalcar que el SARS-CoV-2 puede causar daño a múltiples órganos [17].

El impacto que ha tenido el COVID-19 hasta ahora ha sido inigualable y, además, los síntomas a largo plazo podrían tener un efecto perjudicial [17]. Múltiples estudios estiman que entre un 10 y un 20% de los pacientes que han sufrido COVID-19 presentan sintomatología persistente meses después de la infección inicial, siendo esta condición conocida como SPC-CP. El *National Institute for Health and Care Excellence (NICE)* define como SPC-CP a los síntomas que continúan o se desarrollan después de una infección por COVID -19 y que no se puede explicar por un

diagnóstico alternativo, incluyendo este término como SPC-CP a los síntomas que continúan más allá de las 12 semanas posteriores a la infección, exhibiendo las personas que lo padecen un deterioro en la estructura y en la función de los múltiples órganos nombrados anteriormente [6], [17].

En la literatura médico-científica se definen 2 perfiles de pacientes compatibles con la enfermedad [17], [18]:

- Pacientes que presentan persistencia de síntomas, los cuales no necesariamente han tenido que estar presentes al inicio de la infección, después de 4 semanas desde la infección por SARS-CoV-2, pudiendo tener un curso de los síntomas permanente, remitente o presentando una mejora progresiva.
- Pacientes con SPC-CP, caracterizados por daño irreversible en los tejidos después de 12 semanas desde la infección por SARS-CoV-2. La disfunción permanente podrá darse de distintas formas y grados dependiendo del paciente, así como la sintomatología también podrá ser variante entre sujetos.

Actualmente, no se conocen las causas específicas que provocan el SPC-CP, aunque se han propuesto como posibles mecanismos subyacentes los cambios fisiopatológicos provocados por el SARS-CoV-2, las variaciones inmunológicas posteriores a la interacción dada entre el virus y el huésped y el daño inflamatorio producido como respuesta a la infección aguda [18]. La prevalencia y la sintomatología presentada por el SPC-CP son muy heterogéneas, aunque destacan como síntomas con mayor frecuencia la fatiga (en un 52% de los casos), los síntomas cardiorrespiratorios (en un 30-42% de los casos) y síntomas neurológicos (en un 40% de los casos) [7], [18].

### **2.1.1. Magnitud del problema**

El SPC-CP se presenta en grupos de pacientes muy diversos, pudiendo darse en pacientes hospitalizados con enfermedad COVID-19 de gravedad variable, pacientes que no han sido hospitalizados pero que sí han presentado sintomatología característica o incluso en pacientes que han pasado la enfermedad de manera asintomática [18].

Existen varios estudios recientes sobre la prevalencia del SPC-CP en pacientes tanto hospitalizados como sin hospitalizar [18], [19], pero estimar esta prevalencia de síntomas persistentes no es tarea fácil debido a las diferencias entre las distintas poblaciones analizadas y a los métodos utilizados. Por lo tanto, las conclusiones y evidencias varían enormemente entre diferentes estudios [18], aunque se puede concluir que además de los síntomas físicos o cognitivos, el SPC-CP afecta a la calidad de vida, generando una discapacidad importante con implicaciones económicas y de dificultad de retorno laboral [19]. Por tanto, la magnitud del SPC-CP también queda reflejada en su impacto en el mundo laboral. La incapacidad temporal por enfermedad es reconocida por el médico que examina a la persona mediante tablas informativas del Instituto Nacional de la Seguridad Social (INSS) que definen qué limitaciones pueden conducir a discapacidades laborales [18]. Sin embargo, en el caso específico del SPC-CP no hay protocolos establecidos significando que la vuelta al trabajo se basa en el tipo y la intensidad de

los síntomas, así como la actividad laboral de cada persona. Por tanto, no hay un diagnóstico de incapacidad específico para el SPC-CP aunque este tiene un grave impacto en la capacidad de las personas para volver al trabajo, afectando el ámbito mental, social y económico [18], [19].

### **2.1.2. Mecanismos que contribuyen a la fisiopatología del síndrome post – COVID**

El mecanismo exacto causante de la persistencia de los síntomas y por tanto del SPC-CP es difícil de identificar, pudiendo ser la razón de la persistencia de los síntomas las secuelas derivadas del daño de los órganos, la extensión de la lesión y el tiempo necesario para la recuperación de cada sistema de órganos, la persistencia de la inflamación crónica o la respuesta inmune, la persistencia del virus en el cuerpo, efecto inespecífico de la hospitalización, secuelas de enfermedad crítica, el síndrome de cuidados post-intensivos, las complicaciones relacionadas con la infección, las complicaciones relacionadas con comorbilidades o los efectos adversos de los medicamentos utilizados [20], [21].

Desde el punto de vista de la salud pública, es muy importante distinguir los síntomas residuales a los síntomas provocados por una reinfección [20]. La persistencia de la infección puede deberse a la infección resistente debido a viremia en personas con inmunidad alterada o por una reinfección o recaída [22]. Además, existen otros factores psico-sociales que también contribuyen a los síntomas: problemas psicológicos, estrés postraumático, impacto social y financiero de COVID-19 [20], [22].

### **2.1.3. Manifestaciones clínicas**

Existe una amplia heterogeneidad en los síntomas que presentan los pacientes con SPC-CP, pudiendo destacar como los más frecuentes la presencia de fatiga (52%), los problemas cardiacos y respiratorios (30-42%) y los síntomas neurológicos (40%). Es importante distinguir entre aquellas manifestaciones clínicas producidas como una secuela del COVID-19 y aquellas manifestaciones pertenecientes a la sintomatología del SPC-CP [23].

Evaluando esta sintomatología del SPC-CP se encuentra fatiga intensa, dificultad para respirar, tos, dolor en el pecho, palpitaciones, dolor de cabeza, dolor articular, mialgia y debilidad, insomnio, hormigueo, diarrea, sarpullido o caída del cabello, deterioro del equilibrio y de la marcha, así como problemas neurocognitivos [20]. Las mujeres fueron significativamente más propensas que los hombres a referir ansiedad, bajo estado de ánimo, mialgia, fatiga, trastornos del sueño y deterioro de la memoria. Un índice de masa corporal (IMC) más alto se asoció con mialgia y fatiga [24]. Diferentes investigaciones han identificado dos patrones principales de síntomas en personas con SPC-CP:

- 1) Fatiga, dolor de cabeza y molestias en las vías respiratorias superiores (dificultad para respirar, dolor de garganta, tos persistente y pérdida del olfato) [20], [24].
- 2) Molestias multi-sistémicas, incluyendo fiebre continua y síntomas gastroenterológicos [20], [24].

En pacientes con SPC-CP algunos de los síntomas se informan por primera vez entre 3 y 4 semanas después del inicio de los síntomas agudos [24]. Se puede distinguir estos síntomas de forma más específica según la parte del organismo que se vea involucrada:

- **Síntomas respiratorios.** El pulmón es el principal órgano afectado por la infección por COVID-19 y, en consecuencia, los síntomas respiratorios y las limitaciones a la hora de realizar esfuerzos físicos son frecuentes en el SPC-CP, siendo los síntomas pulmonares más frecuentes tras la infección disnea (33%), tos (20%) y dolor torácico (30%) [20]. Se desconoce la proporción de síntomas que están claramente asociados con secuelas pulmonares y aquellos en ausencia de daño pulmonar persistente [25].
- **Lesiones en la piel y en las mucosas.** Las lesiones en la piel y mucosas, la pérdida de cabello (a menudo transitoria), las aftas bucales, así como las lesiones y sequedad en la piel pueden aparecer después de la infección en casi un tercio de los pacientes. Estos síntomas deben ser evaluados en caso de que estén relacionados con una etiología endocrinológica o inmunológica [20], [25].
- **Síntomas musculoesqueléticos.** Las artralgias son muy comunes pudiendo estar presentes en casi el 50% de los casos. La importancia de este tipo de síntomas recae en que pueden ser discapacitantes, siendo necesario identificar el número y tipo de articulación, el patrón de presentación, especialmente si están asociadas con otros síntomas inflamatorios, verificar la debilidad y la elevación de los biomarcadores musculares. Entre el 20 y el 30% de los pacientes informan de mialgias [25].
- **Síntomas digestivos.** Los síntomas digestivos presentan una frecuencia de entre el 10 y el 17%. Pueden manifestarse en forma de "brotes", con días de malestar abdominal y diarrea, mientras otros días presentan completa normalidad. Parte de la sintomatología probablemente se deba a un cambio en la microbiota [25].
- **Síntomas cardiológicos.** El dolor torácico, presente en el 30% de los casos, requiere que se descarte primero si es consecuencia de una enfermedad coronaria, especialmente en pacientes con factores de riesgo cardiovascular. La inflamación inmunitaria que afecta al miocardio y al pericardio son los mecanismos más comunes de daño cardiovascular que resultan en síntomas clínicos como disnea, fatiga, miocarditis, disminución de la reserva cardíaca, disfunción del sistema renina-angiotensina-aldosterona (RAAS), disfunción autonómica y arritmias [23], [26]. En caso de dolor pleurítico, se debe considerar la posibilidad de serositis. Si ocurren palpitaciones recurrentes, se debe considerar la posibilidad de una taquicardia sinusal inapropiada, así como un Síndrome de Taquicardia Ortostática Postural (POTS) si está asociado con síntomas ortostáticos [25].
- **Trastornos vasculares y hematológicos.** La trombosis es una complicación muy común durante la infección aguda, alrededor del 31.3% del tromboembolismo venoso incluye un 19.8% de trombosis venosa profunda y un 18.9% de trombosis pulmonar. Sin embargo, este porcentaje de prevalencia disminuye después del alta hospitalaria,



alcanzando valores alrededor del 0.8 al 2.5%. Los pacientes con secuelas y movilidad secundaria limitada tienen un alto riesgo de trombosis y deben ser monitorizados [25]. El tromboembolismo agudo asociado a COVID-19 es secundario al estado hiperinflamatorio e hipercoagulable. La hipoxia, la lesión endotelial, la activación plaquetaria y las citoquinas proinflamatorias provocan una tromboembolia desproporcionadamente alta en la fase aguda del COVID-19. Tanto la duración como la gravedad de este estado hiperinflamatorio contribuyen al riesgo de complicaciones trombóticas en el SPC-CP [23]. En cuanto al aspecto hematológico, el COVID-19 se asocia con linfopenia, así como trombocitopenia, aunque en la gran mayoría de los casos estas se resuelven posteriormente (7%). Por lo tanto, si persisten o aparecen más tarde, deben ser estudiadas y se debe descartar un origen inmunomediado, hemolítico o trombótico [25].

- **Complicaciones oftalmológicas.** Los cambios relacionados con la agudeza visual o incluso con la visión borrosa presentan una frecuencia del 27%, siendo los síntomas más frecuentes. Sin embargo, aunque en menor medida, también se describen trastornos graves que podrían estar asociados con otros trastornos como fotopsia, miodesopsia o metamorfopsia [25].
  
- **Síntomas neurológicos.** Diversos estudios sugieren que los trombos microvasculares, la inflamación sistémica y la neurotoxicidad directa mediada por el virus son los posibles mecanismos que contribuyen a la neuropatología en COVID-19 [23]. La pérdida del olfato y del gusto son muy comunes y prolongadas en el tiempo, entre el 16% y el 22% de los casos. Los déficits quimiosensoriales a menudo son los primeros y únicos signos en portadores asintomáticos del virus SARS-CoV-2. Con respecto a la disfunción olfativa, pasados entre un mes y 6 meses, el 15.3% y el 4.7% de los pacientes anosmicos/hipósmicos, respectivamente, no recuperaron el sentido del olfato por completo. El dolor de cabeza, también frecuente en un 33%, es predominantemente frontoparietal y a menudo se controla con analgesia estándar. Sin embargo, hay algunos casos que pueden ser refractarios o asociados con presencia de vómitos, fiebre o interrupción del sueño, siendo necesario que sean examinados preferentemente para descartar problemas graves. Siempre se debe explorar la fuerza física, la sensibilidad, hallazgos relacionados con la disautonomía (taquicardia inapropiada, cambios en la sudoración, ortostasis) que pueden aparecer alrededor del 16% de los pacientes, posibles alteraciones neurocognitivas (problemas de memoria a corto plazo, concentración/atención, orientación) y trastornos del sueño en casi el 40% de los pacientes con SPC-CP [25]. Esta disautonomía, el desacondicionamiento y el trastorno de estrés postraumático pueden contribuir a la niebla mental posterior a la COVID-19. Sin embargo, es importante tener en cuenta que parte del deterioro cognitivo a largo plazo en pacientes que han sufrido COVID-19 también puede ser debido a la duración prolongada de la estancia en la Unidad de Cuidados Intensivos (UCI), ya que la intubación prolongada contribuye significativamente a este deterioro descrito [23].
  
- **Síntomas renales.** Durante la hospitalización, alrededor de un 20% de los pacientes graves por COVID-19 que requieren intubación también necesitan someterse a terapia

de reemplazo renal. Sin embargo, la mayoría de estos pacientes no requieren diálisis al momento de recibir el alta hospitalaria ni en los casos de SPC-CP [23].

- **Síntomas endocrinos.** Las lesiones causadas por el virus, el daño inflamatorio y el daño inmunológico contribuyen a las manifestaciones endocrinas que se presentan en el SPC-CP. Durante la fase aguda y la fase de recuperación posterior a la COVID-19, la inmovilización, el uso de esteroides y la deficiencia de vitamina D podrían contribuir a la desmineralización ósea [23].
- **Trastornos autoinmunes.** Los pacientes con COVID-19 muestran un aumento en las reactividades de autoanticuerpos en comparación con los controles no infectados, con una alta prevalencia de autoanticuerpos contra proteínas inmunomoduladoras, incluyendo citocinas, quimiocinas, componentes del complemento y proteínas de superficie celular. Algunos autoanticuerpos contribuyen a la fisiopatología de COVID-19 al antagonizar la respuesta antiviral innata y pueden contribuir a la gravedad de la enfermedad. Además, se han descrito informes de casos y series de trastornos autoinmunes, no solo con afectación de un solo órgano (tiroiditis, diabetes, etc.), sino también con manifestaciones sistémicas (vasculitis, miopatía inflamatoria o lupus) [20], [25].
- **Síntomas psicológicos.** Existe una alta prevalencia de ansiedad, depresión y trastornos del estado de ánimo, en prácticamente un 40% de los pacientes, existiendo incluso varios casos de síndrome de estrés postraumático [25].

En la Tabla 1 se recoge de forma resumida la presentación clínica que presenta el SPC-CP con más frecuencia.

#### 2.1.4. Tratamiento y seguimiento de la enfermedad

Una vez detectados síntomas prolongados en el tiempo por más de 4 semanas desde la infección (o 12 semanas en caso de secuelas), es fundamental llevar a cabo un estudio completo para caracterizar las variables clínicas y descartar complicaciones u otras etiologías graves que puedan explicar los síntomas prolongados. Por lo general, se establece una tendencia hacia la mejora de los síntomas, pudiendo la fatiga ser uno de los más duraderos. Por el contrario, otros síntomas pueden disminuir en intensidad e incluso desaparecer por completo [25].

Todavía no es posible confirmar el porcentaje de pacientes que mejoran y el tiempo necesario para lograr este objetivo debido a la gran heterogeneidad de los estudios publicados hasta la fecha. Hay algunos estudios prospectivos que han demostrado que la vacunación puede producir una mejora significativa en los síntomas, aunque son necesarias más evidencias para confirmar este enfoque completamente [27].

**Tabla 1.** Resumen de la presentación clínica más frecuente del SPC-CP [20], [23], [25].

RESUMEN DE LA PRESENTACIÓN CLÍNICA MÁS FRECUENTE DEL SPC-CP	
<b>Síntomas respiratorios</b>	Disnea, tos, fatiga, dolor torácico, embolia pulmonar, atelectasia/fibrosis pulmonar, neumonía bacteriana posviral.
<b>Lesiones en la piel y en las mucosas</b>	Lesiones vesiculares, maculopapulares, urticariales o similares a sabañones en las extremidades, pérdida de cabello.
<b>Síntomas musculoesqueléticos</b>	Dolores musculares y debilidad, artralgia.
<b>Síntomas digestivos</b>	Malestar abdominal, diarrea, estreñimiento, vómitos.
<b>Síntomas cardiológicos</b>	Dolor torácico, palpitaciones recurrentes, enfermedad isquémica del corazón, miocarditis posviral, fibrosis miocárdica/cicatrización, insuficiencia cardíaca congestiva, arritmias.
<b>Trastornos vasculares y hematológicos</b>	Trombosis venosa profunda, trombosis pulmonar, tromboembolismo, linfopenia, trombocitopenia.
<b>Complicaciones oftalmológicas</b>	Cambios en la agudeza visual y visión borrosa. En menor medida ftopsia, miodesopsia o metamorfopsia.
<b>Síntomas neurológicos</b>	Dolores de cabeza, anosmia, ageusia, dificultades neurocognitivas, insomnio.
<b>Trastornos autoinmunes</b>	Fiebre, síntomas gastrointestinales, erupción cutánea, dolor en el pecho, palpitaciones, infecciones bacterianas y fúngicas.
<b>Síntomas psicológicos</b>	Ansiedad, depresión, trastornos del estado del ánimo, síndrome de estrés postraumático.

SPC-CP: síndrome post-COVID.

Muchos estudios también abalan que existe una correlación insuficiente entre la gravedad del COVID-19 durante la hospitalización y la carga de síntomas en el posterior seguimiento de los casos de SPC-CP, ya que muchas de las definiciones del SPC-CP muestran una variedad de síntomas que son distintos de los de los pacientes que se recuperan de un COVID-19 grave que requirieron hospitalización. También se da una ausencia de asociación entre la carga de síntomas y anomalías radiográficas o bioquímicas [28].

Debido a la incertidumbre que sigue habiendo respecto a cómo tratar este síndrome, la comunidad médica y científica demanda el establecimiento de estrategias y protocolos que ayuden a la detección precoz de los síntomas, con el principal objetivo de mejorar la calidad de vida del paciente [28]. Algunas de las medidas que se podrían tomar en los centros de salud serían:

- **Caracterización clínica.** Realizar una evaluación completa de los síntomas persistentes o secuelas que presente el paciente, pudiendo descartar complicaciones o causas graves subyacentes [28], [29].

- **Pruebas diagnósticas.** Realizar pruebas según los síntomas y las complicaciones específicas. Algunos ejemplos de estas pruebas serían: análisis de sangre, pruebas de función pulmonar, pruebas cardíacas, pruebas de función cognitiva o pruebas de imagen, entre otras [28], [29].
- **Derivación y colaboración interdisciplinaria.** Derivar a los pacientes a especialistas según su sintomatología y necesidades. Algunas especialidades destacables serían la neumología, la cardiología, la neurología o la psicología, entre otras [28], [29].
- **Importancia de la salud mental.** Es de vital importancia evaluar y tratar los problemas de salud mental asociados al SPC-CP para mejorar la calidad de vida del paciente y que no deriven en un empeoramiento [28], [29].
- **Vacunación.** En varios estudios se menciona que la vacunación contra el COVID-19 puede tener un impacto positivo en la mejora de los síntomas en algunos pacientes con SPC-CP [27], [29].
- **Seguimiento a largo plazo.** Realizar un seguimiento continuo a largo plazo para evaluar la evolución de los síntomas y las secuelas, pudiendo así ajustar el tratamiento [28], [29].

Estas recomendaciones están sujetas a modificaciones según se vaya obteniendo más evidencia científica sobre el SPC-CP. Por eso mismo, es vital proporcionar información y recursos educativos a los pacientes y familiares de estos pacientes de forma continuada sobre cómo manejar los síntomas. Como en cualquier enfermedad de carácter crónico, otra indicación básica es promover el autocuidado, incluyendo seguir una alimentación saludable, realizar alguna actividad física adecuada y dormir un número mínimo de horas [29].

## 2.2. Estado de la técnica en el procesado de señales biomédicas dentro del contexto del síndrome post-COVID

A la hora de seleccionar las señales biomédicas con las que se trabajan en este trabajo se deben tener en cuenta los síntomas más frecuentes ya mencionados: respiratorios y cardiovasculares. Esta es la principal razón por la que se decide trabajar con la HRV, la SpO<sub>2</sub> y la EDR, aunque también ha influido la sencillez de registro tanto del ECG como de la pulsioximetría. En el caso de haber incorporado, por ejemplo, el flujo aéreo o el EEG, esto hubiera dificultado los registros en las consultas y, además, son señales más incómodas y más propensas a artefactos y pérdida de señal.

### **2.2.1. Señal de variabilidad de la frecuencia cardíaca en pacientes con síndrome post-COVID**

La HRV hace referencia a las fluctuaciones en los intervalos entre latidos cardíacos consecutivos. Es un indicador de la actividad autónoma del sistema nervioso, el cual puede estar influenciado por el estado de salud [30]. Algunos estudios han investigado la relación entre la HRV y el SPC-CP [11], [12], sugiriendo que la HRV puede estar alterada en los pacientes que sufren este síndrome posterior a la infección, en concreto se observa una disminución de la HRV en pacientes con síntomas prolongados en comparación con aquellos que se han recuperado por completo de la infección [11], [12].

Las complicaciones cardiovasculares, como la isquemia miocárdica, el infarto, la miocarditis y las arritmias cardíacas, son secuelas ya comentadas de la infección por COVID-19, con diferentes mecanismos fisiopatológicos sugeridos que implican daño directo al sistema circulatorio debido a la unión de los virus a los receptores de la ACE2 y la inflamación sistémica. Sin embargo, las consecuencias de la infección por COVID-19 en la regulación autonómica del corazón aún no están claras [31].

El sistema nervioso autónomo (SNA) es crucial en la regulación del ritmo cardíaco. La HRV evalúa las funciones simpática y parasimpática en el sistema cardiovascular, reflejando el equilibrio simpático-vagal y la disautonomía. La disautonomía se caracteriza por fallos en el SNA, manifestándose en síntomas como fatiga, hipotensión postural, cambios en la presión arterial, arritmias y deterioro de la función de la vejiga y de los intestinos. La disautonomía puede ser consecuencia de infecciones virales, incluyendo el COVID-19, donde el virus puede afectar al sistema nervioso a través de invasión directa, propagación sináptica o por la sangre. Además, se considera que la neumonía por COVID-19 puede causar daño inmunológico, vascular e hipoxia, contribuyendo a las manifestaciones neurológicas de la enfermedad [30], [31].

Como se ha mencionado anteriormente, los resultados de múltiples estudios han demostrado que los pacientes con SPC-CP muestran una reducción en la HRV (desviación estándar de los intervalos RR (SDNN), raíz cuadrada del error cuadrático medio (RMSSD), potencia de baja frecuencia (LF), potencia de alta frecuencia (HF) o relación entre estas dos últimas (LF/HF)) en comparación con aquellos que no presentan síntomas ni secuelas después de la infección [32].

### **2.2.2. Señal de saturación de oxígeno en sangre en pacientes con síndrome post-COVID**

La pulsioximetría es un método no invasivo para medir los niveles de oxígeno en sangre. Actualmente, no existe evidencia de ninguna investigación científica que relacione de forma directa la señal de oximetría y el SPC-CP. Por el contrario, sí hay referencia de datos obtenidos gracias a esta señal y pacientes con COVID-19 o con SPC-CP.

La oximetría es utilizada como una herramienta para poder evaluar la gravedad de la enfermedad de los pacientes que la presentan. Muchos pacientes durante la infección presentaban niveles bajos de saturación de oxígeno en sangre, lo que indica dificultades para obtener suficiente oxígeno. Por lo tanto, la oximetría es una medida de gran utilidad a la hora de evaluar a los pacientes con COVID-19, ya que permite identificar aquellos que tienen una

disminución de los niveles de oxígeno en sangre significativa, pudiendo necesitar administración de oxígeno suplementario o ingresar en la UCI [33].

La pulsioximetría fue una herramienta fundamental para el control de los pacientes con COVID-19, principalmente debido a la frecuencia en la que aparecía hipoxemia sin sintomatología que la manifestase. Esta técnica también se ha propuesto para ayudar a detectar si es necesario hospitalizar a los pacientes diagnosticados de COVID-19 [34].

### **2.2.3. Señal de respiración derivada del ECG en pacientes con síndrome post-COVID: cuantificación del acoplamiento cardiorrespiratorio**

El término de CRC se refiere a la interacción entre el corazón y los pulmones durante el bombeo de la sangre y la respiración [35]. Teniendo en cuenta que el COVID-19 y el SPC-CP puede afectar a múltiples sistemas del cuerpo, incluidos el cardiovascular y el respiratorio, algunas personas experimentan síntomas cardiorrespiratorios persistentes, como fatiga, dificultad para respirar o taquicardia [25]. Estos síntomas podrían ser estudiados a partir de la relación o influencia entre una señal de caracterización de la actividad cardíaca, típicamente la señal de HRV, y una señal de caracterización de la respiración, típicamente el flujo o los movimientos respiratorios o, alternativamente la señal EDR.

Este acoplamiento puede ser estudiado junto con la HRV, ya que, durante la inspiración, la frecuencia cardíaca suele acelerarse, mientras que, durante la espiración, suele suceder lo contrario, tiende a disminuir. Estos cambios en la frecuencia cardíaca son una manifestación del acoplamiento entre la respiración y el sistema cardiovascular que se conoce como arritmia sinusal respiratoria (*respiratory sinus arrhythmia*, RSA) [36].

Por todo lo comentado anteriormente, se puede concluir que el análisis del CRC puede proporcionar información relevante sobre la salud del paciente. Una mayor magnitud de este acoplamiento está relacionada con un mejor estado de salud, mientras que una menor presencia de RSA sugiere una salud más deteriorada [35]. Su cuantificación es muy utilizada en investigaciones científicas para valorar, entre otras cosas, los trastornos respiratorios y cardíacos [36]. Este es el motivo por el cual, aunque todavía no hay evidencia científica que relacione directamente la magnitud del CRC con el SPC-CP, se ha decidido incluir su análisis en el presente trabajo.

# CAPÍTULO 3. HIPÓTESIS Y OBJETIVOS

## 3.1. Hipótesis

La investigación desarrollada en el presente TFG se sustenta sobre varias hipótesis. En primer lugar, desde un punto de vista clínico, entre los síntomas más frecuentes del denominado SPC-CP están los respiratorios (disnea) y cardiovasculares (arritmias). Con respecto a los síntomas respiratorios, se espera que los pacientes con SPC-CP presenten un patrón alterado de la respiración en comparación a los pacientes que no muestren la persistencia de los síntomas (No SPC-CP). En este sentido, la señal de oximetría proporciona una medida indirecta y no intrusiva de la actividad respiratoria del paciente, con lo que podría ser útil para estudiar el efecto del SPC-CP sobre la respiración. Asimismo, se espera que los pacientes con SPC-CP presenten cambios relevantes en su modulación autonómica, con lo que el análisis de la señal HRV podría aportar información de interés. Desde un punto de vista técnico, los índices cardiorrespiratorios tradicionales podrían ser útiles para identificar y caracterizar estas alteraciones. Sin embargo, dado que los sistemas respiratorio y cardíaco son sistemas dinámicos no estacionarios, se espera que abordar la caracterización de estos pacientes desde una perspectiva no lineal podría aportar información adicional útil para detectar las potenciales alteraciones de los patrones respiratorios y modulación autonómica asociados al SPC-CP. Por último, ya que entre los síntomas más frecuentes se encuentran los respiratorios y los cardiovasculares, se espera una potencial alteración del CRC y, por tanto, de las medidas de cuantificación de su magnitud.

## 3.2. Objetivos

De acuerdo con estas hipótesis, el **objetivo principal** de este Trabajo de Fin de Grado consiste en analizar y caracterizar los patrones cardiorrespiratorios de pacientes con y sin diagnóstico de SPC-CP para identificar los índices autonómicos, oximétricos y de acoplamiento con mayor nivel de asociación con la persistencia de los síntomas y poder diferenciar a estos pacientes de otros sujetos sin diagnóstico de SPC-CP. Para lograr este objetivo general, se han establecido los siguientes **objetivos específicos**:

- Estudiar y describir los antecedentes, variables clínicas y afectación de la calidad de vida de pacientes con y sin diagnóstico de SPC-CP.

- Caracterizar los patrones respiratorios de pacientes con y sin SPC-CP a través del análisis automático de la señal de oximetría en los dominios temporal, espectral y no lineal.
- Caracterizar la modulación autonómica cardíaca de pacientes con y sin SPC-CP a través del procesado automático de la señal HRV en los dominios temporal, espectral y no lineal.
- Caracterizar el CRC de pacientes con y sin SPC-CP mediante medidas de coordinación y sincronismo en los dominios temporal, espectral y no lineal.
- Analizar las potenciales diferencias entre grupos de pacientes en función del padecimiento de COVID-19 y de la temporalidad de los síntomas post-COVID: (i) pacientes que no han sufrido COVID-19, (ii) pacientes con diagnóstico previo de COVID que actualmente no presentan sintomatología, (iii) pacientes con diagnóstico previo de COVID diagnosticados actualmente de SPC-CP que mantienen la persistencia de los síntomas durante más de 12 semanas.
- Identificar la combinación óptima de variables clínicas e índices derivados del procesado de las señales biomédicas bajo estudio que mejor discrimina a los pacientes con SPC-CP.
- Implementar y validar un modelo predictivo de *screening* automático de pacientes con SPC-CP en base a sus patrones respiratorios, modulación autonómica y CRC.



## CAPÍTULO 4. SUJETOS Y SEÑALES

En este capítulo se describen la población de estudio y las variables que componen la base de datos generada para esta investigación. Igualmente, se presentan y describen los distintos tipos de señales que se han analizado. En primer lugar, se exponen los criterios de inclusión y exclusión de la población bajo estudio y se presenta la estructura de la base de datos generada, indicando los datos extraídos de la historia clínica electrónica y los índices derivados de los registros biomédicos que se han utilizado. A continuación, se describen las señales y el protocolo de registro de las mismas a los diferentes sujetos incluidos en el estudio.

### 4.1. Diseño del estudio. Criterios de inclusión y exclusión.

El estudio llevado a cabo se caracteriza por ser prospectivo, longitudinal y observacional. La población de estudio estará compuesta por pacientes con diagnóstico previo de COVID-19 que acudieron de forma consecutiva a las consultas de seguimiento del SPC-CP por haber presentado una infección por COVID-19 en el Servicio de Neumología del Hospital Universitario Río Hortega de Valladolid, entre febrero y abril de 2023, de acuerdo a los siguientes criterios de inclusión:

- Diagnóstico previo de COVID-19, tanto por la prueba de reacción en cadena de la polimerasa (PCR) como por test de antígenos, con al menos 3 semanas de persistencia de los síntomas desde la fecha de confirmación del diagnóstico.
- Ambos sexos.
- Edad comprendida entre los 18 y los 80 años.

Se consideró SPC-CP a todos los pacientes con diagnóstico previo de COVID-19 y una persistencia de los síntomas superior a 12 semanas.

Además, se crearán dos grupos de control pareados en sexo y edad. Estos sujetos serán reclutados entre los pacientes que, de forma general, acuden a las consultas del Servicio de Neumología en el mismo periodo de tiempo. Por lo tanto, finalmente se crearán tres grupos de pacientes bajo estudio, definidos de la siguiente forma:

- Grupo 0: pacientes sin diagnóstico previo de infección por COVID-19 (grupo de control puro).
- Grupo 1: pacientes con diagnóstico previo confirmado de COVID-19 que actualmente no presentan sintomatología (grupo de control COVID).
- Grupo 2: pacientes con diagnóstico clínico confirmado de SPC-CP.

Como criterios de exclusión, aplicables en los tres grupos de estudio, se han fijado los siguientes:

- No presentar patologías cardiacas o respiratorias, como enfermedad pulmonar obstructiva crónica (EPOC) o insuficiencia respiratoria, entre otras, anteriores a la infección.
- No tomar medicación que afecten al sistema nervioso autónomo, como betabloqueantes, beta2-adrenérgicos y anticolinérgicos, entre otros.

Se solicitará la firma de consentimiento informado a cada paciente, informándoles de los detalles de la investigación a través de una hoja de información al paciente. El protocolo del estudio fue aprobado por el Comité de Ética de la Investigación con Medicamentos (CEIm) del Hospital Universitario Río Hortega de Valladolid (Referencia 23-PI012).

## 4.2. Protocolo de adquisición de datos

Las condiciones generales en que se realizaron todos los registros contaban con un entorno controlado, en la consulta de Neumología. El paciente se encontraba sentado, en reposo. Primeramente, se tomó la tensión al paciente para después proceder a registrar de forma simultánea el ECG y la pulsioximetría. La duración de las pruebas fue de unos siete minutos, para poder obtener al menos cinco minutos de registro limpio tras el preprocesado.

### 4.2.1. Variables clínicas

Entre las variables recogidas en la base de datos se encuentran los datos extraídos de la historia clínica electrónica de cada paciente. A continuación, se especifican las características recogidas:

**Variables socio-demográficas:** edad, sexo, duración de los síntomas, altura, peso, IMC, tabaquismo, tratamientos previos relacionados con enfermedades respiratorias y/o cardiacas, tensión arterial sistólica (TAS), tensión arterial diastólica (TAD).

**Comorbilidades:** presencia o ausencia de hipertensión arterial (HTA), diabetes mellitus (DM) y apnea del sueño (AOS).

**Relacionadas con el COVID-19:** fecha de diagnóstico, hospitalización, duración de la hospitalización, unidad de vigilancia intensiva (UVI), duración de la UVI, número de vacunas, fecha de vacunación, número de vacunas de cada tipo administradas (Pfizer, AstraZeneca, Janssen, Moderna).

**Tests de evaluación del paciente:** la escala modificada del *Medical Research Council* (mMRC) y el *COPD Assessment Test* (CAT).

- Escala mMRC: permite al paciente graduar cuantitativamente su propia disnea de forma visual y sencilla. Esta escala divide la gravedad en cinco ítems, en orden ascendente de

gravedad: grado 0, ausencia de disnea excepto al realizar ejercicio intenso; grado 1, disnea al andar deprisa en llano, o al andar subiendo una pendiente poco pronunciada; grado 2, la disnea le produce una incapacidad de mantener el paso de otras personas de la misma edad caminando en llano, o tener que parar a descansar al andar en llano al propio paso; grado 3, la disnea hace que tenga que parar a descansar al andar unos 100 metros o después de pocos minutos de andar en llano; y grado 4, la disnea impide al paciente salir de casa o aparece con actividades como vestirse o desvestirse [37].

- Test CAT: evalúa el impacto de la enfermedad pulmonar obstructiva crónica (EPOC) y la calidad de vida. El CAT valora mediante ocho ítems cuatro síntomas relativos a la disnea: la tos, la expectoración, las limitaciones para actividades de la vida diaria y las limitaciones sobre el sueño. Las puntuaciones de cada apartado se gradúan entre 0 (menor gravedad) y 5 puntos (mayor gravedad). La suma total deriva en una puntuación que puede ir desde 0 (mejor calidad de vida) hasta 40 puntos (peor calidad de vida). Esta puntuación se puede dividir dependiendo del impacto que genere al paciente: impacto bajo (1-10 puntos), impacto medio (11-20 puntos), impacto alto (21-30 puntos), impacto muy alto (31-40 puntos) [38].

#### 4.2.2. Señal de variabilidad de la frecuencia cardiaca

La señal HRV es una señal derivada que se extrae típicamente del ECG. La señal de ECG es un registro gráfico de la actividad eléctrica del corazón, obtenida mediante electrodos colocados sobre la piel del paciente. Esta actividad eléctrica es producida cuando el corazón se contrae y se relaja, siendo medible a través de pequeñas variaciones en el voltaje generadas en la superficie del cuerpo. El ECG es una herramienta diagnóstica de gran utilidad para evaluar el estado de salud del corazón y detectar posibles problemas cardíacos, siendo los más comunes a detectar infartos, arritmias o bloqueos. Además, también se utiliza para evaluar la eficacia de los tratamientos cardíacos y para monitorizar a los pacientes en situaciones críticas [39].

El dispositivo utilizado para realizar los registros de ECG fue el Biosignalsplux (Plux Wireless Biosignals S.A., Lisboa, Portugal) con el sensor específico de tres derivaciones para realizar ECG a una frecuencia de muestreo de 1 KHz, dispositivo el cual se muestra en la Figura 1. Los electrodos fueron colocados en las dos clavículas y en la cadera izquierda del paciente (montaje de derivaciones bipolares de extremidades). La señal de ECG aporta la señal HRV, se obtuvo mediante el módulo *heart rate variability* del software Opensignals (r)evolution proporcionado junto con el equipo de registro.

La HRV hace referencia a la variación existente en los intervalos de tiempo producidos entre los latidos consecutivos del corazón, es decir, la variación en la serie temporal del intervalo R-R de un ECG. La señal HRV se obtiene a partir del registro de los intervalos R-R, que representan el tiempo transcurrido entre dos ondas R consecutivas en el ECG [40]. Esta señal es una herramienta útil para evaluar la actividad del sistema nervioso autónomo, el cual controla la frecuencia cardíaca y la variabilidad. La variabilidad se modula principalmente a través de las dos ramas del sistema nervioso autónomo: el sistema nervioso simpático (aumenta la frecuencia cardíaca) y el sistema nervioso parasimpático (disminuye la frecuencia cardiaca) [40], [41].



**Figura 1.** Dispositivo Biosignalsplux (Plux Wireless Biosignals S.A., Lisboa, Portugal).

La señal de HRV es utilizada en numerosas aplicaciones médicas, como por ejemplo en el diagnóstico de enfermedades cardiovasculares, la evaluación del estado de salud del paciente, el análisis del estrés y la ansiedad, y la predicción de posibles eventos cardiovasculares [41].

#### **4.2.3. Señal de saturación de oxígeno en sangre**

La  $SpO_2$  indica la cantidad de oxihemoglobina que posee el paciente, es decir, la cantidad de oxígeno en sangre que es transportada hacia los tejidos. Esta medida se expresa en porcentaje, es decir, es relativa ya que lo que realmente se está midiendo es la oxihemoglobina en comparación a la hemoglobina sin unir (desoxihemoglobina) presente en la sangre [42], [43]. Un sujeto normal presenta una saturación en torno a un valor del 96%, mientras que los sujetos con enfermedades respiratorias pueden presentar desde valores bajos mantenidos de saturación media y basal (relacionados con una hipoxemia crónica), hasta desaturaciones transitorias que dependerán del tipo de patología [44]. En los registros realizados, se ha definido como desaturación a una caída de la  $SpO_2$  del 2%, 3% o 4%, teniendo en cuenta una duración mínima de 10 segundos.

Para medir la  $SpO_2$  de forma no invasiva se utiliza un pulsioxímetro. Este dispositivo se coloca en el dedo del paciente y mide longitudes de onda de luz para determinar la proporción de los niveles actuales de hemoglobina oxigenada [43]. Es posible determinar la saturación de oxígeno en la sangre gracias a que el color de esta cambia dependiendo de cuánto oxígeno esté unido a la hemoglobina. Este cambio de color se debe a las propiedades de las moléculas de hemoglobina y se puede cuantificar midiendo la cantidad de luz roja que pasa a través de la sangre. Cuando la hemoglobina pierde oxígeno, la sangre se vuelve más azulada y permite pasar menos luz roja. Los oxímetros utilizan este cambio de color para medir la saturación de oxígeno en la sangre a través de la espectrofotometría, expresándolo en términos de saturación [42].

Los registros de  $SpO_2$  registrados fueron adquiridos mediante el pulsioxímetro de muñeca NONIN WristOx2 3150 (Nonin Medical Inc., MN, USA), dispositivo el cual se muestra en la Figura 2. La frecuencia de muestreo utilizada para todos los sujetos fue de 1 Hz.



**Figura 2.** Pulsioxímetro de muñeca NONIN WristOx2 3150 (Nonin Medical Inc., MN, USA).

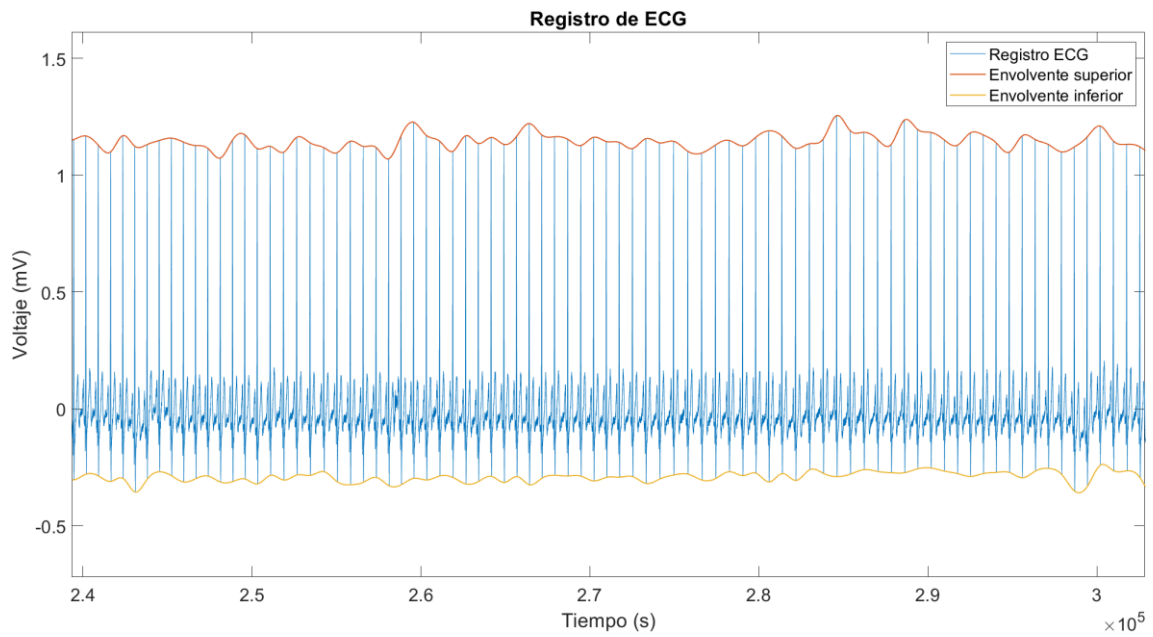
#### **4.2.4. Señal de respiración derivada del ECG: cuantificación del acoplamiento cardiorrespiratorio**

Como se ha desarrollado anteriormente, el término de CRC se refiere a la interacción entre el corazón y los pulmones durante el bombeo de la sangre y la respiración [35]. Debido a que algunos pacientes con SPC-CP presentan síntomas cardiorrespiratorios persistentes, como fatiga, dificultad para respirar o taquicardia [25], se considera de utilidad analizar el acoplamiento existente entre la señal HRV y la señal EDR.

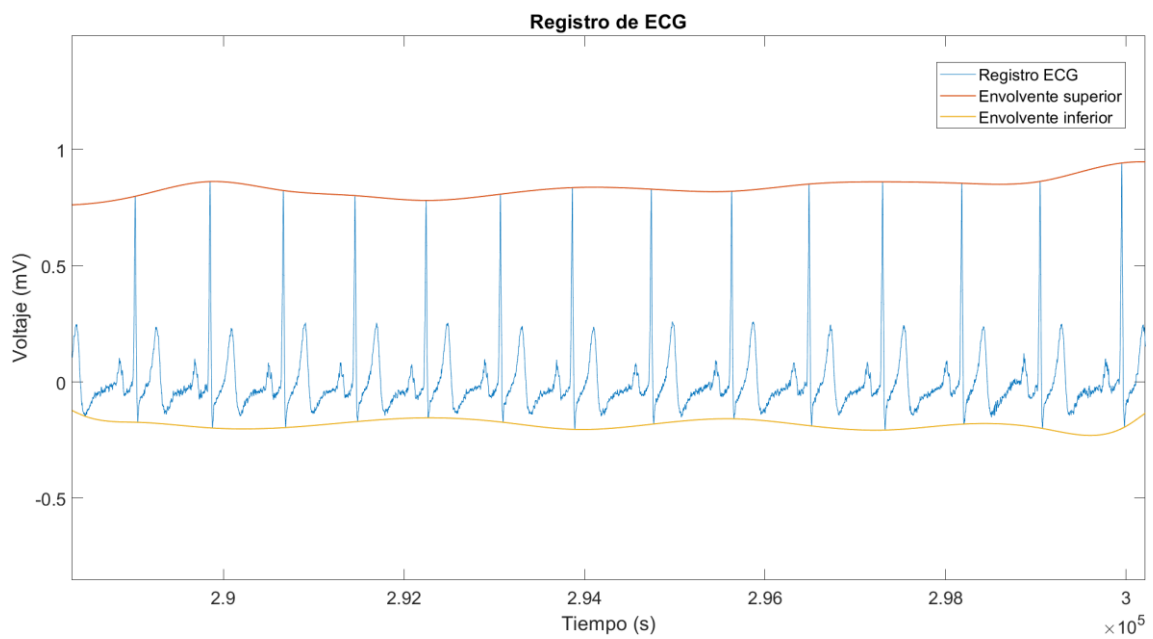
La señal HRV utilizada será la ya preprocesada explicada en el apartado anterior. Tanto esta como la de EDR tendrán una longitud de cinco minutos, eliminándose el primer minuto y las muestras restantes hasta el final del registro.

En el caso de la señal EDR, esta será extraída de los diferentes registros de ECG. La respiración se puede extraer del ECG porque esta produce cambios en la amplitud y en la frecuencia del complejo QRS. Es decir, que la señal de respiración modula al ECG. Detectando y cuantificando esta modulación, será posible obtener la señal de respiración a partir del ECG. En este TFG se han cuantificado los cambios en la amplitud de la onda Q, interpolando para obtener la envolvente superior con la misma frecuencia de muestreo que el ECG original. Además, se ha seguido un procedimiento análogo para obtener la envolvente inferior, que también está modulada por la respiración. En la Figura 3 y en la Figura 4 se muestran dos ejemplos de ECG con las envolventes superiores e inferiores. Ambas envolventes han sido analizadas por separado, para estudiar cuál de ellas ofrecía una mejor estimación de la respiración y cuantificación del CRC.

Antes de proceder al análisis y a la extracción de variables es necesario que ambos registros a comparar tengan la misma longitud. Como la frecuencia de muestreo no es la misma, siendo la señal EDR de 1000 muestras por segundo (la misma que la del ECG original) y la de la HRV muy inferior y no homogénea, será necesario remuestrear ambas señales. El valor escogido como frecuencia de muestreo para ambas señales fue de 25 Hz, tal y como recomiendan las guías actualmente para la señal respiratoria más utilizada, la de flujo aéreo [45].



**Figura 3.** Ejemplo 1 de un registro de ECG ampliado marcando su envoltente superior en naranja y su envoltente inferior en amarillo.



**Figura 4.** Ejemplo 2 de un registro de ECG ampliado marcando su envoltente superior en naranja y su envoltente inferior en amarillo.

# CAPÍTULO 5. METODOLOGÍA

En este capítulo se presentan los métodos, técnicas y algoritmos empleados durante el desarrollo del presente estudio. En primer lugar, se describen los métodos utilizados para la extracción de variables de las tres señales biomédicas bajo estudio, teniendo en cuenta métodos complementarios en el dominio temporal, espectral y no lineal. Una vez extraídas las características, se detalla el procedimiento implementado para la selección de variables, basado en GAs, y, finalmente, el algoritmo de clasificación multiclase empleado para identificar los 3 grupos de pacientes bajo estudio (control, COVID y SPC-CP). Por último, se presentan los test estadísticos utilizados durante el análisis descriptivo de las variables, así como las métricas utilizadas para la comprobación del rendimiento diagnóstico del clasificador.

## 5.1. Limpieza de datos y preprocesado de señales biomédicas

### 5.1.1. Detección de outliers

Una vez se completó la base de datos se analizaron los diferentes pacientes para buscar si las características de alguno dispersaban mucho de las del resto de sujetos de su mismo grupo. Para ello, se aplicó un método de detección de *outliers* multivariante basado en el cómputo de la distancia robusta de Mahalanobis, descartando un porcentaje predefinido de observaciones (pacientes) en función de su distancia con respecto al centro de datos. Este porcentaje fue definido mediante una sensibilidad del 5%.

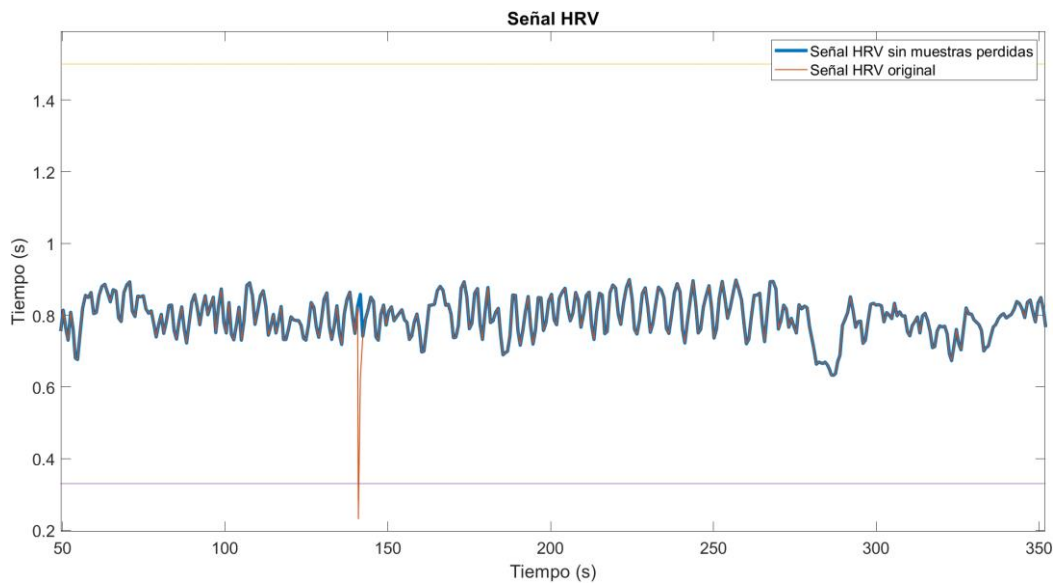
### 5.1.2. Preprocesado de la señal de variabilidad de frecuencia cardiaca

A la hora de trabajar con la señal de HRV, se realizó un preprocesado en el que primeramente se acertaba el tiempo de registro para que durase exactamente cinco minutos (se eliminaba el primer minuto de registro y el tiempo sobrante al final). Posteriormente, se inspeccionó cada registro para eliminar las muestras fisiológicamente imposibles. Concretamente, se descartaron aquellas muestras que cumplían alguna de las siguientes condiciones:

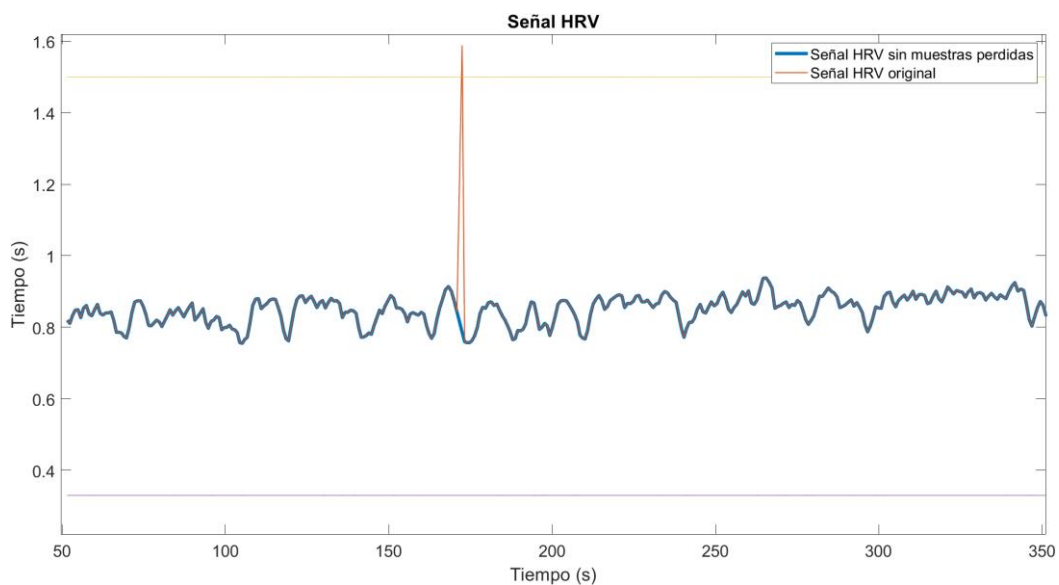
- Muestras cuyos valores, los cuales representan intervalos RR, sean inferiores a 0.33 segundos [46], ya que esto implica que se ha detectado un falso latido. La muestra será eliminada y el valor asignado a la muestra consecutiva (siempre y cuando sea válida) será la suma del intervalo eliminado más el valor de la propia muestra. En la Figura 5 se

observa un ejemplo de una muestra eliminada por tener un valor inferior a 0.33 segundos.

- Muestras cuyos valores sean superiores a 1.50 segundos [46], ya que en este caso esto implica que en el proceso de detección del complejo QRS se ha perdido un latido válido. Se añadirá una nueva muestra en el punto intermedio de las dos muestras válidas, cuyo valor será asignado mediante interpolación lineal. En la Figura 6 se observa un ejemplo de una muestra eliminada por tener un valor superior a 1.50 segundos.



**Figura 5.** Registro HRV preprocesado para eliminar las muestras con valores menores a 0.33.



**Figura 6.** Registro HRV preprocesado para eliminar las muestras con valores mayores a 1.50.



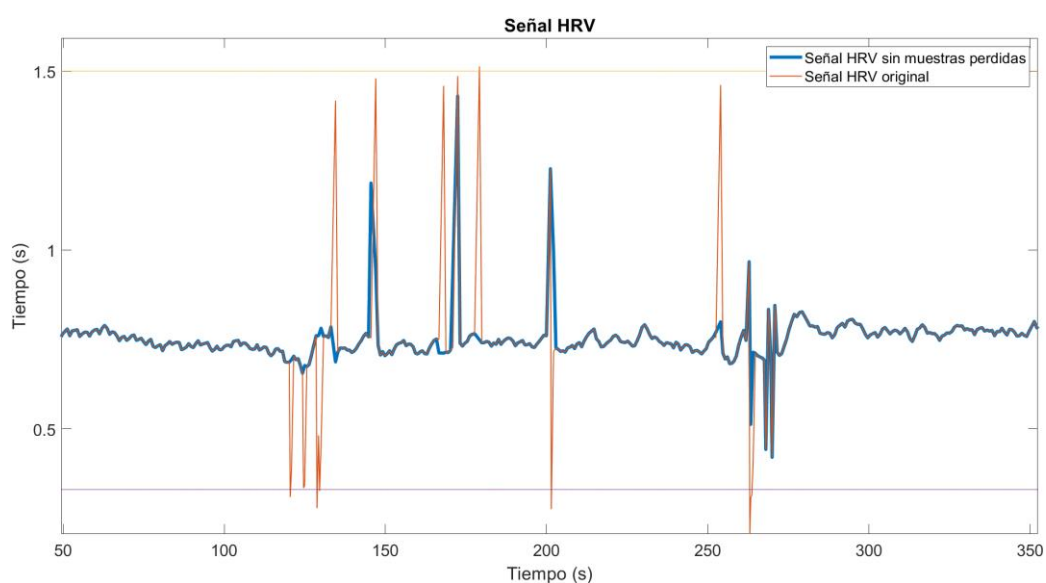
- Muestras cuyos valores tengan una diferencia mayor a 0.66 segundos con el valor del intervalo anterior [46]. En este caso se interpolarán los valores de las posiciones donde esta diferencia mayor de 0.66 esté presente.

Una vez implementadas las condiciones anteriores, se deberían obtener señales HRV libres de estas muestras fisiológicamente no válidas. Sin embargo, al comprobar los registros preprocesados se pudo observar que en algunos casos había valores muy próximos a los umbrales de 0.33 y al 1.50 definidos en la literatura, pero que al no alcanzar exactamente dicho umbral no eran eliminados. Esto fue comprobado revisando de forma simultánea la señal de HRV y el ECG original, pudiendo observar que se mantenían algunos latidos falsos o se saltaban algunos latidos válidos. La Figura 7 muestra un registro donde se pueden observar muestras con valores no fisiológicos que no han sido eliminados.

Para corregir este problema, se implementó una segunda fase de preprocesado adicional de la señal HRV, definiendo un nuevo protocolo basado en la desviación estándar del registro:

- Para las muestras cercanas a 0.33 segundos, se estableció que si el valor de la muestra a valorar  $HRV[i]$  menos 0.33 segundos era menor que la desviación estándar del registro, el valor no sería fisiológicamente válido. Para eliminar la muestra de forma natural (sin introducir falsos negativos), se siguió el mismo proceso establecido para eliminar las muestras inferiores a 0.33.
- Para las muestras cercanas a 1.50 segundos, se estableció que si 1.50 segundos menos el valor de la muestra a valorar  $HRV[i]$  era menor que la desviación estándar del registro, el valor no sería fisiológicamente válido. Para eliminar la muestra de forma natural se siguió el mismo proceso establecido para eliminar las muestras superiores a 1.50 segundos.

Una vez eliminadas todas las muestras que no se correspondían a intervalos RR válidos de la señal de HRV, los registros estarían listos para el procesado y la extracción de características.



**Figura 7.** Registro de HRV preprocesado con muestras fisiológicamente no válidas.

### 5.1.3. Preprocesado de la señal de saturación de oxígeno en sangre

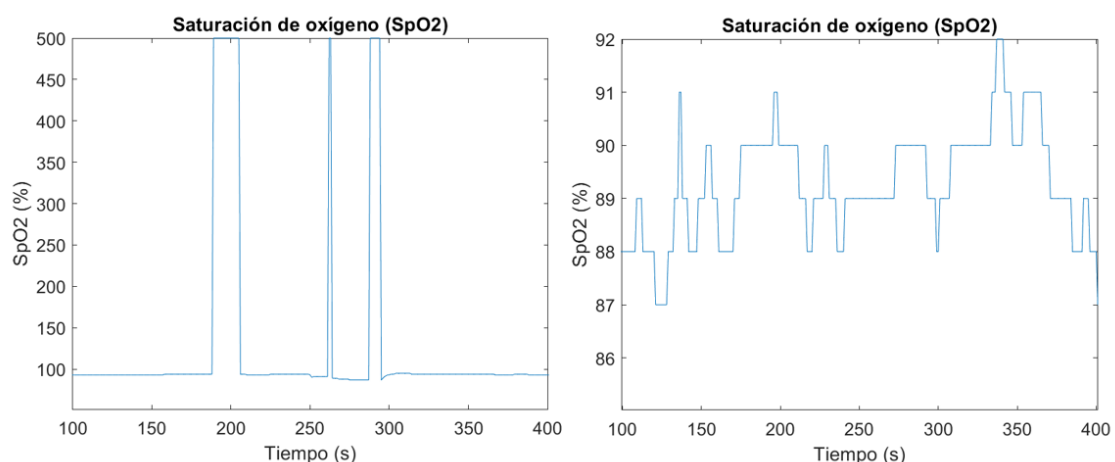
A la hora de trabajar con la señal de  $SpO_2$ , se realizó un preprocesado en el que primeramente se acortaba el tiempo de registro para que durase exactamente cinco minutos (se eliminaba el primer minuto de registro y el tiempo sobrante al final). Posteriormente, se eliminaron las muestras perdidas, es decir, los valores en el registro con un valor atípico de 500. En la Figura 8 se puede observar dos registros de oximetría, uno con artefactos caracterizados por el valor 500 y otro preprocesado sin la presencia de estos artefactos.

## 5.2. Estimación de la densidad espectral de potencial

El análisis espectral es comúnmente utilizado en el procesamiento de señales biomédicas. Se emplean diversas técnicas, siendo las más frecuentes aquellas basadas en la transformada Wavelet, la transformada discreta de Fourier (*Discrete Fourier Transform*, DFT) y sus variantes, como la transformada corta de Fourier (*Short Time Fourier Transform*, STFT) y el periodograma de Welch [47]. Estas técnicas permiten descomponer las señales en componentes frecuenciales, lo que facilita la identificación y el estudio de patrones periódicos relevantes en los datos biomédicos [48].

En el presente trabajo se utiliza la transformada de Welch para obtener una estimación de la densidad espectral de potencia a (*Power Spectral Density*, PSD) de la señal de  $SpO_2$ . Este método fue elegido debido a sus múltiples ventajas, como la reducción del sesgo y la varianza, la mayor resolución en frecuencia, la capacidad de supresión de ruido y la menor carga computacional [40], [41].

El método de Welch para obtener la PSD implica dividir la señal en segmentos o ventanas, calcular el periodograma modificado de cada ventana y luego promediarlos para obtener la PSD estimada final [49].



**Figura 8.** Preprocesado de una señal de oximetría. A la izquierda, un registro original con artefactos. A la derecha, un registro preprocesado sin artefactos.

El objetivo de dividir la señal en segmentos es conseguir reducir el sesgo y la varianza de la estimación de la PSD, proporcionando resultados más precisos. Además, mediante esta técnica es posible lograr una mayor resolución en frecuencia, pudiendo detectarse componentes espectrales más detallados en la señal. También al realizar el promedio de los periodogramas modificados se consigue disminuir el ruido y los artefactos, mejorando así la calidad de la estimación de la PSD especialmente en señales ruidosas o con interferencias [49]. Estas características hacen que la transformada de Welch sea una herramienta ampliamente utilizada para estimar la PSD en el análisis espectral de señales biomédicas [49].

A continuación, se detalla el algoritmo para obtener el estimador de la PSD utilizando el método de Welch [50]:

1. La señal temporal se divide en  $M$  segmentos solapados ( $overlap = D$ ) de longitud  $L$ , dando lugar a  $K$  segmentos que recorren toda la secuencia inicial. Esta secuencia temporal se muestra en la Figura 9.

$$\begin{aligned} X_1(j) &= X(j) & j &= 0, \dots, L-1 \\ X_2(j) &= X(j+D) & j &= 0, \dots, L-1 \\ X_K(j) &= X(j+(K-1) \cdot D) & j &= 0, \dots, L-1 \end{aligned} \quad (5.1)$$

2. Se selecciona un tipo de ventana de la misma longitud que los segmentos de señal ( $W(j), j = 0, \dots, L-1$ ) y se forman las secuencias  $(X_1(j)W(j), \dots, X_K(j)W(j))$  sobre las que se calcula la DFT:

$$A_K(n) = \frac{1}{2} \sum_{j=0}^{L-1} e^{-2K_i j n / L} \quad (5.2)$$

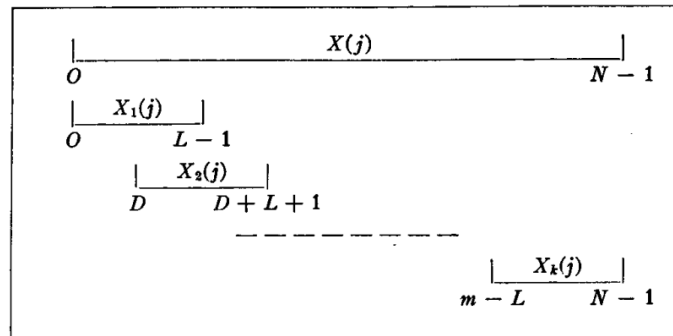


Figura 9. Secuencia temporal original segmentada [50].

3. Los  $K$  periodogramas modificados correspondientes a cada uno de los segmentos se obtienen mediante las siguientes ecuaciones:

$$I_k(f_n) = \frac{L}{U} |A_K(n)|^2, \quad k = 1, \dots, K \quad (5.3)$$

$$f_n = \frac{n}{L}, \quad n = 0, \dots, L/2 \quad (5.4)$$

$$U = \frac{1}{L} \sum_{k=1}^{L-1} W^2(j) \quad (5.5)$$

4. Finalmente, la PSD es calculada promediando los periodogramas:

$$\hat{P}(f_n) = \frac{1}{K} \sum_{k=1}^K I_k(f_n) \quad (5.6)$$

En este contexto, el estimador de Welch se considera asintóticamente consistente [50].

Una vez realizado este proceso, se procede a describir los parámetros espectrales escogidos para caracterizar el contenido espectral de una señal biomédica.

### **Potencia total**

La potencia total (PT) de la señal en todo el espectro es un parámetro descriptivo muy común en el ámbito del análisis espectral. Para su cálculo se debe hallar el área total bajo la curva de la PSD, es decir, se deben sumar los coeficientes de la PSD en todo el rango de frecuencias del espectro, es decir, hasta  $f_s/2$  [51].

$$P_T = \sum_{f=0 \text{ Hz}}^{f_s/2} PSD(f_j), \quad j = 1, 2, \dots, f_s/2 \quad (5.7)$$

siendo NFFT el número de puntos que constituyen la estimación de la PSD y  $f_j$  el valor de frecuencia correspondiente a cada uno de esos puntos.

### **Frecuencia mediana**

La frecuencia mediana (MF) se define como el componente espectral donde se alcanza el 50% de la PT de la señal [51], indicando el valor de este parámetro cómo se distribuye la potencia del espectro en la banda de interés a analizar [52], siendo en este caso todo el registro. Matemáticamente la MF se expresa de la siguiente forma [51]:

$$0.5 \sum_{f_j=0 \text{ Hz}}^{0.5 f_s} PSD(f_j) = \sum_{f_j=0 \text{ Hz}}^{MF} PSD(f_j) \quad (5.8)$$

siendo  $f_s$  la frecuencia de muestreo y  $f_j$  el valor de frecuencia correspondiente a cada uno de los puntos de la DFT.

### **Entropía espectral**

La entropía espectral (SE) mide el grado de irregularidad en el dominio de la frecuencia de una señal, es decir, es una medida de distribución de la potencia en el espectro, la cual se encuentra relacionada con la entropía de Shannon [53]. Valores pequeños de SE implican una concentración significativa de la potencia de la señal en determinadas frecuencias, es decir, una o varias componentes periódicas dominantes en el dominio del tiempo, mientras que valores elevados de SE implican una distribución de potencia más homogénea en la banda de interés y, por tanto, ausencia de periodicidades dominantes. Para calcular la SE es necesario normalizar primero la PSD, debiendo ser 1 la suma de sus componentes frecuenciales [52]. El valor normalizado de la PSD ( $p_j$ ) se calcula mediante la siguiente expresión [51]:

$$p_j = \frac{PSD(f_j)}{\sum_{f_j=0Hz}^{0.5f_s} PSD(f_j)} \quad (5.9)$$

Una vez realizada esta normalización, la SE se calcula mediante la siguiente ecuación [51]:

$$SE = - \sum_j p_j \ln(p_j) \quad (5.10)$$

de forma que el índice  $j$  recorrerá las frecuencias de la banda de interés de la señal.

## **5.3. Análisis no lineal**

Los métodos lineales no son capaces de capturar toda la información de las señales biológicas debido a que su comportamiento es no lineal y no estacionario [51]. Es por ello que se decide aplicar medidas no lineales de irregularidad, variabilidad y complejidad para poder obtener información adicional y complementaria de la dinámica de las señales a analizar.

### **5.3.1. Entropía muestral**

La entropía muestral (SampEn) es una medida no lineal de irregularidad en series temporales de datos, cuyos valores más elevados corresponden a secuencias más irregulares [54]. La familia de parámetros  $SampEn(m, r, N)$  se define como el logaritmo negativo de la probabilidad condicional de que dos secuencias que son similares (dentro del ancho de tolerancia  $r$ ) para  $m$  puntos contiguos sigan siendo similares cuando aumenta la longitud de la serie a  $m+1$  puntos [55]. Por lo tanto, antes de calcular la SampEn se deben fijar estos dos parámetros mencionados:

- La longitud  $m$ , la cual determina el tamaño de los vectores comparados.

- La ventana de tolerancia  $r$ , la cual suele ser normalizada utilizando la desviación estándar (*Standard Derivation, SD*) de la serie original.

Una vez definidos, se aplicaría la ecuación que calcula la SampEn de una serie de  $N$  muestras[51]:

$$SampEn(m, r, N) = -\ln \left[ \frac{A^m(r)}{B^m(r)} \right] \quad (5.11)$$

siendo  $A^m$  y  $B^m$  la media de  $(m)$ -segmentos y  $(m + 1)$ -segmentos de longitud  $X_m(i)$  ( $1 \leq i \leq N - m + 1$ ) respectivamente y siendo  $d[X_m(i), X_m(j)] \leq r$ . A continuación, también se definen matemáticamente los términos de esta ecuación [54]:

$$A^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} A_i^m(r) \quad (5.12)$$

$$B^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} B_i^m(r) \quad (5.13)$$

$$d[X_m(i), X_m(j)] = \max_{k=0, \dots, m-1} (|x(i+k) - x(j+k)|) \quad (5.14)$$

### 5.3.2. Medida de la tendencia central

La medida de la tendencia central (CTM) es un parámetro que mide el grado de variabilidad de una serie temporal, haciendo uso para su cálculo de diagramas de dispersión los cuales representan diferencias de segundo orden. Estas están formadas a partir del desplazamiento temporal de la señal original  $x(n)$  [55]:  $[x(i+2) - x(i+1)]$  vs.  $[x(i+1) - x(i)]$ .

La CTM se calcula seleccionando una región circular de radio  $r$ , alrededor del origen, contando el número de puntos que caen dentro del círculo y dividiendo por el número total de puntos. Esta normalización implica que el valor de CTM se encontrará siempre dentro del rango de 0 a 1. Se puede definir la CTM de forma matemática de la siguiente forma [55], [56]:

$$CTM = \frac{1}{N - 2} \sum_{i=1}^{N-2} \delta(d_i), \quad (5.15)$$

siendo

$$\delta(d_i) = \begin{cases} 1, & \text{si } [(x(i+2) - x(i+1))^2 + (x(i+1) - x(i))^2]^{1/2} < r \\ 0, & \text{en caso contrario} \end{cases} \quad (5.16)$$

### 5.3.3. Complejidad de Lempel-Ziv

La complejidad de Lempel – Ziv (LZC) es una medida no paramétrica de complejidad, cuyos valores más altos corresponden a series de datos de alta complejidad. Para calcular LZC, la señal original se convierte primero en una secuencia binaria comparando cada muestra con un umbral predefinido  $Td$ . Recorriendo la secuencia de símbolos desde su origen, cada vez que se encuentra una nueva subsecuencia se incrementa un contador de complejidad  $c(n)$ , siendo necesario normalizar este contador con respecto al número teórico máximo de subsecuencias que podrían obtenerse con esos símbolos [55], [57].

En el ámbito del análisis de señales biomédicas los símbolos utilizados suelen variar entre secuencias binarias o ternarias, es decir, entre 2 o 3 símbolos. La asignación de uno de estos dos valores se determina mediante el uso de umbrales. Si la secuencia resultante es binaria, se establece un umbral, mientras que, si es ternaria, se requieren dos umbrales [58].

Mediante las siguientes expresiones, se convierte la serie de datos en una secuencia de símbolos binaria  $P$  a través del umbral  $Td$ :

$$P = s(1), s(2), \dots, s(n) \quad (5.17)$$

siendo

$$s(i) = \begin{cases} 0, & \text{si } x(i) < Td \\ 1, & \text{en caso contrario} \end{cases} \quad (5.18)$$

Como se ha comentado anteriormente, será necesario examinar la secuencia  $P$  de izquierda a derecha aumentando el contador  $c(n)$  para cada subsecuencia de caracteres detectada. Este proceso es explicado a continuación [57], [58]:

1. Sean  $S$  y  $Q$  dos subsecuencias de  $P$ , y  $SQ$  la concatenación de ambas. Para expresar la concatenación  $SQ$  a excepción del último símbolo se utiliza  $SQ_\pi$ , siendo  $\pi$  el operador que señala la eliminación del último carácter de la secuencia.  $v(SQ_\pi)$  denota al conjunto de todas las subsecuencias de  $SQ_\pi$ . Las condiciones iniciales son:

$$c(n) = 1; \quad S = s(1); \quad Q = s(2)$$

siendo por tanto  $SQ_\pi = s(1)$ .

2. Si  $Q$  pertenece a  $v(SQ_\pi)$  no se incrementa  $c(n)$  ya que  $Q$  no es una subsecuencia nueva. Esto es debido a que  $S = s(1), s(2), \dots, s(r)$  y  $Q = s(r + 1)$ , por lo tanto  $SQ_\pi = s(1), s(2), \dots, s(r)$ .
3.  $Q$  ahora es  $s(r + 1), s(r + 2)$ , comprobándose otra vez si pertenece a  $SQ_\pi$ .
4. Se repiten los pasos 1, 2 y 3 hasta que  $Q$  no pertenezca a  $SQ_\pi$ . Por lo tanto, se incrementa en una unidad a  $c(n)$  cuando  $Q = s(r + 1), s(r + 2), \dots, s(r + i)$  no sea una subsecuencia de  $SQ_\pi = s(1), s(2), \dots, s(r + i - 1)$ .

5.  $S$  pasaría a ser  $S = s(1), s(2), \dots, s(r + i)$  y  $Q = s(r + i - 1)$ .

En el momento en que  $Q$  es el último carácter,  $c(n)$  reflejará la medida de la complejidad. El valor del parámetro  $c(n)$  depende de la longitud de la serie temporal de la que se obtiene, es por ello que hace falta realizar una normalización [55]:

$$LZC = \frac{c(n)}{b(n)}, \quad (5.19)$$

siendo  $b(n)$  el límite superior teórico de  $c(n)$  [55]:

$$b(n) = \frac{n}{\log_2(n)} \quad (5.20)$$

El valor de LZC estará comprendido por tanto entre 0 y 1, siendo aquellos valores cercanos a 1 los que representen una mayor complejidad en la serie temporal que se esté analizando, al contener un mayor número de subsecuencias [58].

## 5.4. Cuantificación del sincronismo entre señales biomédicas

A la hora de caracterizar la modulación de la frecuencia cardiaca a través de la respiración también se han tenido en cuenta enfoques de distinta naturaleza, utilizando medidas lineales en el dominio del tiempo, parámetros espectrales y métricas no lineales. De esta forma, la información obtenida será analizada desde varios puntos de vista proporcionando información de la forma más completa.

### 5.4.1. La secuencia de correlación

La secuencia de correlación cruzada (MCC) es una medida utilizada para evaluar la relación temporal entre dos señales. Se basa en aplicar la MCC entre segmentos de las señales, identificando así patrones de sincronización o desfase entre ellas [59].

La fórmula básica para calcular la MCC entre dos señales continuas,  $x(t)$  e  $y(t)$ , es la siguiente [59]:

$$MCC(t) = \int [x(\tau) \cdot y(t + \tau)] d\tau, \quad (5.21)$$

siendo:

- $MCC(t)$  el valor de correlación cruzada en el tiempo  $t$ .
- $x(t)$  y  $y(t)$  las señales continuas a comparar.
- $\tau$  el desfase entre las señales.



Valores de MCC cercanos a 1 son indicativos de una fuerte correlación entre las señales, mientras que valores cercanos a 0 indica una falta de correlación entre ellas [59]. En este TFG se ha utilizado como métrica del CRC el valor de la secuencia MCC en el origen, es decir, para un desfase nulo entre ambas señales (HRV y EDR).

#### 5.4.2. Magnitud de la coherencia espectral

La coherencia espectral (MSC) es una métrica utilizada para analizar la relación entre dos señales en el dominio de la frecuencia, evaluando su similitud. La MSC se calcula mediante el análisis de la PSD de cada una de las dos señales de interés. Este cálculo implica dividir la densidad espectral de la señal cruzada, conocido como cross-spectral density, entre la raíz cuadrada del producto de las densidades espectrales de las dos señales individuales. La señal cruzada se entiende como la transformada de Fourier de la autocorrelación cruzada de las dos señales que se están analizando [60]. La expresión de cálculo de la MSC se indica a continuación:

$$MSC_{ij}(f) = \left| \frac{S_{ij}(f)}{(S_{ii}(f) S_{jj}(f))^{1/2}} \right|, \quad (5.22)$$

$$S_{ij}(f) = \langle x_i(f) x_j^*(f) \rangle \quad (5.23)$$

siendo  $S_{ij}(f)$  el *cross-spectrum* donde \* significa conjugación compleja y  $\langle \rangle$  significa valor esperado.

El resultado de la MSC es un valor entre 0 y 1, indicando un valor cercano a 1 una alta similitud o sincronismo entre las señales en una frecuencia determinada, es decir, una fuerte coherencia entre ellas. Por el contrario, un valor cercano a 0 representa una baja relación entre las señales en esa frecuencia específica [60].

La MSC es especialmente útil en el análisis de señales biomédicas, destacando su uso en ECG, donde se puede investigar la sincronización entre diferentes regiones del sistema cardiovascular o con otro sistema, como puede ser el respiratorio [60], [61].

#### 5.4.3. Información mutua

La métrica de la información mutua (MI) evalúa la relación no lineal entre dos señales, siendo comúnmente utilizada en el ámbito del procesamiento de señales y el análisis de datos. Esta métrica cuantifica la cantidad de información que una señal proporciona sobre la otra, basándose en el hecho de que, si dos señales están fuertemente relacionadas, la información que se pueda extraer de una de las señales puede predecir la otra de manera más concisa [62]. El cálculo de la MI de dos señales discretas denominadas X e Y se basa en la siguiente expresión [62]:

$$MI(X, Y) = \sum_{ij} p(i, j) \log \frac{p(i, j)}{p_x(i)p_y(j)} \quad (5.24)$$

siendo:

- $MI(X, Y)$  la representación de la MI entre las señales X e Y.
- $p(i, j)$  la probabilidad conjunta de que las señales X e Y tomen los valores x e y, respectivamente.
- $p_x(i)$  y  $p_y(j)$  las probabilidades marginales de las señales X e Y, respectivamente.

El cálculo de la MI implica además estimar las distribuciones de probabilidad de las señales que se están analizando y calcular las probabilidades conjuntas y marginales correspondientes. Las técnicas de estimación de densidad de probabilidad pueden realizar estos cálculos (método del histograma o el método de *kernel*). El valor de la MI varía de 0 a un valor positivo, indicando el 0 una dependencia nula entre las señales y un valor mayor una mayor dependencia o relación no lineal [62].

## 5.5. Caracterización de la señal de variabilidad de la frecuencia cardiaca

A la hora de extraer las variables de la señal de HRV se han tenido en cuenta características de distinta naturaleza: parámetros temporales, parámetros espectrales y parámetros no lineales. De esta forma, la información obtenida gracias a esta señal será analizada desde varios puntos de vista, lo que se espera de lugar a características complementarias.

### 5.5.1. Análisis temporal de la señal HRV

A la hora de estudiar la secuencia de la HRV en el dominio del tiempo se hace uso de diferentes medidas estadísticas, como pueden ser la media o la desviación típica. Estos parámetros se aplican directamente sobre las muestras de la señal HRV sin interpolar [63]. En total se extrajeron tres parámetros:

- ***Average NN Interval (AVNN)***: media de la señal de HRV en el registro completo.
- ***Standard Deviation of NN Intervals (SDNN)***: desviación típica de la señal de HRV en el registro completo.
- ***Root Mean Square of the Differences of Successive NN Intervals (RMSSD)***: valor cuadrático medio de las diferencias entre valores adyacentes de la señal de HRV en el registro completo.

El intervalo NN hace referencia a la convergencia de los intervalos de tiempo entre latidos cardíacos sucesivos, también conocidos como intervalos RR.

### 5.5.2. Análisis espectral de la señal HRV

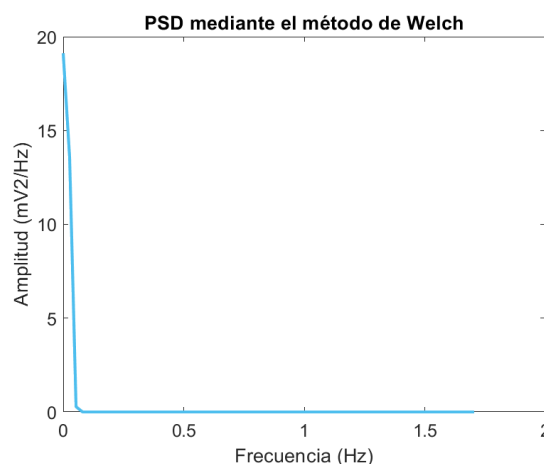
El análisis espectral es comúnmente utilizado en el procesamiento de señales biomédicas. Se emplean diversas técnicas que permiten descomponer las señales en componentes frecuenciales, lo que facilita la identificación y el estudio de patrones y características relevantes en los datos biomédicos [48]. En el presente trabajo se utiliza la transformada de Welch para obtener una estimación de la PSD de la señal de HRV. La técnica de Welch se considera un enfoque no paramétrico y utiliza la FFT como parte del proceso de cálculo [50].

Antes de proceder a extraer la PSD es necesario realizar una interpolación de la señal debido a que el procedimiento de obtención de la señal HRV tiene como resultado un muestreo no uniforme, es decir, no está equiespaciada en su eje de abscisas. Para conseguir un muestreo homogéneo se realizó una interpolación mediante *splines* cúbicos a una frecuencia de remuestreo de 3.41 Hz [46]. Este registro interpolado solo será utilizado a la hora de extraer las características espectrales de la señal de HRV, puesto que la transformada de Fourier en la que se suele basar el análisis espectral de esta señal asume un muestreo uniforme de la serie temporal.

Una vez que se tiene el registro interpolado es necesario definir los parámetros del método de Welch para la estimación de la PSD. Se utilizaron ventanas de 30 muestras y teniendo en cuenta que la frecuencia de muestreo es de 3.41 Hz, cada ventana equivaldría a 10.23 segundos de la señal HRV. El solapamiento utilizado fue del 50% y 128 puntos para obtener la DFT. En la Figura 10 se observa la PSD extraída de un registro de HRV.

Las características se aplican sobre bandas espectrales clásicas, las cuales son divididas de la siguiente manera [63]:

- **Banda de muy bajas frecuencias** (*Very Low Frequency, VLF*): comprendida entre 0.003 y 0.04 Hz. La interpretación fisiológica de esta región refleja la modulación vagal (actividad parasimpática) y el efecto del sistema renina angiotensina.
- **Banda de bajas frecuencias** (*Low Frequency, LF*): comprendida entre 0.04 y 0.15 Hz. Esta región de frecuencias guarda relación con el control simpático de la presión arterial.



**Figura 10.** PSD de un registro de HRV calculada mediante el método de Welch.

- **Banda de altas frecuencias (*High Frequency*, HF):** comprendida entre 0.15 y 0.4 Hz. Estas frecuencias reflejan el ritmo respiratorio. Se considera que la actividad en esta banda está relacionada con el control parasimpático de la frecuencia cardiaca.

Las características de la señal HRV en el dominio de la frecuencia se extraen en forma de medidas de potencia sobre estas bandas espectrales, siendo la potencia calculada como el área bajo la PSD sin normalizar. Las características extraídas son [63]:

- $P_{VLF}$ : potencia total en la banda de VLF.
- $P_{LF}$ : potencia total en la banda de LF.
- $P_{HF}$ : potencia total en la banda de HF
- $P_{LF/HF}$ : relación de la potencia absoluta en la banda de LF y la potencia absoluta en la banda HF, también conocido como balance simpático-vagal.

Las tres primeras variables fueron normalizadas, siendo la cuarta calculada una vez que la  $P_{LF}$  y la  $P_{HF}$  ya estaban normalizadas. Para esta normalización se valoró dividir las variables entre la PT o dividir las variables entre la suma de  $P_{LF}$  y  $P_{HF}$ , siendo elegida finalmente esta segunda normalización ya que presentaba diferencias más significativas a la hora del análisis estadístico.

### 5.5.3. Análisis no lineal de la señal HRV

Los métodos lineales no son capaces de capturar toda la información de las señales biológicas debido a que su comportamiento es no lineal y no estacionario [51]. Es por ello que se deciden aplicar medidas no lineales de irregularidad, variabilidad y complejidad para poder obtener información adicional y complementaria de la dinámica de la señal HRV.

Los métodos no lineales son aplicados sobre la señal temporal de HRV original una vez se han eliminado los artefactos. A continuación, se exponen las particularidades elegidas para la señal de HRV en cada una de las variables. Para ello se debe tener en cuenta que, aunque depende de la naturaleza de la condición bajo estudio y de las características del registro (duración y *setting*), se asume de forma general que mayor irregularidad o variabilidad en registros HRV de corta duración y en reposo estaría relacionada con un mejor estado de salud, equiparando salud con una mayor capacidad de adaptación a los cambios de la regulación autonómica cardiaca.

#### Entropía muestral

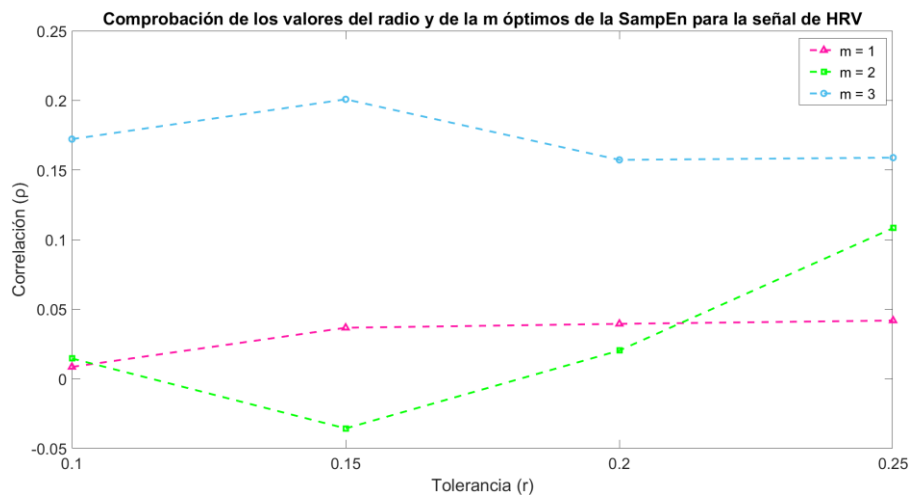
Será necesario establecer el valor de  $m$  y el valor de la tolerancia  $r$  para el cálculo de la SampEn para la señal de HRV. Para ello, se busca una correlación negativa entre los valores de esta variable y la variable clínica CAT escogida como referencia, ya que cuanto más variabilidad en esta señal y, por tanto, mejor estado de salud, mayores serán los valores de la SampEn, mientras que cuanto más salud (menor afectación a la calidad de vida), menores serán las puntuaciones en el cuestionario CAT. La correlación se comprobó sobre el 25% de la población bajo estudio.

En la Figura 11 se muestra la evolución de la correlación entre SampEn y CAT para los diferentes valores de  $m$  y  $r$  típicamente empleados en la literatura en el contexto del análisis de la señal HRV. Es posible observar que la correlación más negativa se obtiene con un valor de  $m$  igual a 2 y un valor de tolerancia de 0.15.

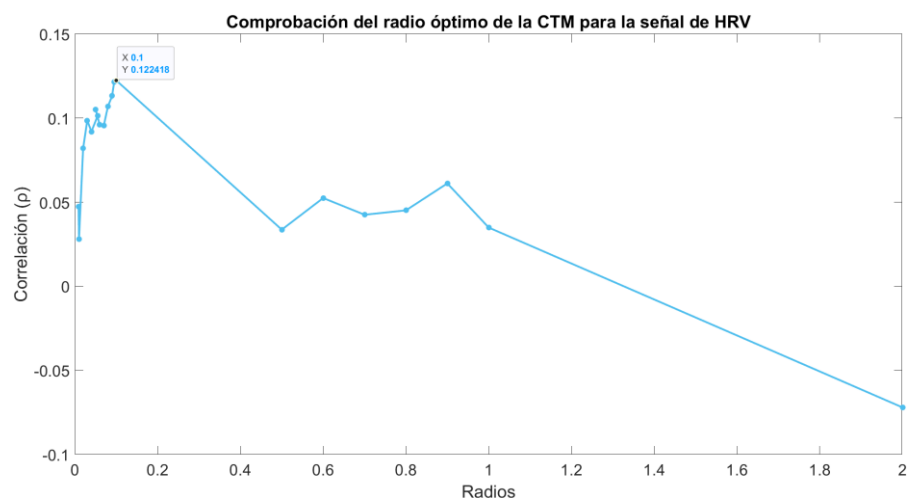
### Medida de la tendencia central

En el caso de la CTM será necesario establecer un radio óptimo para su cálculo. Para ello se busca una correlación positiva entre los valores de la CTM y la variable clínica CAT escogida con referencia, ya que cuanto más variabilidad (mejor estado de salud), menores serán los valores de la CTM y cuanto más salud (menor afectación a la calidad de vida), menores serán los valores de CAT. La correlación se comprobó sobre el 25% de la población bajo estudio.

La Figura 12 muestra la evolución de la correlación entre la medida de CTM y el cuestionario CAT para diferentes valores del radio. Como se puede observar en la gráfica, la correlación más positiva se obtiene para un valor de radio de 0.122418.



**Figura 11.** Comprobación de los valores óptimos de  $m$  y  $r$  para la SampEn para la señal de HRV.



**Figura 12.** Comprobación del valor óptimo del radio para la CTM para la señal de HRV.

### **Complejidad de Lempel – Ziv**

Par el cálculo de LZC fue necesario establecer el número de símbolos y la estrategia para definir el umbral. Para el caso de la señal de HRV, se decidió emplear 2 símbolos y basar el umbral de decisión en la mediana de la serie temporal, debido a su capacidad para minimizar la influencia de los valores atípicos comunes en los registros biomédicos.

## **5.6. Caracterización de la señal de saturación de oxígeno en sangre**

A la hora de extraer las variables de la señal de SpO<sub>2</sub> se han tenido en cuenta características de distinta naturaleza, encontrando entre ellas índices clásicos, parámetros espectrales y parámetros no lineales. De esta forma, la información obtenida gracias a esta señal será analizada desde enfoques complementarios, lo que proporcionará a priori una caracterización más completa.

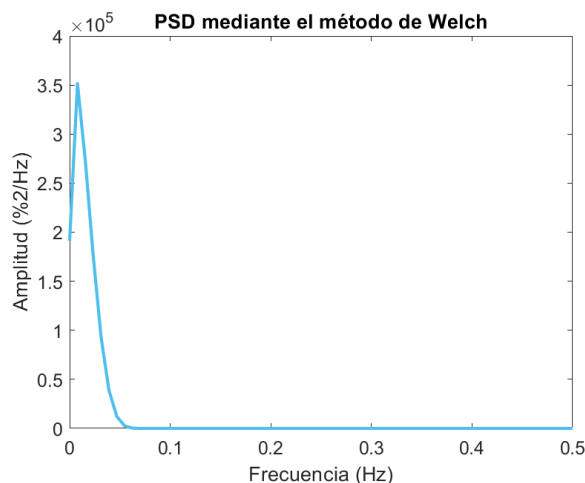
### **5.6.1. Índices de oximetría clásicos**

Los índices de oximetría convencionales típicamente utilizados en la práctica clínica fueron proporcionados por el software de análisis del dispositivo de registro (nVision versión 6.5.1.2, Nonin Medical Inc., MN, USA), los cuales se detallan a continuación:

- Valor medio, valor basal y valor mínimo de la señal de SpO<sub>2</sub>.
- Porcentaje de tiempo acumulado con valores por debajo de 95% (*Cumulative Time below 95%*, CT95) y de 90% (CT90).
- Índice de eventos de desaturaciones de oxígeno (*Oxygen Desaturation Index*, ODI) del 2% (ODI2), 3% (ODI3) y 4% (ODI4). Estas desaturaciones vienen determinadas por el dispositivo como una caída de la SpO<sub>2</sub> prefijada del 2%, 3% o 4% respecto a un umbral de saturación basal, teniendo en cuenta una duración mínima de 10 segundos.

### **5.6.2. Estimación de la densidad espectral de potencial**

Para la obtención de la PSD y de las características espectrales de la señal de oximetría, en este trabajo se utilizaron ventanas de 30 muestras. Teniendo en cuenta que la frecuencia de muestreo es de 1 Hz, cada ventana equivaldría a 30 segundos de la señal de SpO<sub>2</sub>. El solapamiento utilizado fue del 50% y se emplearon 128 puntos para obtener la DFT. En la Figura 13 se puede observar la PSD calculada en un registro de oximetría.



**Figura 13.** PSD de un registro de la SpO<sub>2</sub> calculada mediante el método de Welch.

### 5.6.3. Análisis no lineal

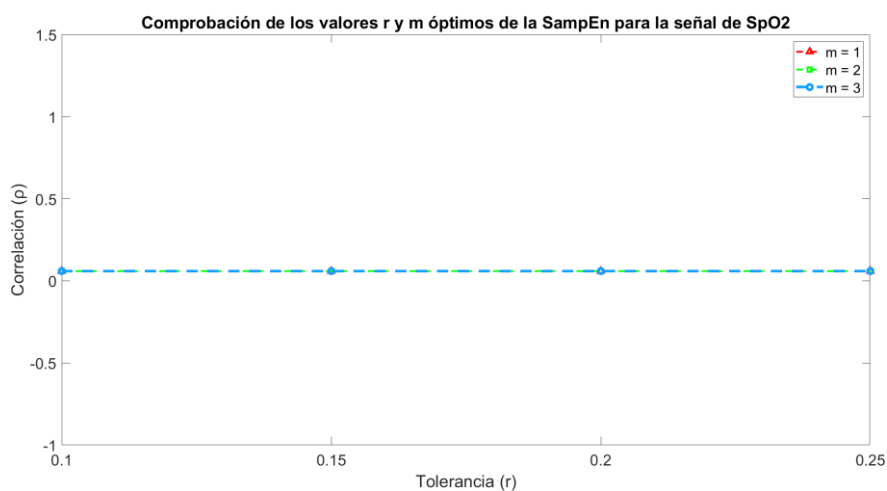
#### Entropía muestral

Es fundamental tener en cuenta que  $SampEn(m,r,N)$  es un conjunto de parámetros, cuyas comparaciones deben llevarse a cabo con valores fijos de  $m$ ,  $r$  y  $N$ . El intervalo de valores utilizable en la SampEn para la señal de SpO<sub>2</sub> que consigue una buena reproductibilidad estadística son  $m$  igual a 1 y a 2, y  $r$  entre 0.1 y 0.25 veces la SD de la serie temporal [64].

En el caso de la señal de SpO<sub>2</sub>, un estado de salud estaría relacionado con mayor estabilidad en los valores del registro, mientras que cambios acusados (descensos continuados, desaturaciones y resaturaciones) implicarían un peor estado de salud desde un punto de vista respiratorio.

Para establecer los valores óptimos de los parámetros  $m$  y  $r$  de cálculo de la SampEn para un registro de oximetría, se trató de optimizar la correlación entre las medidas de entropía y una variable calidad de vida CAT como variable representativa del estado de los pacientes con y sin SPC-CP (valores bajos en pacientes con mayor calidad de vida, i.e., con menor sintomatología; valores elevados en sujetos con peor calidad de vida, i.e., con mayor persistencia de síntomas). Concretamente, se buscó una correlación positiva con los valores aportados por la variable CAT, ya que cuanto más estabilidad (salud), menores serán los valores de la SampEn, y cuanto más salud, menores serán los valores de CAT. La correlación se comprobó sobre el 25% de la población bajo estudio.

En la Figura 14 se muestra los valores óptimos de  $m$  y  $r$  para la SampEn. En este caso la correlación es la misma para todas las combinaciones de valores, de forma que se seleccionaron de forma arbitraria un valor de 1 para el parámetro  $m$  y un valor de 0.1 para  $r$ .

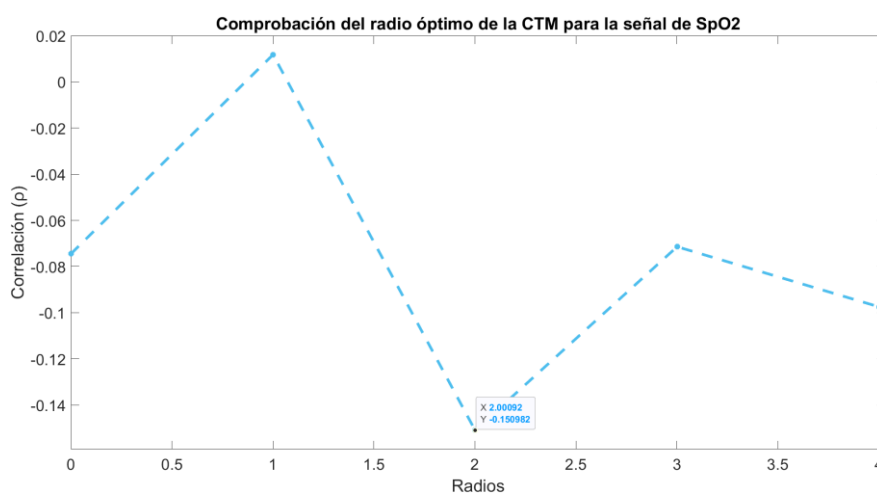


**Figura 14.** Comprobación de los valores óptimos de  $m$  y  $r$  para la SampEn para la señal de  $SpO_2$ .

### Medida de la tendencia central

Una mayor variabilidad en la serie de datos implicará que menos puntos se encuentren dentro de la región circular, aproximándose el valor de CTM al valor cero. En caso de que haya un mayor número de puntos dentro de la circunferencia se obtiene el caso contrario, habrá menos variabilidad en la señal y, por tanto, el valor del CTM será más cercano a 1. Como se ha explicado anteriormente, en el caso de la señal de  $SpO_2$ , la salud está relacionada con la estabilidad: cuanta más estabilidad en los valores de oximetría más salud. A la hora de seleccionar el valor óptimo del radio de cálculo de la CTM se vuelve a utilizar el test CAT para optimizar la correlación con el estado de salud de los pacientes. En este caso, para establecer el radio óptimo de la CTM se busca una correlación negativa, ya que cuanta más estabilidad, mayores serán los valores de la CTM y cuanta más salud, menores serán las puntuaciones del test CAT. La correlación se comprobó sobre el 25% de la población bajo estudio.

Como se puede comprobar en la Figura 15, el radio elegido tendrá un valor de 2.00092 ya que con él se alcanza la correlación más negativa.



**Figura 15.** Comprobación del valor óptimo de  $r$  para la CTM para la señal de  $SpO_2$ .



### **Complejidad de Lempel-Z**

Para este trabajo se decidió establecer 2 símbolos y el umbral de decisión se basó en la mediana de la serie temporal, debido a su capacidad para minimizar la influencia de los valores atípicos comunes en los registros biomédicos.

## **5.7. Caracterización del acoplamiento cardiorrespiratorio**

A la hora de caracterizar la modulación de la frecuencia cardíaca a través de la respiración también se han tenido en cuenta enfoques de distinta naturaleza, utilizando medidas lineales en el dominio del tiempo, parámetros espectrales y métricas no lineales. De esta forma, la información obtenida será analizada desde varios puntos de vista proporcionando información de la forma más completa.

Anteriormente han sido explicadas métricas de sincronismo en el dominio temporal, espectral y no lineal, métricas las cuales fueron aplicadas para la caracterización del CRC. A mayores, se aplicaron dos medidas comúnmente empleadas para cuantificar la magnitud de la RSA en el dominio del tiempo:

**Promedio de la diferencia entre el intervalo RR más la largo y el más corto para cada respiración.** Una de las principales manifestaciones de interacción/modulación del sistema respiratorio sobre el cardíaco es la RSA. La magnitud de RSA se puede calcular como el valor medio de la diferencia entre los dos intervalos R-R (el más largo y el más corto) dentro de una sola respiración, promediando finalmente con respecto a todos los ciclos respiratorios contenidos en el tiempo de registro [65].

**Promedio de las diferencias entre la frecuencia cardíaca y la frecuencia respiratoria para cada ciclo respiratorio.** Un enfoque frecuentemente utilizado para caracterizar la RSA se basa en la diferencia entre las frecuencias cardíaca y respiratoria en una sola respiración, promediando finalmente con respecto a todos los ciclos respiratorios incluidos en la ventana de registro. Para evitar el solapamiento de la relación entre las frecuencias cardíaca y respiratoria, se requiere una relación de al menos 2:1. Si no se cumple esta condición, el RSA debe calcularse sobre duraciones más largas en lugar de una sola respiración [65].

## **5.8. Selección de características. Algoritmos genéticos**

Una vez que se han extraídos todas las variables definidas en el protocolo del estudio (clínicas y derivadas del procesamiento de señal), cada paciente queda caracterizado a través de un amplio conjunto de variables. Computacionalmente, realizar una selección de variables es beneficioso para optimizar el modelo que se quiere entrenar y que en un primer momento se encuentra representado por un número de variables relativamente grande. Mediante esta selección se conseguirá por tanto reducir el coste computacional, reducir el tiempo de entrenamiento y facilitar la comprensión de los datos [66]. Esta reducción de la información de entrada al modelo

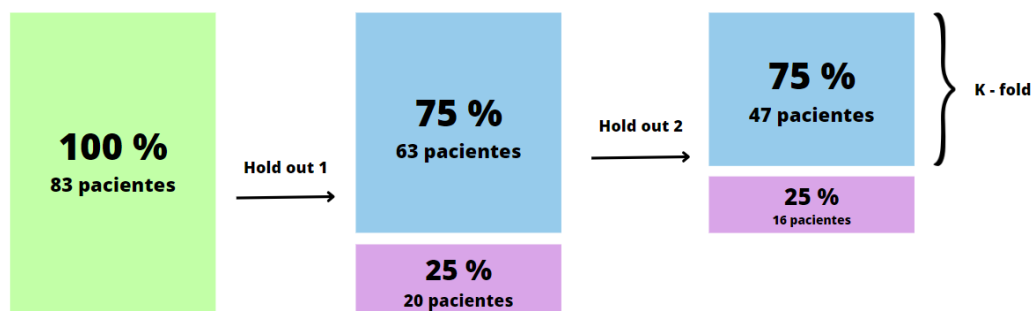
también presenta numerosas ventajas desde el punto de vista del rendimiento predictivo que van a conseguir que el modelo se entrene mejor [67], [68]:

- Eliminar información redundante proveniente de variables con una alta correlación entre ellas.
- Identificar características que se complementen proporcionando simultáneamente información más útil para la clasificación.
- Evitar el sobre-entrenamiento (*overfitting*).
- Mejorar la predicción final del clasificador.

La selección del subconjunto óptimo para el clasificador se puede realizar mediante múltiples métodos, pero en este trabajo se optó por utilizar una metodología directa (*wrapper methods*). De esta forma, el proceso de selección es guiado por una métrica de rendimiento predictivo (precisión, kappa, área bajo la curva ROC, entre otras) obtenida con el algoritmo de clasificación que se empleará en la posterior etapa de reconocimiento de patrones [68].

Para realizar este proceso de selección lo primero que se tuvo en cuenta fue el tamaño de la población bajo estudio y, dado que la base de datos obtenida no era lo suficientemente grande como para obtener subconjuntos individuales de entrenamiento, validación y test representativos, se plantearon unas particiones de esta población combinando estrategias *hold-out* y de validación cruzada para poder llevar a cabo todas las pruebas y optimizaciones requeridas. Primeramente, se divide a la población en un conjunto de entrenamiento con el 75% de los individuos y en un conjunto de test el 25% restante, realizando esta partición mediante una estrategia *hold-out*. Nos quedarían por tanto 63 pacientes en el conjunto de entrenamiento y 20 en el conjunto de test. Sobre este conjunto de entrenamiento se va a realizar una segunda partición idéntica a la anterior, obteniendo esta vez 47 pacientes en el conjunto de entrenamiento y 16 en el conjunto de test.

El 25% de los pacientes resultado del Hold-out 1 se reserva para realizar un test independiente del modelo de red neuronal final, mientras que el 75% restante se empleará para seleccionar las características óptimas. La segunda partición mediante Hold-out 2 se realiza sobre este 75% de la población inicial, siendo empleada para optimizar los hiperparámetros de cada red neuronal perceptrón multicapa (MLP) diseñada durante el cálculo de la función de ajuste a optimizar en el proceso de selección de variables. En la Figura 16 se representa como se fue dividiendo la población bajo estudio para obtener los subconjuntos explicados.



**Figura 16.** Esquema de la partición de la población bajo estudio en conjuntos de entrenamiento y test para la selección de características.

Una vez explicado cómo se va a ir dividiendo la población, se procede a explicar el método de selección de características, el cual se va a basar en un método de optimización estocástico perteneciente a la computación evolutiva: los GAs. Los GAs son enfoques de búsqueda heurística aplicables a un amplio rango de problemas de optimización. Un GA evalúa en paralelo un conjunto de soluciones potenciales que representan una solución al problema de optimización que se quiere resolver. Una solución es un candidato potencial dentro del problema de optimización. Su representación tiene un papel importante, ya que mediante ella se determina la elección de los operadores genéticos. Las representaciones suelen ser listas de valores y generalmente se basan en conjuntos de símbolos, pudiendo ser continuos (vectores) o secuencias de bits. En el caso de problemas combinatorios, las soluciones suelen consistir en símbolos que aparecen en forma de lista. Los operadores genéticos producen nuevas soluciones dentro del espacio de búsqueda y permiten avanzar en la inspección del espacio de soluciones. La representación de una solución de forma codificada, que está sujeta al proceso evolutivo, se suele denominar genotipo o cromosoma empleando la terminología evolutiva [15].

Como se ha comentado, los GAs trabajan sobre cadenas binarias, siendo estas cadenas de la misma longitud para todos los individuos de la población de soluciones potenciales. El tamaño de la población es fijo según se van recorriendo las generaciones, dependiendo ambos valores (tamaño de la población y número de generaciones) de la complejidad del problema bajo estudio (número de grados de libertad a optimizar). Se pueden definir las siguientes etapas, posteriormente representadas en forma de diagrama de flujo [15]:

1. Inicialización de la población de soluciones potenciales.
2. Evaluación del ajuste (*fitness*) de cada uno de los individuos que forman la población de soluciones potenciales.
3. Aplicar una estrategia de selección para elegir a los padres (*parents*).
4. Realizar operaciones genéticas de variación (cruce y mutación) para obtener la descendencia (*children*) entre los padres previamente seleccionados.
5. Crear la nueva población de soluciones potenciales.
6. Repetir las etapas 2 a 5 hasta que se cumpla alguna de las condiciones de parada.

En la etapa 5, es importante decidir qué estrategias de selección de padres y reemplazo de individuos van a ser utilizadas, además de fijar algunos parámetros [15], [69]:

- Tamaño de la población: es el número de soluciones potenciales, siendo en este trabajo un valor aleatorio entre 25 y 30.
- Población inicial: es el punto de partida del GA. Es conveniente inicializarlo de forma pseudo-aleatoria, forzando a que el número de variables iniciales sea pequeño, para explorar convenientemente todo el espacio de búsqueda y evitar soluciones finales tendientes al *overfitting*. En este TFG se parte de soluciones potenciales con 5 variables iniciales.
- Probabilidad de cruce (*Pc*): es utilizada para controlar la probabilidad de que se produzca una operación de cruce o recombinación durante la reproducción de los individuos en una población, es decir, determina la frecuencia con la que se utilizará el operador de

cruce en la reproducción de la población. En este trabajo,  $P_c$  toma un valor de 0.8, típicamente empleado en la literatura [70].

- Probabilidad de mutación ( $P_m$ ): es utilizada para controlar la probabilidad de que se produzca una mutación en un gen durante la reproducción de los individuos en una población. La mutación es un operador genético que introduce cambios aleatorios en la información genética de un individuo, permitiendo explorar nuevas soluciones mediante la introducción de variaciones aleatorias en los genes existentes que no podrían obtenerse únicamente con la operación de cruce. La  $P_m$  determina por tanto la frecuencia con la que se utilizará el operador de mutación durante la generación de nuevos individuos hijos. En este trabajo, la  $P_m$  toma un valor de 0.01, típicamente empleado en la literatura [70].
- Criterio de parada: en este trabajo se establece como criterio de finalización del proceso de optimización el alcanzar un número máximo de generaciones, típicamente 100 [70].

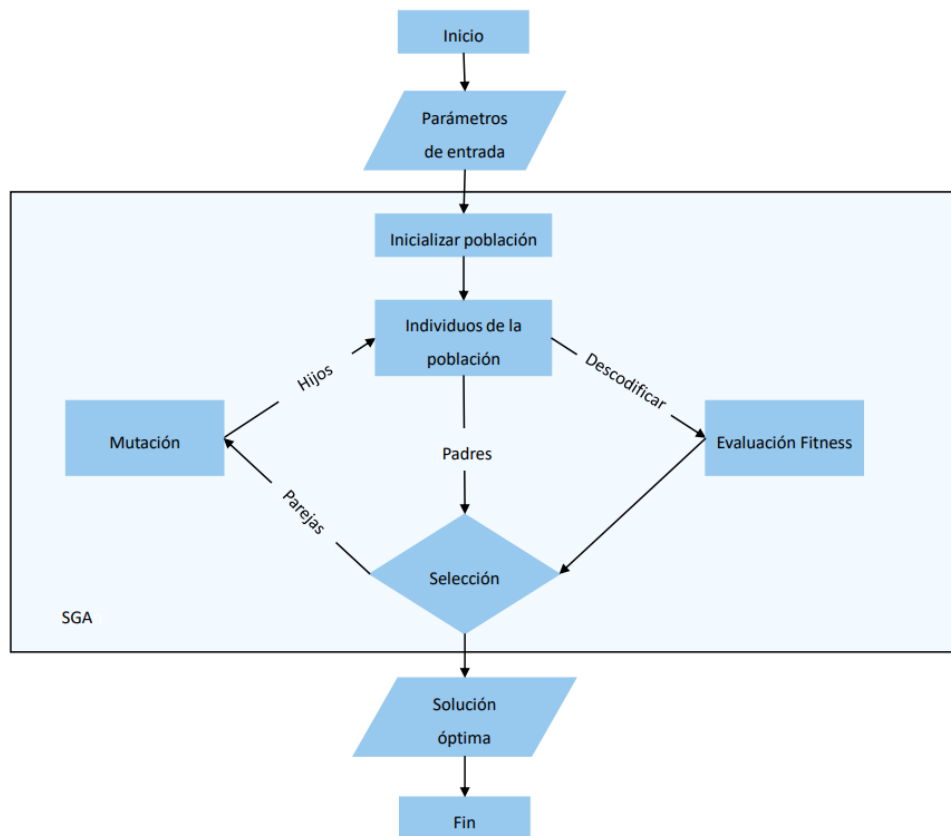
Para realizar la selección final, primeramente, será necesario generar una función de ajuste, también conocida como función *fitness*. Esta función será la encargada de evaluar el ajuste de los individuos de una población de soluciones potenciales, siendo independiente de la estructura general del GA (el diagrama de flujo mostrado en la Figura 17) y solo dependiendo del problema de optimización particular [15].

En este trabajo, la función *fitness* será la precisión (tasa global de aciertos: Acc) de un clasificador multiclase (3 clases) implementado mediante una red neuronal perceptrón multicapa, que será entrenada empleando el subconjunto de pacientes del último bloque de la Figura 15 (75% resultante del Hold-out 2: 47 pacientes). Este conjunto de entrenamiento será dividido utilizando la técnica de validación cruzada *10-fold* ( $K = 10$  folds) para optimizar los hiperparámetros de la red neuronal (número de nodos ( $N_h$ ) en la capa oculta y la tasa de aprendizaje ( $\alpha$ ) que posteriormente se van a utilizar para entrenar la red neuronal dentro de la función de cálculo del *fitness* utilizando el conjunto completo de entrenamiento de 47 pacientes, para posteriormente ser validado mediante el conjunto de test de 15 pacientes resultante del proceso de Hold-out 2. De este último paso se obtiene una precisión (Acc) para la solución potencial bajo estudio, siendo éste el valor de salida de la función de *fitness*.

En los procedimientos de optimización mediante GAs, también es común por parte del diseñador especificar el porcentaje de elitismo, ya que es conveniente solo reemplazar a los peores individuos manteniendo fijos los mejores. Hay que tener en cuenta que, para evitar una convergencia hacia un óptimo local en las etapas iniciales, este valor no puede ser muy elevado [15]. En este trabajo, el valor del elitismo se estableció en 1 individuo.

La salida del algoritmo de optimización será un vector de 0's y 1's de la misma longitud que la dimensión del espacio de búsqueda (número de variables de la base de datos de partida), donde los 1's representan qué variables forman parte del subconjunto óptimo final de entrada a la etapa de clasificación (red neuronal), mientras que los 0's indican que esa variable debe ser descartada.

En la Figura 17, se muestra mediante un diagrama de flujo el funcionamiento de un GA.



**Figura 17.** Diagrama de flujo de un algoritmo genético [69].

### 5.9. Clasificación de características. Red neuronal perceptrón multicapa

Una MLP es una red neuronal artificial en la que se construye un modelo inspirado en cómo el cerebro humano almacena y estructura la información mediante la interconexión de sus neuronas (nodos de la red). Entre sus componentes se encuentran estructuras de diferentes niveles o capas (entrada, capas ocultas y salida), encontrándose todas estas capas interconectadas a través de sus unidades más simples, llamadas perceptrones. Cada perceptrón presenta conexiones con perceptrones de otras capas, caracterizándose cada uno por una función de activación  $g_{act}()$ . Las conexiones mencionadas entre unidades están asociadas mediante pesos adaptativos ( $w_{ij}$ ), expresándose la salida del perceptrón mediante la siguiente ecuación [16]:

$$z_j = g_{act} \left( \sum_{i=1}^d w_{ij} \cdot x_i + b_j \right), \quad (5.25)$$

siendo:

- $x_i$  cada una de las componentes del vector de características de entrada.
- $d$  la dimensión del vector de entrada (número de características).
- $z_j$  el valor de activación.
- $b_j$  el sesgo.

La función de activación suele ser la función sigmoide. En este trabajo se ha empleado la función *softmax*, la cual se puede definir como una salida sigmoide normalizada, describiendo así una distribución de probabilidad [16].

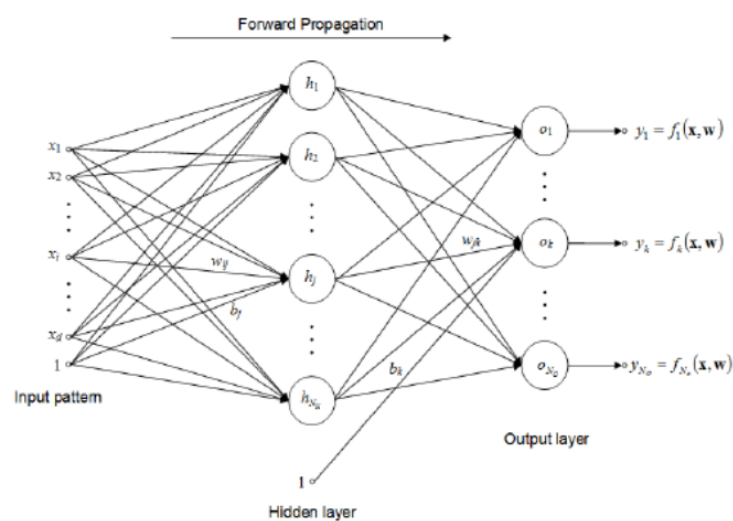
En una red MLP, los valores de las variables de entrada se introducen a las neuronas de la primera capa, para que ésta pueda calcular su valor de salida según la ecuación descrita anteriormente. A su vez, esta salida actúa como una nueva entrada para la siguiente capa, que una vez más vuelve a calcular su salida. En la Figura 18 se muestra un ejemplo de una MLP genérica.

El valor de salida de la red neuronal del ejemplo se puede definir tomando las salidas como función explícita de las entradas y los pesos [16]:

$$y_k = f_k(x, w) = g_l \left\{ \sum_{j=1}^{Nh} w_{jk} \cdot g_t \left( \sum_{i=1}^d w_{ij} \cdot x_i + b_j \right) + b_k \right\}, k = 1, \dots, No, \quad (5.26)$$

siendo:

- $d$  el número de características del vector de entrada.
- $Nh$  el número de nodos de la capa oculta.
- $w_{jk}$  los pesos que conectan la salida de los nodos de la capa oculta con el nodo de la capa de salida.
- $w_{ij}$  los pesos adaptativos que conectan la característica  $i$  del vector de entrada con el nodo  $h_j$  de la capa oculta.
- $b_j$  bias del nodo  $j$  de la capa oculta.
- $g_t()$  la función de activación de la capa oculta.
- $g_l()$  la función de activación de la capa de salida.
- $No$  el número de nodos de la capa de salida (número de clases a identificar en los enfoques multiclase).



**Figura 18.** Ejemplo de una MLP genérica  $Nh$  nodos en una única capa oculta y  $No$  salidas [16].

En este trabajo se han utilizado tres nodos en la capa de salida, una para cada uno de los tres grupos de pacientes definidos: (i) grupo de control sin diagnóstico previo de COVID; (ii) grupo de pacientes diagnosticados de COVID sin persistencia de los síntomas; (iii) grupo de pacientes con SPC-CP.

La red neuronal debe obtener los pesos óptimos de la red  $w_{ij}$  mediante un proceso de aprendizaje, inicializando aleatoriamente estos pesos primero y después ajustándolos para un conjunto de datos de entrenamiento definidos por pares  $(x^n, t^n)$  correspondientes al patrón de entrada y al valor de salida asociado. La función de densidad de probabilidad de estos datos se expresa como [16]:

$$p(x, t) = p(t|x) \cdot p(x), \quad (5.27)$$

siendo:

- $p(t|x)$  la densidad de probabilidad de  $t$  condicionado a  $x$ .
- $p(x)$  la densidad de probabilidad de  $x$ .

Por lo tanto, la red MLP pretende conseguir predecir el target o valor real  $t$  a partir de las características o variables del vector  $x$ , modelando para ello  $p(t|x)$  [16]. Para este trabajo se ha aplicado una red neuronal MLP para realizar una clasificación multiclase, en concreto, una clasificación ternaria, representando  $t$  el grupo al que pertenece: grupo 0, grupo 1 o grupo 2.

Tras el proceso de selección de variables se obtendrá un subconjunto del espacio de características original recogido en la base de datos del estudio, siendo ahora estas variables seleccionadas la entrada al clasificador. La función del clasificador será analizar estas características y asignar a cada paciente a una de las tres categorías definidas (0: control; 1: COVID; 2: SPC-CP). Como se ha indicado, para implementar esta clasificación se propone el modelo de red neuronal MLP debido a su capacidad para analizar problemas complejos. A continuación, se describe el procedimiento empleado para optimizar los hiperparámetros del modelo final de red neuronal.

En este TFG se ha utilizado una red MLP con una única capa oculta y la función de activación *softmax*. Los pesos han sido optimizados mediante el algoritmo *scaled conjugate gradient* (SCG) con un número máximo de 100 iteraciones. El modelo de red final ha sido entrenado empleando los valores óptimos de  $N_h$  y  $\alpha$  obtenidos con el conjunto de entrenamiento resultante del Hold-out 1 (segundo bloque representado en la Figura 16), aplicando sobre los 63 pacientes y una estrategia de *leave-one-out cross-validation*, es decir, validación cruzada dejando uno fuera para simular la métrica de validación de los hiperparámetros. Para minimizar la influencia de la inicialización aleatoria de la red neuronal, tratando de mantener un coste computacional adecuado, este proceso se repitió en 5 realizaciones distintas. Una vez obtenidos los hiperparámetros óptimos, la red MLP final se entrenó con los 63 pacientes que formaban el conjunto de entrenamiento de la primera partición Hold-out 1. Las métricas de rendimiento predictivo finales se calcularon utilizando el conjunto independiente de test de 20 pacientes resultante de la primera partición Hold-out 1.

## 5.10. Análisis estadístico

Una vez recopiladas las variables clínicas definidas en el protocolo y extraídas las características procedentes de las tres señales analizadas, es común realizar un análisis descriptivo de las mismas en términos de la mediana y rango intercuartil (IQR), así como realizar un análisis estadístico para poder establecer si hay diferencias significativas entre los grupos bajo estudio. Igualmente, se definen las métricas de rendimiento predictivo empleadas para evaluar cómo de adecuado fue el clasificador diseñado para implementar el *screening* de pacientes con SPC-CP. La implementación de las etapas de extracción de características de las señales biomédicas, selección de variables y clasificación de patrones mediante la red neuronal se ha realizado empleando Matlab® 2022a. Igualmente, el *toolbox* de estadística de este mismo software se empleó para realizar el análisis descriptivo de las variables y aplicar los test estadísticos.

### 5.10.1. Test no paramétrico de Kruskal Wallis

El test de Kruskal-Wallis (KW) es una prueba estadística no paramétrica, la cual determina si existen diferencias significativas entre las medianas de tres o más grupos independientes. Este test es utilizado cuando los datos no cumplen con los supuestos de normalidad y homogeneidad de varianza, que son comprobados mediante los test de Kolmogorov-Smirnov y Levene, respectivamente. La hipótesis nula del test de KW establece que las medianas de todos los grupos son iguales, al contrario que la hipótesis alternativa, la cual determina que al menos una de las medianas es diferente a las demás. El procedimiento para realizar un test de Kruskal-Wallis es el siguiente [71]:

1. Selección de los grupos: es necesario tener al menos tres grupos independientes. Cada uno de los grupos debe tener además una muestra independiente de observaciones.
2. Organización de los datos: todos los datos de los grupos tendrán que estar combinados y ordenados de menor a mayor.
3. Asignación de rangos: es necesario asignar rangos a los datos ordenados. En caso de que haya valores repetidos se asigna el promedio de los rangos correspondientes.
4. Suma de los rangos al cuadrado corregida: se calcula la suma de los rangos al cuadrado para cada grupo. Este valor se ajusta según el tamaño de la muestra y el número de grupos.
5. Cálculo del estadístico de prueba: el estadístico de prueba del test de KW está basado en la suma de los rangos al cuadrado corregida y sigue una distribución chi-cuadrado con  $k-1$  grados de libertad, siendo  $k$  el número de grupos.
6. Determinación del valor  $p$ : mediante la distribución chi-cuadrado se determina el valor  $p$  correspondiente al estadístico de prueba calculado. En caso de que el valor  $p$  sea menor que el nivel de significancia predefinido, el cual generalmente toma un valor de



0.05, se rechaza la hipótesis nula y, por lo tanto, se puede afirmar que hay diferencias significativas entre las medianas de los grupos.

El test de KW es una herramienta útil cuando se trabaja con datos no paramétricos y se desea comparar múltiples grupos, como es en el caso de este trabajo y por ello es utilizado para buscar diferencias estadísticas significativas entre los tres grupos para las variables bajo estudio. Tras aplicar el test de Kruskal-Wallis, en aquellas variables que arrojaron diferencias estadísticamente significativas, se realizó un análisis post-hoc en busca de las diferencias estadísticas entre los pares de grupos concretos bajo estudio (*pairwise analysis*), empleando para ello el test no paramétrico de Mann-Whitney. Para las variables categóricas, las diferencias estadísticas entre grupos se evaluaron empleado el test de Chi2. En este análisis se aplicó la corrección de Fisher (*Fisher's least significant difference*), con un umbral de significación de 0.05.

## 5.10.2. Medidas de la capacidad predictiva de un test diagnóstico

### 5.10.2.1. Parámetros para la evaluación de un clasificador multiclase

A la hora de expresar la utilidad de un clasificador se pueden utilizar diferentes métricas que permiten evaluar cómo de fiable es este test. Comúnmente, la gran mayoría de estas métricas se derivan de la denominada matriz de confusión. En el presente TFG se analizan 3 grupos de pacientes, con lo que se construirá una matriz de tres por tres. En esta matriz se comparan las predicciones realizadas por el clasificador automático bajo estudio con los targets reales (*gold standard*) para cada una de las categorías bajo estudio. A continuación, se exponen las métricas empleadas en el presente TFG.

#### Precisión (*Accuracy*)

La precisión (Acc) es una métrica comúnmente utilizada para evaluar el rendimiento de las redes neuronales en problemas de clasificación multiclase. El Acc representa la proporción de ejemplos clasificados correctamente con respecto al total de ejemplos evaluados, es decir, indica cómo de bien se está realizando la tarea de clasificación. Normalmente esta métrica es expresada como un valor entre 0 y 1, donde el 1 representa una precisión perfecta y el 0 una precisión nula [72].

$$Acc = \frac{\text{Número de instancias correctamente predichas}}{\text{Número total de instancias}} \quad (5.28)$$

Aunque este índice representa una métrica importante para evaluar el rendimiento de una red neuronal multiclase, es necesario tener en cuenta ciertas limitaciones. La principal limitación es que en el caso de que las clases se encuentren desbalanceadas en el conjunto de datos, un alto Acc no tiene por qué ser representativo del rendimiento real de la red neuronal [72]. Sin embargo, esta limitación no será relevante para este trabajo, dado que los tres grupos se encuentran bastante balanceados entre sí.

**Índice Kappa**

El índice Kappa ( $K$ ) es una métrica utilizada para evaluar la concordancia entre las categorías asignadas por dos o más observadores. En este TFG, los observadores serán el clasificador automático (red neuronal) y el especialista que emitió el diagnóstico de los pacientes en base a los procedimientos clínicos estándar. A diferencia de la precisión, el índice  $K$  es una medida más robusta para evaluar el rendimiento de un modelo predictivo, ya que no tiene en cuenta (descarta) el acuerdo que podría darse por simple azar [72], [73].

Este índice es utilizado especialmente en situaciones en las cuales las clases del conjunto de datos presentan un desequilibrio en su tamaño. Para calcularlo es necesario construir una matriz de confusión que muestre la cantidad de aciertos y la cantidad de errores producidos al clasificar cada clase. A partir de la matriz de confusión, se calcula el índice  $K$  mediante la siguiente ecuación [73]:

$$K = \frac{P_o - P_e}{1 - P_e}, \quad (5.29)$$

siendo:

- $P_o$  proporción de aciertos observados entre las etiquetas predichas y las etiquetas reales.
- $P_e$  proporción esperada de aciertos por azar.

Matemáticamente,  $P_o$  y  $P_e$  se describirían como [72], [73]:

$$P_o = \frac{n_{11} + n_{12} + n_{13}}{N}, \quad (5.30)$$

siendo:

- $n_{11}$ ,  $n_{22}$  y  $n_{33}$  los valores de la diagonal principal de la matriz de confusión, es decir, el número de aciertos totales de las tres clases bajo estudio.
- $N$  el número total de instancias en el conjunto de datos.

$$P_e = \frac{n_{1f} \cdot n_{1c} + n_{2f} \cdot n_{2c} + n_{3f} \cdot n_{3c}}{N^2}, \quad (5.31)$$

siendo:

- $n_{if}$  la suma de los valores que se encuentran en la fila  $i$  de la matriz de confusión (siendo  $i = 1, 2$  y  $3$ ), representando el número total de ejemplos de la clase correspondiente.
- $n_{ic}$  la suma de los valores que se encuentran en la columna  $i$  de la matriz de confusión (siendo  $i = 1, 2$  y  $3$ ), representando el número total de ejemplo predichos de la clase correspondiente.

El valor de  $K$  varía entre -1 y 1, indicando un valor de 1 una concordancia perfecta entre las etiquetas predichas y las reales, mientras que un valor de 0 indica un acuerdo realizado al azar.

En caso de que el valor fuese negativo significaría que la concordancia sería peor que si hubiese sido realizada al azar [72], [73].

### **Macro-F1 score**

El índice macro-F1 score (MF1) es una métrica utilizada en modelos de aprendizaje automático para evaluar el rendimiento a la hora de realizar una clasificación multiclase. El F1 score combina la precisión y la sensibilidad, proporcionando una medida equilibrada del rendimiento del modelo para todas las clases. Para calcular el MF1 primero se calcula el F1 score para cada clase individualmente, para después promediarlos y así poder obtener un valor global. La fórmula del F1 score se define como [72], [74]:

$$F1_i = 2 \cdot \frac{\text{sensibilidad} \cdot \text{valor predictivo positivo}}{\text{sensibilidad} + \text{valor predictivo positivo}}, \quad (5.32)$$

siendo  $i$  en nuestro caso la categorías 0, 1 y 2.

El MF1 se calcula mediante promedio no ponderado de los F1 score individuales, contribuyendo cada clase de igual forma al cálculo de la puntuación F1 global, sin tener en cuenta el tamaño de la clase:

$$MF1 = \frac{1}{3} \sum_{i=1}^3 F1_i \quad (5.33)$$

Es importante tener en cuenta que este índice es útil cuando todas las clases tienen la misma relevancia y se desea evaluar el rendimiento general del modelo sin tener en cuenta el desequilibrio de clases [74].

#### **5.10.2.2. Parámetros para la evaluación de un clasificador binario**

En los problemas multiclase, es común simular la simplificación a problemas binarios combinando las categorías hasta confirmar matrices de confusión de dos clases, de las cuales se derivarán sus propias métricas de rendimiento. En el presente TFG, las dos matrices de confusión construidas serán:

- Grupo 0 vs. (Grupo 1 + Grupo 2): “pacientes que no han pasado el COVID-19” frente a “pacientes que han pasado el COVID-19”.
- (Grupo 0 + Grupo 1) vs. Grupo 2: “pacientes que no presentan SPC-CP” frente a “pacientes que presentan SPC-CP”.

A continuación, se explican los parámetros que se pueden encontrar dentro de la matriz de confusión de dos clases, teniendo en cuenta que la nueva clase 0 será la categoría con etiqueta negativa y la clase 1 la categoría con etiqueta positiva [74]:

- **Verdaderos positivos (TP):** número de sujetos positivos que han sido clasificados de forma correcta como patológicos mediante el test bajo estudio.
- **Falsos negativos (FN):** número de sujetos positivos que han sido clasificados de forma incorrecta como no patológicos mediante el test bajo estudio.
- **Verdaderos negativos (TN):** número de sujetos negativos que han sido clasificados de forma correcta como no patológicos mediante el test bajo estudio.
- **Falsos positivos (FP):** número de sujetos negativos que han sido clasificados de forma incorrecta como patológicos mediante el test bajo estudio.

La distribución de estos índices en la matriz de confusión es la mostrada en la

Tabla 2 [74]. A partir de estos elementos encontrados en la matriz de confusión se pueden calcular las siguientes métricas [74]:

- **Sensibilidad (Se):** porcentaje de sujetos positivos que han sido correctamente clasificados por el test bajo estudio.

$$Se = \frac{TP}{TP + FN} \quad (5.34)$$

- **Especificidad (Sp):** porcentaje de sujetos negativos que han sido correctamente clasificados por el test bajo estudio.

$$Sp = \frac{TN}{TN + FP} \quad (5.35)$$

- **Valor predictivo positivo (PPV):** proporción de sujetos clasificados como positivos por el test bajo estudio que realmente son patológicos.

$$PPV = \frac{TP}{TP + FP} \quad (5.36)$$

**Tabla 2.** Distribución de una matriz de confusión de dos clases.

CLASE DE DATOS		Clases predichas	
		0	1
Clases reales	0	TN	FP
	1	FN	TP

TN: verdaderos negativos; FP: falsos positivos; FN: falsos negativos; TP: verdaderos positivos.

- **Valor predictivo negativo (NPV):** proporción de sujetos clasificados como negativos por el test bajo estudio que realmente son no patológicos.

$$NPV = \frac{TN}{TN + FN} \quad (5.37)$$

- **Razón de verosimilitud positiva (LR+):** ratio del porcentaje de sujetos positivos correctamente clasificados respecto al porcentaje de negativos erróneamente clasificados por el test bajo estudio.

$$LR^+ = \frac{Se}{1 - Sp} \quad (5.38)$$

- **Razón de verosimilitud negativa (LR-):** ratio del porcentaje de sujetos positivos erróneamente clasificados respecto al porcentaje de negativos correctamente clasificados por el test bajo estudio.

$$LR^- = \frac{1 - Se}{Sp} \quad (5.39)$$

- **Precisión (Acc):** porcentaje global de sujetos correctamente clasificados.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.40)$$



## CAPÍTULO 6. RESULTADOS

A la hora de realizar el reclutamiento de pacientes se tuvieron en cuenta criterios de inclusión y exclusión y una vez completada la base de datos, se realizó la técnica de detección de *outliers* descrita en el capítulo anterior para observar si era necesario descartar a alguno de los pacientes.

Los métodos descritos en el capítulo anterior han servido para extraer un conjunto de características de la señal de SpO<sub>2</sub>, de la señal de HRV y de la señal de EDR. Estas características están definidas en el dominio temporal, espectral y no lineal. Además de estas características, en la composición del modelo también se han utilizado variables clínicas. Se procede a explicar primeramente los resultados obtenidos buscando diferencias significativas entre las variables en los distintos grupos bajo estudio. En aquellas características en las que se observan diferencias estadísticamente significativas, se representa su distribución de datos correspondiente mediante diagramas de bloques cajas (*boxplots*) o de barras, dependiendo de la naturaleza de los datos (continuos o categóricos, respectivamente).

Como resultado del proceso de evaluación de la red MLP, se muestran los valores de las métricas obtenidas en el conjunto independiente de test, tanto para el enfoque multiclase (3 clases) como binario (2 clases). También se muestran las variables seleccionadas que finalmente componen el subconjunto óptimo de variables resultante del procedimiento de optimización mediante algoritmos genéticos.

### 6.1. Población bajo estudio

Un total de 105 pacientes cumplieron los criterios de inclusión y firmaron el consentimiento informado. De ellos doce fueron descartados debido a problemas relacionados con la adquisición y almacenamiento de los diferentes registros, dos fueron descartados por presentar patologías incompatibles con el estudio detectadas durante la revisión posterior de la historia clínica y otros dos debido a que los registros de ECG presentaban muchos artefactos, de tal forma que el tiempo de artefactos hizo que el tiempo de registro válido fuese inferior a los cinco minutos definidos en el protocolo. Tras la detección de *outliers* se descartaron 6 pacientes: 1 paciente sin COVID (grupo control puro), 1 paciente del grupo de control COVID y 4 pacientes del SPC-CP. En total, fueron analizados 83 registros (48 hombres y 35 mujeres).

A continuación, en la Figura 19 se muestra el diagrama de flujo descriptivo de la población de estudio final.

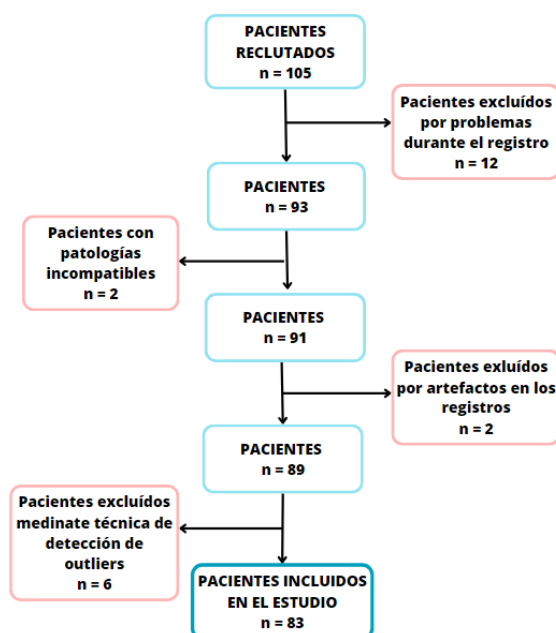


Figura 19. Diagrama de flujo de pacientes que forman parte de la población bajo estudio.

## 6.2. Análisis descriptivo de la población bajo estudio

La Tabla 3.1. y la Tabla 3.2. muestran los valores medianos y rango intercuartil de cada una de las variables clínicas recopiladas para cada paciente. Es importante destacar que no se observaron diferencias estadísticamente significativas entre los grupos bajo estudio en términos de la edad, sexo o IMC, lo que descarta la influencia de estas variables confusoras en las potenciales diferencias asociadas a otras variables. Se pueden observar diferencias significativas en cuanto a la duración de los síntomas, tabaquismo, tensión diastólica (TAD), test de disnea mMRC y test de calidad de vida CAT.

Tabla 3.1. Mediana, rango intercuartil y p - valor de las características clínicas para los tres grupos bajo estudio.

Variables	Todos los sujetos	Grupo 0	Grupo 1	Grupo 2	p – valor
Sujetos	83 (100.00%)	25 (30.12%)	25 (30.12%)	33 (39.76%)	-
Sexo (masculino)	48 (57.83%)	13 (15.66%)	17 (20.48%)	18 (21.69%)	0.4596
Edad (años)	54.00 [16.00]	53 [17.75]	52.00 [16.25]	56.00 [14.50]	0.2363
IMC	28.03 [6.81]	26.89 [6.32]	30.90 [5.50]	27.25 [5.21]	0.1564
Duración de los síntomas					
< 4 semanas	48 (57.83%)	25 (30.12%)	22 (26.51%)	1 (1.20%)	<< 0.01
4 - 12 semanas	4 (4.82%)	0 (0.00%)	3 (3.61%)	1 (1.20%)	
> 12 semanas	31 (37.35%)	0 (0.00%)	0 (0.00%)	31 (37.35%)	

Los datos se expresan en forma de mediana y rango intercuartil (IQR) para las variables continuas y en términos de número y porcentaje para las variables categóricas. IMC: Índice de Masa Corporal.



**Tabla 3.2.** Mediana, rango intercuartil y p - valor de las características clínicas para los tres grupos bajo estudio.

Variables	Todos los sujetos	Grupo 0	Grupo 1	Grupo 2	p – valor
<b>Tabaquismo</b>					
No fumador	37 (44.58%)	9 (10.84%)	14 (16.87%)	14 (16.87%)	<b>&lt;&lt; 0.01</b>
Fumador	13 (15.66%)	11 (13.25%)	2 (2.41%)	0 (0.00%)	
Exfumador	33 (39.76%)	5 (6.02%)	9 (10.84%)	19 (22.89%)	
<b>HTA</b>	28 (33.73%)	8 (9.64%)	8 (9.64%)	12 (14.46%)	0.9188
<b>DM</b>	13 (15.66%)	3 (3.61%)	6 (7.23%)	4 (4.82%)	0.3901
<b>AOS</b>	26 (31.33%)	11 (13.25%)	9 (10.84%)	6 (7.23%)	0.0921
<b>Hospitalización</b>	8 (9.64%)	0 (0.00%)	4 (4.82%)	4 (4.82%)	0.1312
<b>Hospitalización UVI</b>	3 (3.61%)	0 (0.00%)	0 (0.00%)	3 (3.61%)	0.0946
<b>TAS</b>	140.00 [19.75]	138.00 [15.25]	144.00 [19.00]	139.00 [23.00]	0.5472
<b>TAD</b>	79.00 [17.50]	75.00 [10.75]	81.00 [16.75]	80.00 [18.50]	<b>&lt; 0.05</b>
<b>Vacunas</b>	3.00 [2.00]	3.00 [1.25]	3.00 [1.25]	3.00 [1.00]	0.1655
<b>mMRC</b>					
Grado 0	32 (38.55%)	11 (13.25%)	17 (20.48%)	4 (4.82%)	<b>&lt;&lt; 0.01</b>
Grado 1	33 (39.76%)	9 (10.84%)	8 (9.64%)	16 (19.28%)	
Grado 2	10 (12.05%)	5 (6.02%)	0 (0.00%)	5 (6.02%)	
Grado 3	6 (7.23%)	0 (0.00%)	0 (0.00%)	6 (7.23%)	
Grado 4	2 (2.41%)	0 (0.00%)	0 (0.00%)	2 (2.41%)	
<b>CAT</b>	10.00 [11.00]	9.00 [5.00]	5.00 [4.25]	17.00 [14.25]	<b>&lt;&lt; 0.01</b>

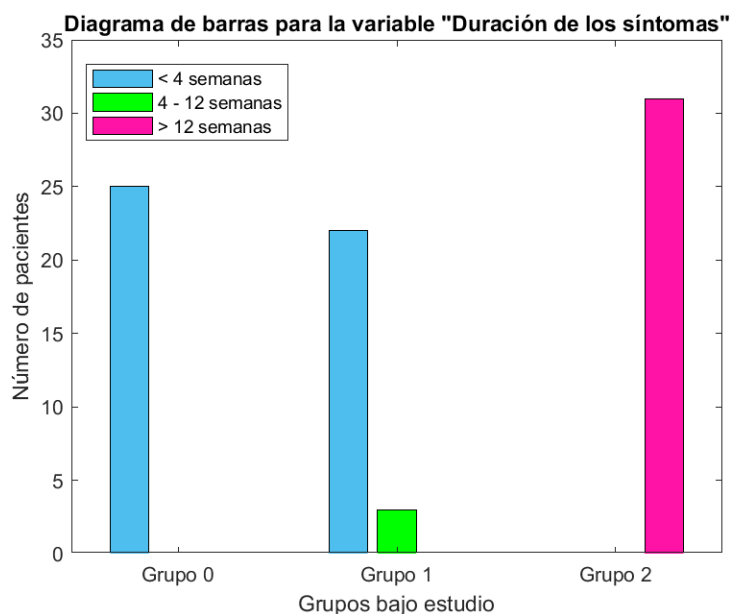
Los datos se expresan en forma de mediana y IQR para las variables continuas y en términos de número y porcentaje para las variables categóricas. HTA: Hipertensión Arterial; DM: Diabetes Mellitus; AOS: Apnea Obstructiva del Sueño; TAS: Tensión Arterial Sistólica; TAD: Tensión Arterial Diastólica; mMRC: tes de disnea mMRC; CAT: test de calidad de vida CAT.

En la tabla anterior se pueden observar los p-valores obtenidos mediante el método de KW referentes al análisis multiclase. A continuación, se presenta el análisis *post-hoc* para las variables que muestran diferencias significativas, desglosado para analizar entre qué grupos concretos se producen estas diferencias.

En el caso de la “duración de los síntomas” es obvio que las diferencias más significativas se den con respecto al grupo SPC-CP, ya que la duración de los síntomas, es decir, presentar síntomas por más de doce semanas, es una de las principales razones para diagnosticar SPC-CP. Además, la mayoría de los clasificados en el grupo 1 presentan los síntomas en un tiempo inferior incluso a las cuatro semanas, haciendo que sea indistinguible con el grupo 0. En la Tabla 4 se muestra el análisis *post-hoc* de esta variable y en la Figura 20 el diagrama de barras correspondiente.

**Tabla 4.** Análisis post-hoc de la variable “Duración de los síntomas”.

Análisis variable “Duración de los síntomas”		p – valor
Grupo 0	Grupo 1	0.8582
Grupo 0	Grupo 2	<< 0.01
Grupo 1	Grupo 2	<< 0.01

**Figura 20.** Diagrama de barras para la variable "Duración de los síntomas".

En el caso de la TAD se observa que las diferencias significativas se producen entre el grupo 0 y el grupo 1, aunque se puede interpretar que hay tendencias también entre el grupo 0 y el grupo 2, es decir, diferencias entre el grupo de aquellos que no han pasado el COVID-19 y aquellos con SPC-CP. En la Tabla 5 se muestra el análisis *post-hoc* de esta variable y en la Figura 21 el diagrama de bloques correspondiente.

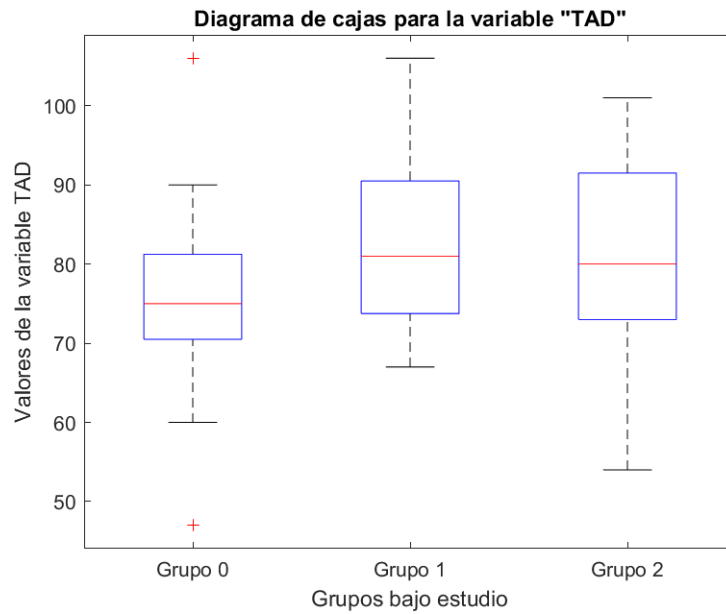
En los dos test clínicos realizados (mMRC: escala de disnea; CAT: cuestionario de calidad de vida) se observaron diferencias entre los dos grupos control (con y sin diagnóstico previo de COVID-19) y el grupo SPC-CP. La Tabla 6 y la Figura 22 se refieren al primer test mencionado, mientras que la

Tabla 7 y la Figura 23 recogen la información del test CAT.

**Tabla 5.** Análisis *post-hoc* de la variable “TAD”.

Análisis variable “TAD”		p – valor
Grupo 0	Grupo 1	< 0.05
Grupo 0	Grupo 2	0.0969
Grupo 1	Grupo 2	0.9138

TAC: tensión arterial diastólica.

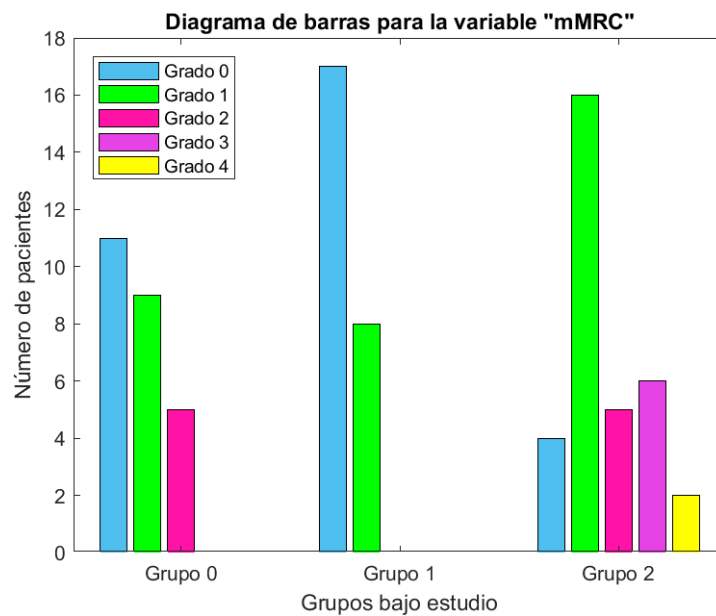


**Figura 21.** Diagrama de cajas para la variable "TAD".

**Tabla 6.** Análisis *post-hoc* de la variable "mMRC".

Análisis variable "mMRC"		p – valor
Grupo 0	Grupo 1	0.1403
Grupo 0	Grupo 2	< 0.05
Grupo 1	Grupo 2	<< 0.01

mMRC: tes de disnea mMRC

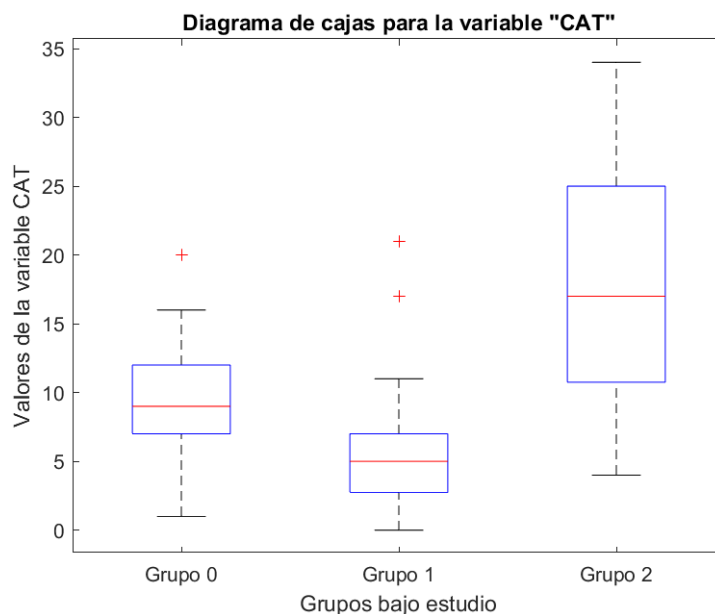


**Figura 22.** Diagrama de barras de la variable "mMRC".

**Tabla 7.** Análisis *post-hoc* de la variable "CAT".

Análisis variable "CAT"		p – valor
Grupo 0	Grupo 1	0.0856
Grupo 0	Grupo 2	< 0.01
Grupo 1	Grupo 2	<< 0.01

CAT: test de calidad de vida CAT.

**Figura 23.** Diagrama de cajas para la variable "CAT".

### 6.3. Caracterización de la señal de variabilidad de la frecuencia cardiaca

En el caso de esta señal, no se encontraron diferencias significativas entre los grupos bajo estudio para ninguna de las variables analizadas. En la Tabla 8.1. y en la Tabla 8.2. se muestran los resultados obtenidos.

**Tabla 8.1.** Mediana, rango intercuartil y p-valor de las características de la señal de HRV para los tres grupos bajo estudio.

Variables	Todos los sujetos	Grupo 0	Grupo 1	Grupo 2	p – valor
AVNN	0.7972 [0.1006]	0.7948 [0.0695]	0.7972 [0.1306]	0.8032 [0.1127]	0.6669
SDNN	0.0435 [0.0231]	0.0477 [0.0281]	0.0391 [0.0277]	0.0435 [0.0164]	0.4081
RMSDD	0.0300 [0.0227]	0.0387 [0.0284]	0.0226 [0.0232]	0.0275 [0.0179]	0.1890

Los datos se expresan en forma de mediana y IQR. AVNN: *average NN intervals*; SDNN: *standard derivation of NN inetrvals*; RMSDD: *root mean square of the differences os successive NN intervals*.

**Tabla 8.2.** Mediana, rango intercuartil y p-valor de las características de la señal de HRV para los tres grupos bajo estudio.

Variables	Todos los sujetos	Grupo 0	Grupo 1	Grupo 2	p – valor
<b>P<sub>VLF</sub></b>	0.5004 [0.7456]	0.7525 [1.7405]	0.4660 [0.6570]	0.4743 [0.6419]	0.1396
<b>P<sub>LF</sub></b>	0.0130 [0.0177]	0.0169 [0.0404]	0.0093 [0.0169]	0.0115 [0.0146]	0.1223
<b>P<sub>HF</sub></b>	0.0006 [0.0029]	0.0011 [0.0102]	0.0004 [0.0016]	0.0005 [0.0015]	0.1459
<b>P<sub>LF/HF</sub></b>	20.0891 [20.4962]	13.5511 [19.5238]	23.3972 [18.3152]	21.6448 [20.8460]	0.1156
<b>SampEn</b>	11.0984 [8.0084]	11.1203 [2.3432]	10.8466 [8.2503]	11.1203 [2.5900]	0.4552
<b>CTM</b>	0.0447 [0.0465]	0.0362 [0.0467]	0.0521 [0.0590]	0.0516 [0.0428]	0.4744
<b>LZC</b>	0.6593 [0.2535]	0.6744 [0.2880]	0.5843 [0.2766]	0.6639 [0.2213]	0.1580

Los datos se expresan en forma de mediana y IQR. P<sub>VLF</sub>: potencia total en la banda de VLF; P<sub>LF</sub>: potencia total en la banda de LF; P<sub>HF</sub>: potencia total en la banda de HF; P<sub>LF/HF</sub>: relación de la potencia absoluta en la banda de LF y la potencia absoluta en la banda HF, también conocido como balance simpático-vagal; SampEn: entropía muestral; CTM: medida de la tendencia central; LZC: complejidad de Lempel-Ziv.

#### 6.4. Caracterización de la señal de saturación de oxígeno en sangre

En la Tabla 9.1 y en la Tabla 9.2 se muestran los valores medianos de los índices oximétricos extraídos de la señal de SpO<sub>2</sub> para cada grupo bajo estudio. La única variable donde se pueden observar diferencias significativas es en la CTM.

**Tabla 9.1.** Mediana, rango intercuartil y p - valor de las características de la señal de SpO<sub>2</sub> para los tres grupos bajo estudio.

Variables	Todos los sujetos	Grupo 0	Grupo 1	Grupo 2	p – valor
<b>SpO<sub>2</sub> media</b>	94.40 [1.90]	94.40 [1.85]	94.20 [1.72]	94.60 [1.78]	0.5079
<b>SpO<sub>2</sub> basal</b>	94.00 [2.00]	94.00 [1.25]	94.00 [2.50]	94.00 [3.00]	0.5415
<b>SpO<sub>2</sub> mínima</b>	93.00 [3.00]	93.00 [2.00]	92.00 [2.50]	93.00 [3.00]	0.5779
<b>CT95</b>	55 [85.50]	55 [87.93]	55.70 [83.30]	54.80 [85.78]	0.9551
<b>CT90</b>	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.2546
<b>ODI2</b>	17.30 [22.23]	16.40 [11.25]	18.50 [16.45]	18.70 [23.83]	0.3013
<b>ODI3</b>	7.10 [9.30]	6.80 [8.70]	8.40 [10.08]	4.90 [14.48]	0.6419

Los datos se expresan en forma de mediana y IQR. SpO<sub>2</sub>: saturación de oxígeno en sangre; CT95: *Cumulative Time below 95%*; CT90: *Cumulative Time below 90%*; ODI2: *Oxygen Desaturation Index (2%)*; ODI3: *Oxygen Desaturation Index (3%)*.

**Tabla 9.2.** Mediana, rango intercuartil y p - valor de las características de la señal de SpO<sub>2</sub> para los tres grupos bajo estudio.

Variables	Todos los sujetos	Grupo 0	Grupo 1	Grupo 2	p – valor
<b>ODI4</b>	0.00 [0.00]	0.00 [0.00]	0.00 [7.55]	0.00 [0.00]	0.2781
<b>PT</b>	11420576.81 [46810.14]	1141079.69 [46354.69]	1142613.29 [45296.90]	1142076.81 [57880.54]	0.7996
<b>MF</b>	0.0156 [0.00]	0.0156 [0.00]	0.0156 [0.00]	0.0156 [0.00]	-
<b>SE</b>	1.6820 [0.00038]	1.6822 [0.00043]	1.6820 [0.00030]	1.6820 [0.00034]	0.1878
<b>SampEn</b>	0.1450 [0.0688]	0.1432 [0.1101]	0.1248 [0.0688]	0.1511 [0.0547]	0.2056
<b>CTM</b>	0.9933 [0.0134]	0.9922 [0.0285]	1.0000 [0.0076]	0.9933 [0.0168]	<b>&lt; 0.05</b>
<b>LZC</b>	0.3017 [0.1371]	0.2743 [0.1234]	0.3017 [0.1440]	0.3017 [0.1509]	0.9097

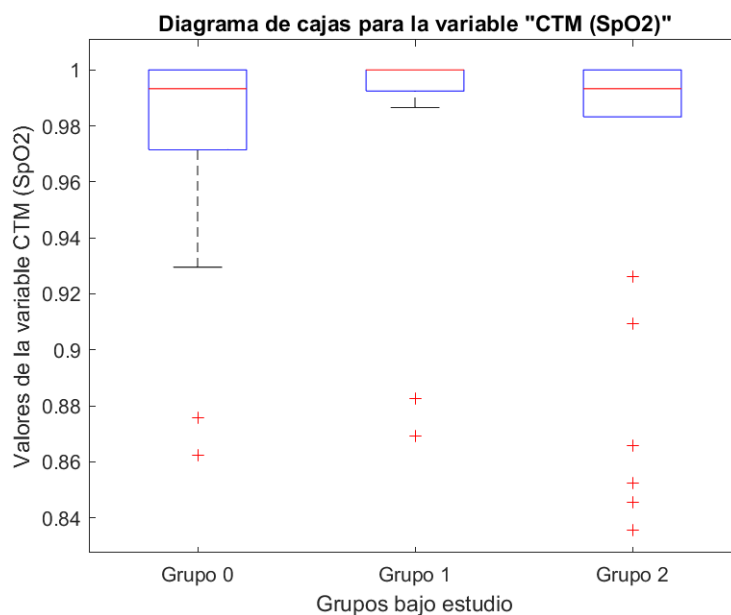
Los datos se expresan en forma de mediana y IQR. ODI4: *Oxygen Desaturation Index* (4%); PT: potencia total, MF: frecuencia mediana; SE: entropía espectral; SampEn: entropía muestral; CTM: medida de la tendencia central; LZC: complejidad de Lempel-Ziv.

Como se ha comentado anteriormente, la CTM mide la variabilidad de una serie temporal. De acuerdo con ello, esperaríamos que en la señal de SpO<sub>2</sub>, cuanto mayor fuese el grado de enfermedad, mayor sería la variabilidad en la señal y, por tanto, menores serían los valores de CTM (ya que la salud viene representada por mayor estabilidad). Aunque no se observó una tendencia creciente homogénea con el grado de enfermedad (control, sin síntomas persistentes, SPC-CP), el grupo SPC-CP presentó valores de CTM significativamente menores (menor estabilidad en la señal y, por tanto, peor condición) que el grupo con COVID sin síntomas persistentes. En la Tabla 10 se muestra el análisis *post-hoc* realizado sobre la variable CTM y en la Figura 24 el diagrama de bloques representativo de esta variable.

**Tabla 10.** Análisis post-hoc de la variable "CTM".

Análisis variable "CTM"		p – valor
Grupo 0	Grupo 1	<b>&lt; 0.05</b>
Grupo 0	Grupo 2	0.8684
Grupo 1	Grupo 2	<b>&lt; 0.05</b>

CTM: medida de la tendencia central.



**Figura 24.** Diagrama de cajas para la variable "CTM (SpO<sub>2</sub>)".

## 6.5. Caracterización del acoplamiento cardiorrespiratorio

La caracterización de la magnitud del CRC se realizó dos veces: en primer lugar, analizando el sincronismo de la señal de HRV con la EDR obtenida mediante la envolvente superior del ECG definida mediante los picos R y, en segundo lugar, analizando el sincronismo de la señal de HRV con la EDR obtenida mediante la envolvente inferior. Esta caracterización se muestra en la Tabla 11 y en la Tabla 12. Al realizar este análisis, únicamente se obtuvieron diferencias significativas en términos de la variable MCC calculada con la señal EDR obtenida mediante la envolvente inferior del ECG.

**Tabla 11.** Mediana, rango intercuartil y p - valor de las características del CRC (EDR mediante la envolvente superior) para los tres grupos bajo estudio.

Variables	Todos los sujetos	Grupo 0	Grupo 1	Grupo 2	p - valor
Diferencia RR	0.0256 [0.0285]	0.0392 [0.0399]	0.0206 [0.0204]	0.0256 [0.0257]	0.0940
Diferencia FC - FR	1.2546 [0.1645]	1.2546 [0.0965]	1.2546 [0.2108]	1.2440 [0.1898]	0.6565
MCC	- 0.0026 [0.0685]	0.0131 [0.0831]	- 0.0177 [0.0676]	- 0.0081 [0.0527]	0.2073
MSC	0.8162 [0.0726]	0.8293 [0.0759]	0.8188 [0.0865]	0.7983 [0.0504]	0.8887
MI	0.3156 [0.1361]	0.3156 [0.1082]	0.3252 [0.1377]	0.3156 [0.1166]	0.8212

Los datos se expresan en forma de mediana y IQR. FC: frecuencia cardiaca; FR: frecuencia respiratoria; MCC: correlación cruzada; MSC: coherencia espectral; MI: información mutua.

**Tabla 12.** Mediana, rango intercuartil y p - valor de las características del CRC (EDR mediante la envolvente inferior) para los tres grupos bajo estudio.

Variables	Todos los sujetos	Grupo 0	Grupo 1	Grupo 2	p – valor
Diferencia RR	0.0232 [0.0221]	0.0259 [0.0307]	0.0229 [0.0129]	0.0222 [0.0144]	0.2138
Diferencia FC – FR	1.2557 [0.1710]	1.2557 [0.0968]	1.2557 [0.2084]	1.2448 [0.1905]	0.7136
MCC	0.0143 [0.0856]	0.0227 [0.0722]	0.0288 [0.1077]	- 0.0042 [0.0755]	< 0.05
MSC	0.8162 [0.0726]	0.8293 [0.0759]	0.8188 [0.0865]	0.7983 [0.0504]	0.8887
MI	0.2792 [0.1284]	0.3305 [0.1030]	0.3305 [0.1371]	0.2647 [0.0749]	0.3466

Los datos se expresan en forma de mediana y IQR. FC: frecuencia cardiaca; FR: frecuencia respiratoria; MCC: correlación cruzada; MSC: coherencia espectral; MI: información mutua.

Un valor de MCC cercano a 1 es indicativo de una fuerte correlación entre las señales, mientras que un valor cercano a 0 indica una falta de correlación entre ellas. El grupo SPC-CP mostró una magnitud (valor absoluto) del coeficiente de correlación mucho menor que el grupo con diagnóstico de COVID-19 pero sin persistencia en los síntomas, es decir, un menor nivel de sincronización entre la actividad respiratoria y la cardiaca. En la Tabla 13 se muestra el análisis *post-hoc* realizado sobre la variable MCC y en la Figura 25 se muestra el diagrama de bloques representativo de esta variable.

## 6.6. Selección de características

Las variables seleccionadas de forma automática mediante el procedimiento de optimización basado en GAs como entrada a la red neuronal fueron las siguientes:

- Presencia de Apnea Obstructiva del Sueño (AOS).
- Potencia de la señal HRV a altas frecuencias ( $P_{HF}$ ).
- Magnitud de la CRC en el dominio de la frecuencia (MSC obtenida mediante la HRV y la EDR de la envolvente superior).
- Test CAT de calidad de vida.

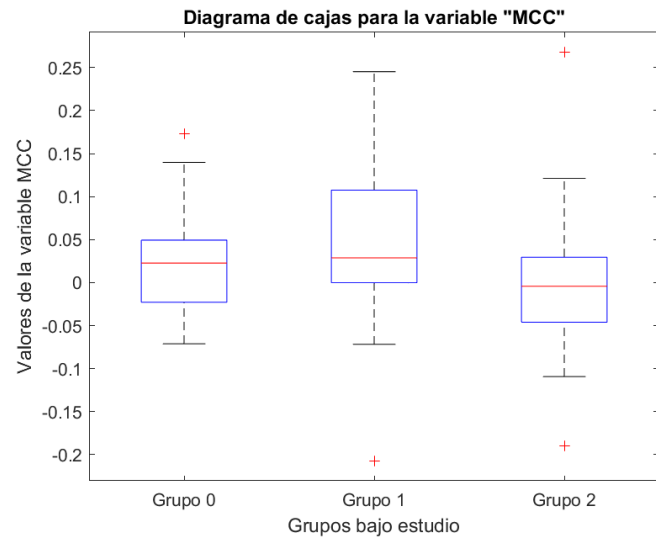
Las distribuciones de datos de las variables seleccionadas se muestran en la Figura 26, la Figura 27, la Figura 28 y la Figura 29.

**Tabla 13.** Análisis *post-hoc* de la variable "MCC (EDR mediante envolvente inferior)".

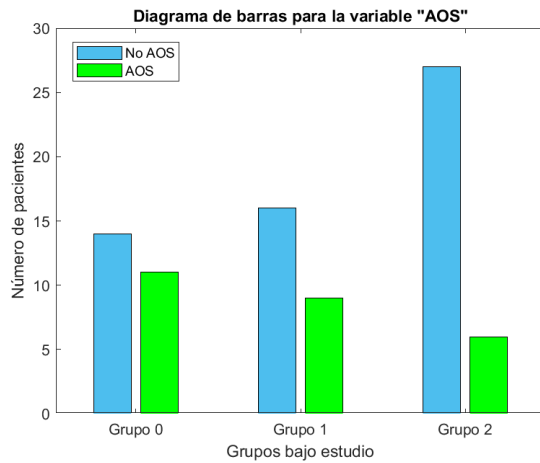
Análisis variable "MCC (EDR inferior)"		p – valor
Grupo 0	Grupo 1	0.9480
Grupo 0	Grupo 2	0.1104
Grupo 1	Grupo 2	< 0.05

MCC: correlación cruzada.

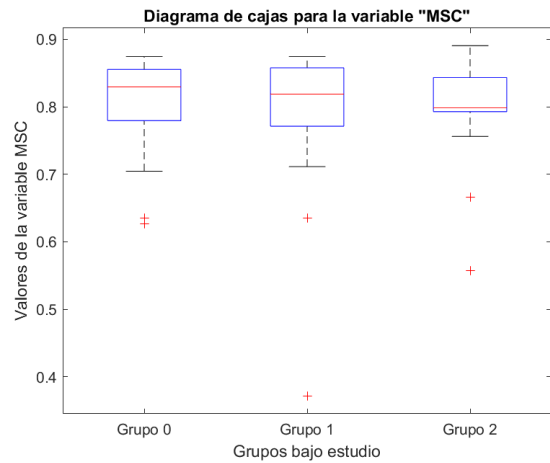




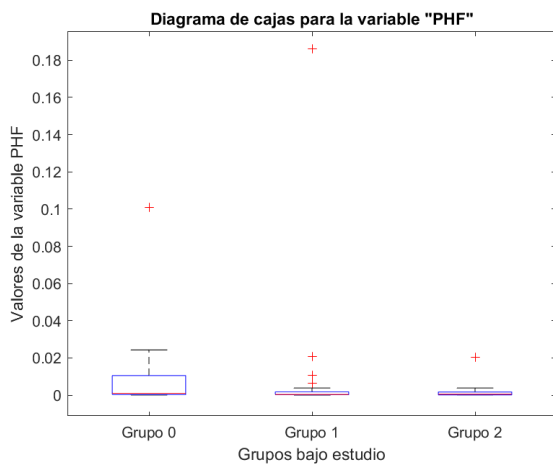
**Figura 25.** Diagrama de cajas para la variable "MCC (EDR mediante envolvente inferior)".



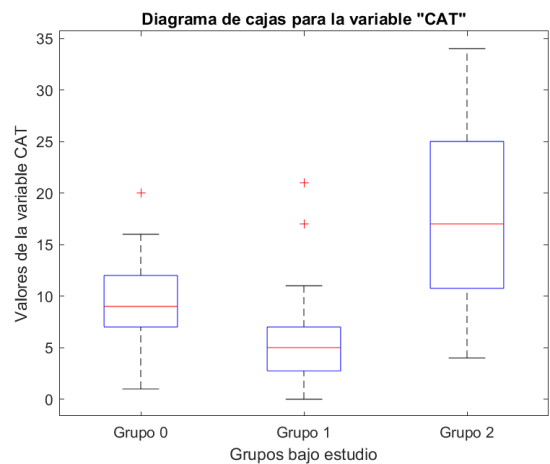
**Figura 26.** Diagrama de barras para la variable "AOS".



**Figura 27.** Diagrama de cajas para la variable "MSC".



**Figura 28.** Diagrama de cajas para la variable "PHF".



**Figura 29.** Diagrama de cajas para la variable "CAT".

Estas variables fueron obtenidas después de evaluar múltiples soluciones potenciales mediante la función de *fitness*, pudiendo ser observado el proceso mediante la gráfica que se adjunta en la Figura 30. En esta gráfica se puede observar la evolución del *fitness* alcanzado por las soluciones potenciales, diferenciando entre la precisión de la mejor solución y la media de la precisión de todas las posibles soluciones para cada una de las iteraciones del algoritmo. Es posible observar que en las primeras iteraciones se va produciendo una mejora notable de precisión, para al final alcanzar un máximo que se mantiene durante 50 generaciones, motivo por el que se alcanza una de las condiciones de finalización del proceso de optimización. Es importante recordar que se eligió un elitismo de uno para no perder la mejor solución en cada nueva generación. Al llegar a la iteración 60 y no presentar mejora, se cumple el criterio de parada, obteniendo como resultado la combinación de las cuatro variables que aportan la métrica de precisión más elevada.

## 6.7. Red neuronal perceptrón multicapa

### 6.7.1. Elección de los hiperparámetros óptimos

Una vez elegidas las variables óptimas para utilizar en el clasificador, se procedió a determinar cuáles eran los hiperparámetros óptimos del modelo de red neuronal final. Para elegirlos, se optó por hacer comprobaciones con distintas combinaciones de  $N_h$  y de  $\alpha$ , para ver con cuál de ellas se obtenía la precisión más alta (se seleccionó la precisión global de 3 clases, *Acc*, como métrica de guiado del proceso de optimización). Finalmente, los valores óptimos fueron  $N_h = 5$  nodes en la capa oculta y  $\alpha = 1$ . En la Figura 31 se muestra la evolución de la métrica de precisión al variar los hiperparámetros de acuerdo a las respectivas particiones y rangos de valores evaluados.

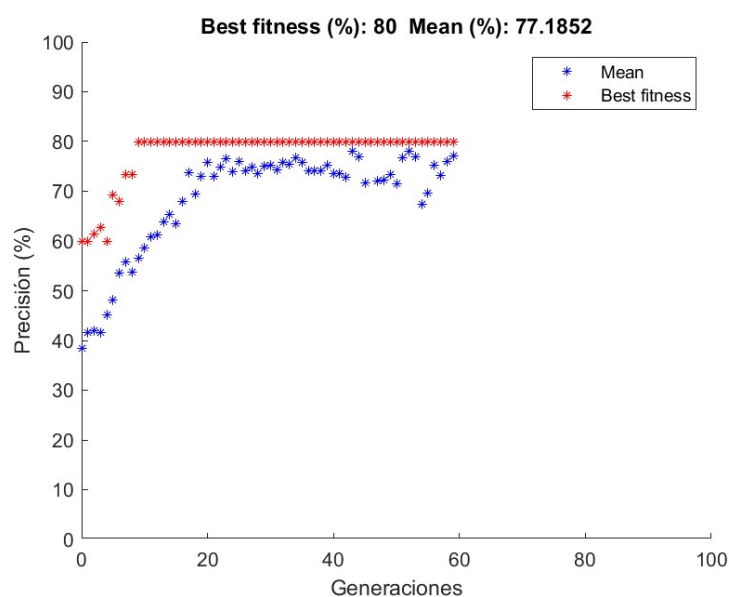
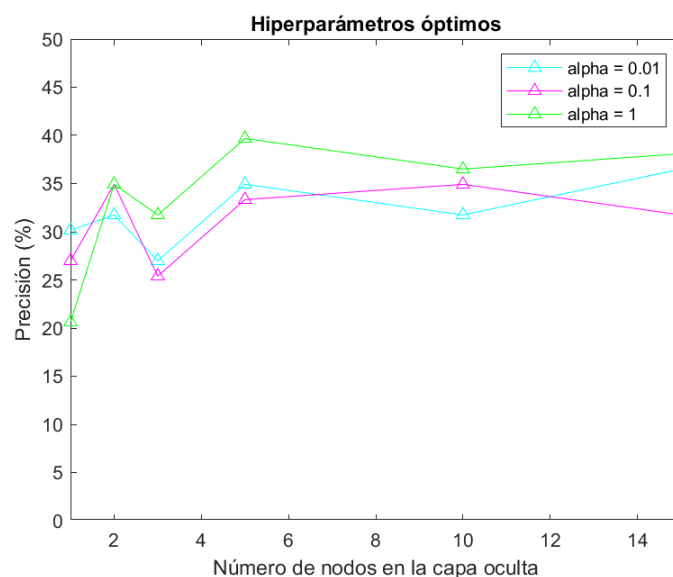


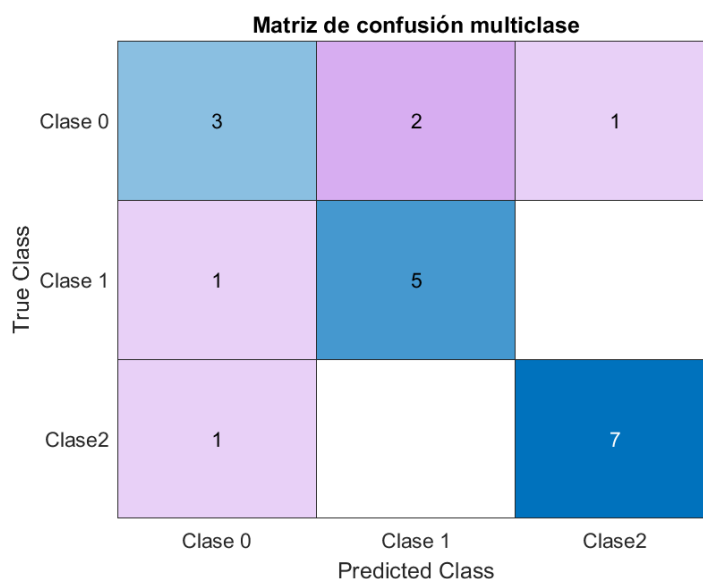
Figura 30. Gráfica del mejor ajuste obtenido con la función *fitness*.



**Figura 31.** Elección de los hiperparámetros óptimos ( $N_h = 5$  y  $\alpha = 1$ ).

### 6.7.2. Evaluación de la capacidad diagnóstica del clasificador

Una vez establecidas las variables de entrada e hiperparámetros óptimos, se procedió a entrenar el modelo final de red neuronal empleando el conjunto completo de entrenamiento compuesto por los 63 pacientes resultantes de la partición Hold-out 1 (ver Figura 16). Su eficacia diagnóstica (*screening* de SPC-CP) se evaluó mediante el conjunto independiente de test de 20 pacientes restantes. Se extrajeron tres métricas de rendimiento para evaluar la capacidad predictiva del clasificador, construyendo primero la matriz de confusión correspondiente únicamente a la clasificación del grupo test. Esta matriz de confusión se muestra en la Figura 32.



**Figura 32.** Matriz de confusión multiclase para evaluar la eficacia del clasificador sobre el conjunto test.

En la matriz de confusión se puede observar que los grupos que mejor clasificó el modelo fueron el 1 y el 2, es decir, los grupos correspondientes a aquellos pacientes que han pasado el COVID-19, ya fuera con o sin síntomas persistentes. También se puede visualizar que entre estos dos grupos no hay confusión a la hora de clasificar, sino que los errores se dan principalmente con los pacientes del grupo 0, es decir, con el grupo correspondiente a aquellos pacientes que no han pasado el COVID-19.

En la Tabla 14 se muestran las métricas de rendimiento de la red neuronal. Es destacable el rendimiento alcanzado en términos de MF1, con un 76.52%. Aunque estos porcentajes podrían no parecer lo suficientemente concluyentes, es importante tener en cuenta que es un clasificador de tres clases y que, por tanto, la posibilidad de asignar erróneamente una categoría siempre va a ser mayor que en uno binario.

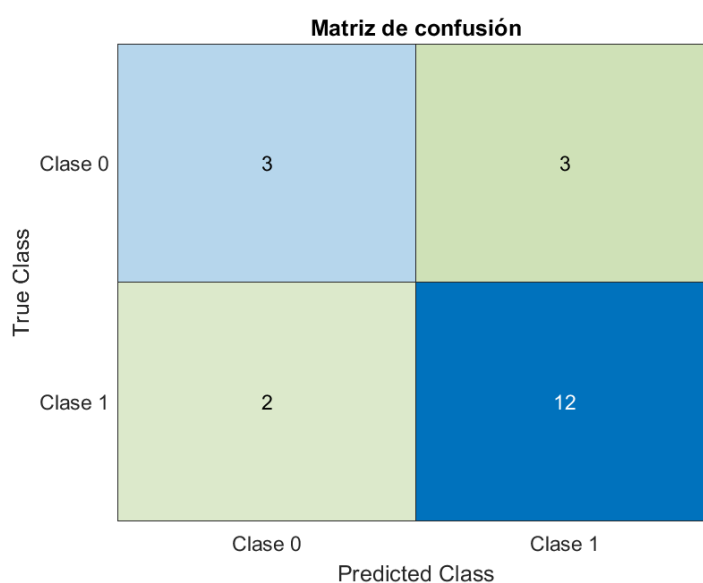
A la vista de estos resultados, se decide agrupar las categorías originales de dos formas diferentes, para evaluar los resultados desde una perspectiva binaria. Además, se calculan las métricas de rendimiento comúnmente derivadas de las matrices de confusión de 2 clases.

La primera forma de reorganizar los grupos de pacientes consistió en fusionar los grupos 1 y 2, es decir, los grupos de pacientes que sí que han pasado el COVID-19. Por lo tanto, la clase 0 continuará siendo el grupo 0 original, mientras que la clase 1 corresponderá a los pacientes de los antiguos grupos 1 y 2. Esta primera agrupación se representa en la matriz de confusión de la Figura 33 y en la Tabla 15 se muestran las métricas obtenidas a partir de esta matriz de confusión.

**Tabla 14.** Métricas utilizadas para evaluar la eficacia del clasificador.

Acc	Kappa	MF1
75.00 %	0.6212	76.52 %

Acc: índice de precisión; Kappa: índice Kappa; MF1: índice Macro F1 Score.



**Figura 33.** Matriz de confusión de clasificación binaria (No COVID vs. (COVID + SPC-CP)).

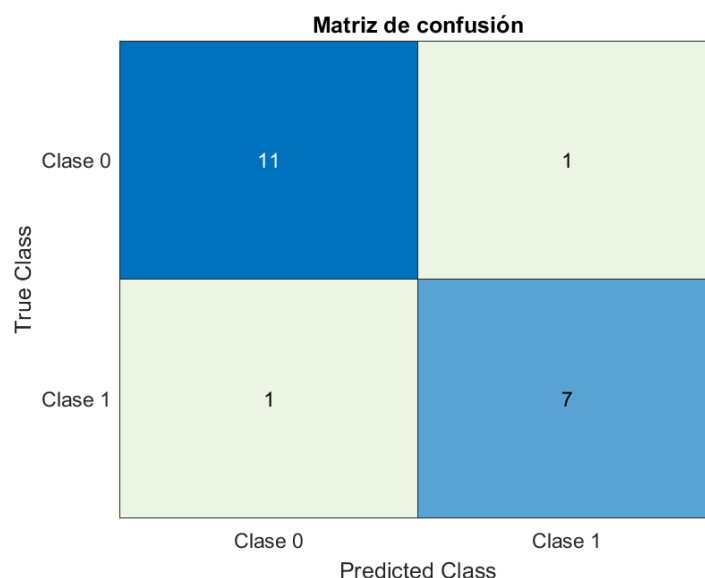
**Tabla 15.** Métricas utilizadas para la evaluación del clasificador agrupando clases (No COVID vs. COVID + SPC-CP).

Se	Sp	PPV	NPV	LR+	LR-	Acc
85.71%	50.00%	80.00%	60.00%	1.71	0.29	75.00%

SE: sensibilidad; Sp: especificidad; PPV: valor predictivo positivo; NPV: valor predictivo negativo; LR+: razón de verosimilitud positiva; LR-: razón de verosimilitud negativa; Acc: índice de precisión.

La segunda forma de organizar los grupos de pacientes originales consistió en fusionar los grupos 0 y 1, es decir, los grupos de pacientes que no presentaban SPC-CP. Por lo tanto, ahora la clase 0 está formada por los pacientes de los antiguos grupos 0 y 1, mientras que la clase 1 estará compuesta por los pacientes del grupo 2 inicial. La Figura 34 muestra la matriz de confusión de estas nuevas categorías binarias y en Tabla 16 muestra las métricas obtenidas a partir de esta matriz de confusión.

Es destacable que la clase que mejor se identifica por la red neuronal es la que representa a los pacientes con SPC-CP, ya que como se puede observar en la Figura 34 solo predice erróneamente dos pacientes, alcanzando una precisión del 90%.

**Figura 34.** Matriz de confusión de clasificación binaria (No COVID + COVID) vs. SPC-CP).**Tabla 16.** Métricas utilizadas para la evaluación del clasificador agrupando clases (No COVID + COVID) vs. SPC-CP).

Se	Sp	PPV	NPV	LR+	LR-	Acc
87.50%	91.67%	87.50%	91.67%	10.50	0.14	90.00%

SE: sensibilidad; Sp: especificidad; PPV: valor predictivo positivo; NPV: valor predictivo negativo; LR+: razón de verosimilitud positiva; LR-: razón de verosimilitud negativa; Acc: índice de precisión.



## CAPÍTULO 7. DISCUSIÓN

En el presente trabajo se han analizado los registros de SpO<sub>2</sub> y ECG realizados a pacientes del Servicio de Neumología, con el objetivo de caracterizar a los pacientes con SPC-CP e identificar las variables con mayor capacidad predictiva a la hora de implementar una red neuronal. De esta manera se podrá discriminar entre estos pacientes y otros individuos con COVID-19 sin síntomas persistentes y controles sin diagnóstico previo de esta enfermedad. A partir de los registros de ECG se obtuvieron las señales de HRV, para estudiar potenciales cambios en la modulación autonómica cardíaca, y de EDR, para analizar la afectación a la interacción cardiopulmonar. El análisis de estas señales se ha estructurado en las tres etapas típicas de reconocimiento automático de patrones: (i) extracción de características y variables clínicas; (ii) selección de características; (iii) clasificación de características.

A continuación, se procede a analizar los resultados y a realizar una comparación con estudios previos centrados en el SPC-CP. Primeramente, se examinará el análisis descriptivo realizado sobre las características extraídas de las señales y las variables clínicas. Posteriormente, se interpretarán las variables seleccionadas mediante el procedimiento de GAs y, a continuación, la capacidad predictiva del modelo final de red neuronal. Se buscarán similitudes y diferencias con otros estudios y, finalmente, se comentarán las limitaciones encontradas en el presente trabajo.

### 7.1. Extracción de características y análisis estadístico

Respecto a las variables clínicas, es esperable encontrar diferencias significativas entre la duración de los síntomas del grupo de SPC-CP y los otros dos grupos, debido a que la duración de los síntomas es parte de la definición de este síndrome. Sin embargo, entre el grupo 0 y el grupo 1 no se observan diferencias estadísticas, ya que los síntomas de la mayoría de los pacientes que han padecido COVID-19 y no han presentado después un SPC-CP no han superado las cuatro semanas. En el caso de la TAD, varios estudios hablan de un incremento de la presión sanguínea y en concreto de la TAD [12], [32], [75], lo que concuerda con los resultados obtenidos en el presente TFG. Se observaron diferencias significativas entre los grupos 0 y 1, aunque se puede observar una tendencia notable entre el grupo 0 y el 2, sin llegar a ser significativa. Respecto a los dos cuestionarios clínicos evaluados, la escala de disnea mMRC y el cuestionario de calidad de vida CAT, ambos presentaron diferencias entre los dos grupos control (con y sin COVID-19) y el grupo de SPC-CP. El primero de los test evalúa la disnea y la discapacidad

asociada, siendo un indicador de exacerbación, mientras que el segundo es utilizado para evaluar la afectación a la calidad de vida de los pacientes con enfermedad crónica respiratoria, típicamente EPOC, aunque su uso es extensible al COVID-19. Es por ello que tiene cabría esperar que los pacientes que presentan SPC-CP obtuviesen puntuaciones mucho más altas en ambos test.

En el caso de la señal de SpO<sub>2</sub>, se decidió llevar a cabo su estudio debido a que la infección aguda por SARS-CoV-2 se caracteriza por una respuesta inmunoinflamatoria y la infiltración de los mediadores inflamatorios en los tejidos pulmonares. Las lesiones pulmonares consiguientes, que pueden identificarse mediante anomalías en la tomografía computarizada torácica, se acompañan de una SpO<sub>2</sub> reducida que puede agravar las respuestas inflamatorias y puede persistir incluso después de la recuperación total [76], [77]. Como se puede observar en la Tabla 10.2 y en la Tabla 11 en esta señal solo se alcanzaron diferencias significativas entre grupos para la variable CTM. Como se ha comentado anteriormente, la CTM cuantifica la variabilidad de una serie temporal, por tanto, en los pacientes con mayor carga de enfermedad se esperaría una señal de SpO<sub>2</sub> con mayor variabilidad y, por tanto, valores más pequeños de CTM (ya que la salud viene representada por la estabilidad). Teniendo esto en cuenta, se esperaría que los valores del grupo de SPC-CP fueran menores al resto de grupos, pero esto solo ocurre respecto al grupo 1, ya que respecto al grupo 0 la mediana de este es inferior al grupo 2. Se encuentran diferencias entre el grupo 0 y el 1 y entre el 1 y el 2.

Al examinar las variables extraídas de la señal de HRV no se encontraron diferencias estadísticamente significativas entre los grupos de pacientes bajo estudio. En este caso sí que se esperaban encontrar tendencias significativas en varias de las variables según lo observado en varias publicaciones previas [11], [75]. Debido a que la banda HF representa el control parasimpático de la frecuencia cardíaca asociado a la modulación respiratoria [63], cabría esperar haber encontrado en esta variable en concreto un patrón significativo. Sin embargo, es muy destacable que, aunque no se observaron diferencias significativas en nuestra muestra en el análisis estadístico univariante para esta variable, la P<sub>HF</sub> de la señal HRV sí que formó parte del subconjunto óptimo de variables seleccionadas mediante GAs. Esto implica que esta variable sí que aporta información complementaria y relevante en la detección del SPC-CP.

Por último, se analizaron las características extraídas para cuantificar la magnitud del CRC, tanto con la EDR de la envolvente superior del ECG como de la obtenida con la EDR de la envolvente inferior. Se encontraron diferencias estadísticamente significativas en la variable MCC obtenida con la EDR de la envolvente inferior entre pacientes con COVID-19 y pacientes con SPC-CP. El análisis de esta variable se decidió llevar a cabo ya que identifica patrones de sincronización o desfase entre las señales cardíaca y respiratoria. El grupo SPC-CP mostró una magnitud (en valor absoluto) del MCC mucho menor que el grupo con diagnóstico de COVID-19 pero sin persistencia en los síntomas, es decir, un menor nivel de sincronización entre la actividad respiratoria y la cardíaca.



## 7.2. Selección y clasificación de características

En el capítulo anterior se determinaron las variables óptimas seleccionadas mediante GAs para ser introducidas en la red MLP y clasificar pacientes entre los tres grupos. Las variables obtenidas tras el procedimiento de selección de características basado en algoritmos genéticos fueron la AOS, la  $P_{HF}$ , la MSC y el test CAT.

En el caso de la variable AOS, varios estudios apuntan a que el SPC-CP está relacionado con alteraciones del sueño. Entre estas alteraciones mencionadas se destacan en los pacientes que presentan un SPC-CP una mala calidad del sueño, somnolencia diurna, insomnio, y trastornos del sueño como la apnea del sueño y el síndrome de las piernas inquietas [78], [79]. Dado que existe una relación bidireccional entre el sueño y el sistema inmunitario, es probable que una mala calidad del sueño influya también en el tiempo de recuperación de las personas que sufren una infección por COVID-19 [79].

La segunda variable seleccionada fue la  $P_{HF}$ , variable de gran interés ya que se relaciona esta banda de frecuencias con el control parasimpático de la frecuencia cardiaca, reflejando las variaciones causadas por la respiración [63]. Además, existen múltiples estudios que al estudiar el SPC-CP mediante la señal HRV encuentran diferencias significativas respecto a grupos control en esta variable concreta [11], [75], lo que concuerda con los resultados obtenidos en el TFG.

La tercera variable seleccionada fue la MSC, en concreto la resultante del análisis del CRC entre la señal de HRV y la señal de EDR obtenida por la envolvente superior en el ECG. La MSC es especialmente útil para analizar la relación entre dos señales biomédicas en el dominio de la frecuencia, evaluando su sincronismo. Destaca en concreto su uso en ECG, donde se puede investigar la sincronización entre diferentes regiones del sistema cardiovascular con otro sistema, en este caso con el respiratorio [60], [61]. En el presente trabajo sería de esperar que aquellos pacientes pertenecientes al grupo SPC-CP presentasen un valor inferior de MSC, indicando una peor interdependencia entre las señales. Como se puede observar en la Figura 27, en el grupo de pacientes con SPC-CP se observa una tendencia hacia valores más bajos en comparación al resto de grupos.

La última variable seleccionada como entrada al clasificador final fue la puntuación correspondiente al test CAT. Como se ha comentado, este test evalúa mediante ocho preguntas, puntuadas cada una de ellas con una puntuación de 0 a 5 puntos de mejor a peor grado, el nivel de influencia que tiene sobre la calidad de vida. Teniendo en cuenta que la EPOC es una enfermedad pulmonar que disminuye el flujo de aire y causa problemas respiratorios, se decidió también evaluar la afectación a la calidad de vida de los pacientes con COVID-19 que formaban parte del estudio mediante este test. La puntuación mínima corresponde a 0 puntos, mientras que la máxima a 40 puntos, correspondiendo la puntuación más alta a la mayor gravedad. Como se observa en la Figura 29, los valores más altos corresponden al grupo con SPC-CP, lo que justificaría la elección de este cuestionario para caracterizar esta condición.

Una vez seleccionadas estas características se procedió a analizar la capacidad diagnóstica del modelo final de red neuronal entrenado. Bajo el enfoque de clasificación multiclase, se obtuvo una precisión global del 75% (Acc), un índice de concordancia kappa de 0.621 (K) y un F1-score promediado entre cada una de las categorías del 76.52% (MF1). Analizando

pormenorizadamente estos resultados, se observó (Figura 32) que los grupos que mejor identificaba la red eran el 1 y el 2, es decir, los grupos de aquellos pacientes que sí que habían pasado el COVID-19. Sin embargo, los pacientes del grupo de control eran incorrectamente clasificados de forma más repetitiva, pudiendo ser debido a que este grupo control es obtenido de las consultas generales del servicio de Neumología y que, por tanto, pueden presentar patologías que, aun siendo compatibles con el estudio, puedan influir negativamente en los resultados obtenidos.

A la vista de estos resultados se optó por realizar un análisis desde un enfoque de clasificación binaria, es decir, agrupando dos de los grupos. Estas agrupaciones se realizaron de dos formas distintas: primero fusionando los grupos que habían pasado el COVID-19 en uno solo y, posteriormente, fusionando los dos grupos control. La precisión alcanzada por el primer enfoque binario es exactamente la misma que el multiclase (Acc 3-clases 75% vs. Acc 2-clases 75%), lo que indica que no hay ganancia predictiva en el hecho de fusionar a los pacientes COVID, ya que lo que se clasifican mal son los controles puros. Por otro lado, es muy destacable la precisión alcanzada en el segundo enfoque binario (Acc 3-clases 75% vs. Acc 2-clases 90%), lo que sugiere que sí que existen unas características clínicas particulares de los pacientes con SPC-CP que se maximizan al diferenciarlos de los pacientes agrupados sin síntomas independientemente de que hayan pasado el COVID o no.

### 7.3. Comparación con otros estudios

Se han encontrado múltiples estudios enfocados en la búsqueda de diferencias significativas entre pacientes con SPC-CP y diferentes grupos control, constituidos en su mayoría por sujetos que no han sufrido infección por COVID-19. La mayoría de estos estudios trabajan a partir de la señal de HRV [11], [75], [80], [81], [82], mientras que alguno, aunque escaso, lo hace mediante la señal de SpO<sub>2</sub> [76], [77], [83]. Sin embargo, ninguno de los estudios consultados valora analizar el CRC para la búsqueda de características significativas del SPC-CP. En la Tabla 17 se resumen las principales características de cada estudio, especificando el año, el tipo de registro biomédico analizado, el número de pacientes en cada grupo a comparar, las variables que cada estudio determina que tienen diferencias significativas y el p-valor de cada una de esas variables. Respecto a las variables, también se recogen aquellas variables clínicas que alcanzaron diferencias estadísticas.

En lo referente a los grupos bajo estudio, tan solo el realizado por *Aranyó et ál.* [80] valoró incluir tres grupos a comparar, recogiendo este tercero a pacientes que habiendo pasado el COVID-19 no presentaban síntomas en la actualidad. El resto de los estudios hacen una comparación entre un grupo de pacientes con SPC-CP y otro grupo con sujetos que no han pasado el COVID-19.

Respecto a las variables relevantes que se repiten con frecuencia en la mayoría de los estudios, se observan las referentes al análisis temporal y espectral de la señal de HRV, destacando la SDNN, la RMSSD y la P<sub>HF</sub>. En los estudios que utilizan la señal de SpO<sub>2</sub>, se repite de forma notable la SpO<sub>2</sub> mínima. En el caso de *Sirayder et ál.* [72], que utiliza la prueba de espirometría, también se alcanzarán diferencias significativas entre los índices espirométricos.

**Tabla 17.** Comparación de las variables significativas obtenidas en diferentes estudios.

Autor	Año	Tipo de registro	Grupo bajo estudio con SPC-CP	Grupo control sin COVID-19	Grupo control COVID-19	Variables significativas	p - valor
Asarcikli LD et ál. [11]	2022	HRV	60	33	-	SDNN RMSSD P <sub>HF</sub> P <sub>LF/HF</sub>	< 0.05 < 0.01 < 0.05 < 0.05
Shah B et ál. [75]	2022	HRV	92	120	-	AVNN SDNN RMSSD	<< 0.01 < 0.05 < 0.05
Al-Hadrawi DS et ál. [76]	2022	SpO <sub>2</sub>	120	36	-	Edad SpO <sub>2</sub> mínima PBT	< 0.01 << 0.01 << 0.01
Al-Hakeim HK et ál. [77]	2022	SpO <sub>2</sub>	120	36	-	Sexo Edad SpO <sub>2</sub> mínima PBT	< 0.05 < 0.05 << 0.01 << 0.01
Sirayder U et ál. [83]	2022	Espirometría SpO <sub>2</sub>	26	26	-	mMRC FVC (%) FEV1 (%) SpO <sub>2</sub> UCI (días)	< 0.01 << 0.01 < 0.05 < 0.05 < 0.05
Aranyó J et ál. [80]	2022	HRV	40	17	19	P <sub>VLF</sub> P <sub>LF</sub> P <sub>HF</sub>	< 0.01 < 0.05 < 0.01
Freire APCF et ál. [81]	2022	HRV	20	18	-	SDNN RMSSD	< 0.05 < 0.05
Kurtoğlu E et ál. [82]	2022	HRV	50	50	-	SDNN RMSSD PT P <sub>VLF</sub> P <sub>LF</sub> P <sub>HF</sub> SampEn	<< 0.01 << 0.01 << 0.01 < 0.01 << 0.01 << 0.01 < 0.01

HRV: variabilidad de la frecuencia cardiaca; SpO<sub>2</sub>: saturación de oxígeno en sangre; SDNN: *standard derivation of NN intervals*; RMSSD: *root mean square of the differences os successive NN intervals*; P<sub>VLF</sub>: potencia total en la banda de VLF; P<sub>LF</sub>: potencia total en la banda de LF; P<sub>HF</sub>: potencia total en la banda de HF; P<sub>LF/HF</sub>: relación de la potencia absoluta en la banda de LF y la potencia absoluta en la banda HF; SampEn: entropía muestral.

Durante la búsqueda de estudios, no se encontró ninguno que implementara algún método de clasificación automática para diferenciar pacientes con SPC-CP de aquellos que no lo padecían.

## 7.4. Limitaciones

Una vez finalizado el presente trabajo, es necesario destacar las principales limitaciones del estudio, que podrían influir en la interpretación y generalización de los resultados alcanzados.

La limitación más relevante es que todos los sujetos proceden del servicio de Neumología, incluidos los controles. Esto supone que aquellos pacientes que se categorizaron como “sanos” por el hecho de no haber sufrido una infección por COVID-19 podrían padecer otras patologías que podrían interferir con las variables del estudio y ser erróneamente clasificados. Sería interesante probar el modelo diseñado con pacientes de otros servicios o de población general para comprobar su validez.

Relacionado con la limitación anterior, al haber realizado todos los registros en el mismo servicio, los dispositivos utilizados para ello fueron los mismos para toda la población bajo estudio. Sería interesante probar este modelo con datos obtenidos de otros equipos de ECG para comprobar si es generalizable a cualquier registro de esta señal.

Otra posible limitación es que, al haber realizado una clasificación entre tres clases, siendo dos de ellas consideradas grupos control, ha sido más complicado alcanzar rendimientos predictivos superiores al 75%. Tal y como se ha comprobado, al agrupar los dos grupos control e intentar diferenciarlos del grupo representante del SPC-CP, los resultados obtenidos por las métricas de clasificación binaria eran más concluyentes.

Evaluando la base de datos construida en este trabajo, la limitación más importante es el tamaño de la misma, ya que al contar con 83 pacientes las diferentes pruebas a la hora de elegir las variables e hiperparámetros óptimos y a la hora de entrenar la red neuronal, los conjuntos de entrenamiento no contaban con un gran número de instancias. El hecho de trabajar con tres clases en una base de datos no lo suficientemente grande también supuso una limitación, porque la representación de cada clase en cada conjunto de entrenamiento era relativamente pequeña. Por otro lado, es importante recalcar que las tres clases se encontraban bastante balanceadas.

Respecto a los registros recogidos, podría ser interesante tomarlos de mayor duración, pudiendo así aplicar las técnicas propuestas en la metodología dividiendo la señal en segmentos. Los métodos de procesado de la señal implementados se seleccionaron debido a su probada validez en otros contextos también centrados en enfermedades respiratorias (AOS, EPOC), no solo referentes al SPC-CP. Por lo tanto, se podrían encontrar nuevas metodologías que derivasen en resultados diferentes a los obtenidos.

Haciendo referencia a la selección de características, hubiese sido beneficioso haber utilizado en este proceso técnicas de *bootstrapping* para remuestrear las instancias a partir del conjunto de datos original, para así haber podido asignar medidas de precisión a los estimadores estadísticos utilizados (error estándar o intervalos de confianza). La principal razón de poder aplicar esta técnica es el contar con un tamaño de población relativamente pequeño realizado la validación del modelo de forma intensiva y evaluando su estabilidad en términos de la probabilidad de *overfitting*.

También es importante tener en cuenta que el COVID-19 ha dejado de ser una enfermedad de declaración obligatoria, desconociendo actualmente en muchos casos si el paciente ha pasado el COVID-19 o no.

Por último, una de las variables seleccionadas fue la presencia de AOS. Se debe tener en cuenta que los pacientes incluidos en el estudio con diagnóstico negativo de AOS acudían en algunos casos a la consulta de sueño, pero no fueron evaluados mediante una prueba de polisomnografía en el momento de su incorporación al estudio, por tanto, no es completamente seguro que no padezcan esta enfermedad.



---

## CAPÍTULO 8. CONCLUSIONES

El SPC-CP actualmente es uno de los desafíos más importantes para la comunidad médica, tanto porque su estudio y comprensión todavía es insuficiente, como por la gran carga asistencial que aún representa el COVID-19. Por esta razón, el presente trabajo se ha centrado en el análisis de la actividad cardíaca (HRV), respiratoria (oximetría) y la interacción entre ambas (CRC), con el principal objetivo de encontrar información relevante que ayude a diagnosticar el SPC-CP.

Para poder llevar a cabo este estudio se realizaron registros de SpO<sub>2</sub> y de ECG a pacientes del servicio de Neumología del Hospital Universitario Río Hortega de Valladolid, analizando finalmente una población compuesta por 83 sujetos. La población bajo estudio estaba constituida por tres grupos distintos de sujetos: dos de ellos formados por participantes correspondientes a los dos grupos de control sin SPC-CP (unos sin diagnóstico previo y otros con diagnóstico confirmado de COVID-19) y el otro por pacientes correspondientes al SPC-CP.

La metodología aplicada se dividió en tres etapas: extracción de características, selección de variables y reconocimiento automático de patrones. En la fase de extracción referente a la señal de SpO<sub>2</sub> se aplicaron técnicas espectrales y no lineales, además de extraer los índices clásicos de esta señal en el dominio del tiempo. En el caso de la señal de HRV, se aplicaron técnicas temporales, espectrales y no lineales. Por último, para el estudio del acoplamiento también se aplicaron técnicas temporales, espectrales y no lineales, siempre con el objetivo de obtener información complementaria de diferentes enfoques de procesado. Para la selección de características se utilizó la técnica de GAs, mientras que para la clasificación automática de los pacientes en los tres grupos de estudio se empleó una red neuronal MLP.

Finalmente, los resultados obtenidos son comparados con otros estudios que aplican metodologías similares de caracterización de la actividad cardíaca o respiratoria. Sin embargo, debido a la falta de estudios, no se ha podido realizar una comparación directa del modelo diagnóstico desarrollado con otros modelos de detección de SPC-CP, lo que sugiere la novedad del trabajo.

## 8.1. Contribuciones

Las principales contribuciones que derivan de este estudio se exponen a continuación:

- Análisis multiclase incluyendo un grupo de pacientes con COVID con síntomas durante el registro, pero de duración inferior a las 12 semanas, observándose diferencias estadísticamente significativas entre estos sujetos y los pacientes con SPC-CP para diferentes variables clínicas y derivadas del procesado de la señal de oximetría y del estudio del CRC.
- Estudio de la influencia del SPC-CP en el CRC, obteniendo además diferencias significativas en términos de la MCC.
- Implementación de un método automático de clasificación de pacientes que podría ser útil en el *screening* de pacientes con SPC-CP, alcanzando una precisión global de 3-clases moderada-alta.
- Identificación de un subconjunto óptimo de variables complementarias, clínicas (AOS, CAT) y derivadas del procesado de señales biomédicas desde diferentes enfoques también complementarios ( $P_{HF}$ , MSC), que podrían suponer un avance en la caracterización y gestión de los pacientes con SPC-CP.

## 8.2. Principales conclusiones del estudio

A continuación, se explican a continuación las conclusiones obtenidas con la realización del trabajo:

- La caracterización de patrones respiratorios en los dominios temporal, espectral y no lineal solamente aportó información relevante estadísticamente significativa a través de la variable CTM correspondiente a la señal de  $SpO_2$ . Los pacientes con SPC-CP mostraron una variabilidad de la señal de oximetría significativamente mayor que los pacientes con diagnóstico de COVID-19 sin síntomas persistentes, lo que sugiere mayor afectación en el intercambio de gases.
- La caracterización de la modulación autonómica cardíaca a través del procesado de la señal HRV no mostró ningún índice autonómico significativo.
- La caracterización del CRC mediante medidas de coordinación y sincronismo en el dominio espectral y no lineal aportó tendencias significativas en las variables MCC.
- Entre las variables clínicas analizadas, se encuentran potenciales diferencias significativas entre los distintos grupos de pacientes en el caso de la duración de los síntomas, la TAD, el test mMRC y el test CAT. Los pacientes con SPC-CP mostraron tensiones, niveles de disnea y afectación de los síntomas a la calidad de vida significativamente mayores que los grupos sin persistencia de los síntomas.



- Las variables AOS, CAT, PHF y MSC han demostrado aportar información complementaria para la identificación de pacientes con SPC-CP.
- Las redes neuronales MLP han demostrado ser útiles en el *screening* de pacientes con SPC-CP, alcanzando una precisión (3-clases) del 75% en la diferenciación de sujetos sin diagnóstico previo de COVID-19, con COVID-19 sin síntomas a largo plazo y de pacientes con SPC-CP, así como una precisión (2-clases) del 90% en el *screening* general de pacientes con SPC-CP frente a controles agrupados sin diagnóstico confirmado de esta condición.

Tras comparar los resultados alcanzados en este estudio con los de otros resultados publicados dentro del mismo contexto de SPC-CP, es posible concluir que el presente trabajo contribuye a aumentar el conocimiento del SPC-CP en la ayuda al diagnóstico y gestión de pacientes con esta condición clínica.

### 8.3. Líneas futuras de investigación

De acuerdo con las conclusiones obtenidas, se procede a exponer una serie de propuestas para futuras líneas de investigación que podrían mejorar el trabajo realizado en el presente estudio y ampliar el conocimiento alcanzado.

El tamaño de la base de datos es uno de los aspectos a mejorar, ya que la utilizada en el presente estudio fue de 83 sujetos. Aunque los grupos se encuentran bastante balanceados, sería más conveniente disponer de más sujetos de cada grupo para conseguir una mayor generalización. De esta forma, también se podría dividir la población en grupos de test y entrenamiento independientes en todas y cada una de las etapas de optimización, sin tener que reutilizar parte de ellos para la selección de características y la elección de los hiperparámetros óptimos. Sería conveniente que el grupo control estuviese constituido por sujetos voluntarios que no tuviesen ningún tipo de patología que pudiese influir sobre la generalización de los resultados obtenidos.

Teniendo en cuenta que en el análisis estadístico realizado sobre la señal de HRV no se han obtenido resultados significativos, se plantea realizar una nueva evaluación con técnicas diferentes a las utilizadas. En lo referente a los métodos espectrales, se propone usar la transformada Wavelet o la STFT. Igualmente, ante la relevancia observada del SPC-CP sobre el acoplamiento cardiorrespiratorio, existen numerosos métodos no lineales que podrían aplicarse para profundizar en la caracterización de la afectación sobre el CRC, como la causalidad de Granger, la predicción no lineal, entre otros.

Los registros utilizados fueron de corta duración, por lo que podría resultar de utilidad tomar registros de mayor duración, para tener más información de las señales. También hay que tener en cuenta que al preprocesar las señales en varias ocasiones el tiempo de registro efectivo se vio reducido por la presencia de artefactos, motivo por el cual sería más beneficioso trabajar con registros de mayor duración.

La aplicación de técnicas eXplainable Artificial Intelligence (XAI) para la interpretación de los modelos diseñados, lo que contribuirá a comprender las decisiones tomadas por los sistemas automáticos y, por tanto, a extender y generalizar su uso.

# ANEXOS

A continuación, se recoge el código desarrollado para el preprocesado y la extracción de variables. También se adjunta el código utilizado para programar la red MLP. En ambos casos primero se adjunta el código del *script* principal y después las funciones que son llamadas en este *script*.

## ANEXO 1. Código para el preprocesado y extracción de variables de la señal de saturación de oxígeno en sangre

```
% Primero se cargan todos los registros de la señal de SpO2.
path_origin = 'C:\Users\Neumologia\Documents\TFG MARIA\REGISTROS PULSIOXIMETRIA\pacientes';
files = dir('*.txt');
Identificadores = [];

frecuencia = [];
no_lineal = [];
se = [];

for i = 1:length(files)
    fprintf('...File: %s (%i/%i)\n', files(i).name(2:end-4), i, size(files,2))

    paciente = load(files(i).name);
    Identificadores1(i) = (string(files(i).name(2:end-4))).';
    Identificadores = Identificadores1.';

    %%% ELIMINACIÓN DE ARTEFACTOS %%%
    saturacion = saturacion_oxigeno(paciente);

    %%% TRANSFORMACIÓN LONGITUD DE LA SEÑAL %%%
    longitud = length(saturacion);
    datos_300 = duracion300(saturacion);

    %%% ANÁLISIS EN EL DOMINIO DE LA FRECUENCIA %%%
    % Primero se calcula la densidad espectral de potencia (PSD) mediante el
    % método de Welch
    window = 30;
    noverlap = 0.5*window;
```

```

nfft = 128;
fs = 1;

[pxx, f] = pwelch(datos_300,window,noverlap,nfft,fs);
figure(1)
plot(f,pxx)
title('Método de Welch')
xlabel('Frecuencia (Hz)')
ylabel('Amplitud (mV)')

[frecuencia(i,1), frecuencia(i,2), frecuencia(i,3)] = analisis_frecuencia(pxx, f);
PT = frecuencia(:,1); MF = frecuencia(:,2); SE = frecuencia(:,3);

%%%% ANÁLISIS NO LINEAL %%%
tam = 0;
radio_ctm = 2.00092;
m = 1;
radio_se = 0.1*std(datos_300);

[no_lineal(i,1), no_lineal(i,2), no_lineal(i,3)] =
analisis_no_lineal(datos_300, tam, m, radio_se, radio_ctm);
SampEn = no_lineal(:,1); CTM = no_lineal(:,2); LZC = no_lineal(:,3);
end

% Una vez extraídas todas las variables, estas son guardadas en una tabla de
Excel.
variables = table(Identificadores, PT, MF, SE, SampEn, CTM, LZC);

filename = 'variablesSpO2.xlsx';
writetable(variables,filename,'Sheet',1,'Range','A1')

```

### **Función: saturacion oxigeno**

```

function saturacion = saturacion_oxigeno(paciente)
% Representación gráficamente la señal de saturación de oxígeno una vez
eliminados los artefactos que tenían. En el caso particular del dispositivo
utilizado los artefactos vendrán representados con el valor 500.

saturacion = paciente(:,2); % La primera columna corresponde a la
frecuencia de pulso.
saturacion((saturacion(:,1) == 500), :) = [];

figure()
plot(saturacion)
title('Saturación de oxígeno (SpO2)')
xlabel('Tiempo (s)')
ylabel('Amplitud (mV)')
end

```

**Función: duracion300**

```
function datos = duracion300(datos)
    % Todas las señales tienen que tener una longitud de 300 muestras, es
    % decir, de 5 minutos. Los datos fuera de este tamaño serán eliminados (primer
    % minuto de registro y el tiempo sobrante del final).

    datos(1:60, :) = [];
    datos(301:length(datos), :) = [];
end
```

**Función: análisis frecuencia**

```
function [PT, MF, SE] = analisis_frecuencia(pxx,f)
    % Cálculo de todas las métricas relevantes para el análisis ESPECTRAL.

    % POTENCIA TOTAL: suma de los coeficientes de la PSD en todo el espectro.
    PT = sum(pxx);
    % FRECUENCIA MEDIANA
    %Inicializamos la variable que irá almacenando la potencia acumulada por
    % las sucesivas componentes espectrales.
    A = 0;
    i = 1;

    % Buscamos la componente frecuencial (Fm) en la que se alcanza la mitad de
    % la potencia total de la PSD (PT/2).
    while A < PT/2
        A = A + pxx(i);
        frecuencia_mediana = f(i);
        i = i + 1;
    end

    MF = frecuencia_mediana;

    % ENTROPIA ESPECTRAL
    % Calculamos la PSD normalizada para convertirla en una función de densidad
    % de probabilidad.
    fdp = pxx/PT;

    % Cálculo de la entropía espectral.
    SE = (-sum(fdp.*log(fdp)));
end
```

**Función: análisis no lineal**

```
function [SampEn, CTM, LZC] = analisis_no_lineal(datos, tam, m, radio_se,
radio_ctm)
    % Cálculo de todas las métricas relevantes para el análisis NO LINEAL.
```

```

% SampEn: análisis no lineal para medir la irregularidad de la serie
temporal.
SampEn = f_calculaSampEn(datos, m, radio_se);

% CTM: análisis no lineal para medir la variabilidad de la serie temporal.
CTM = f_calculaCTM(datos,radio_ctm);

% Lempel-Ziv: análisis no lineal para cuantificar la complejidad de la
serie temporal.
umbral_db = median(datos);
num_simbolos = 2;
LZC = lzcomplexity_tramas(datos,umbral_db,num_simbolos,tam);
end

```

### **Función: f\_calculaSampEn**

```

% ----- %
%           H   Y   D   R   A           %
% ----- %
% Function 'sampen' computes the Sample Entropy of a given signal. %
% %
% Input parameters: %
% - signal:      Signal vector with dims. [1xN] %
% - m:          Embedding dimension (m < N). %
% - r:          Tolerance (percentage applied to the SD). %
% - dist_type:  (Optional) Distance type, specified by a string. %
%              Default value: 'chebychev' (type help pdist for %
%              further information). %
% %
% Output variables: %
% - value:      SampEn value. %
%              %
% ----- %
% Versions: %
% - 1.0:       (21/09/2018) Original script. %
% - 1.1:       (09/11/2018) Upper bound is added. Now, SampEn is %
%              not able to return Inf values. %
% ----- %
% Script information: %
% - Version:   1.0. %
% - Author:    V. Martínez-Cagigal %
% - Date:      21/09/2018 %
% ----- %
% References: %
% [1] Richman, J. S., & Moorman, J. R. (2000). Physiological %
% time-series analysis using approximate entropy and sample %
% entropy. American Journal of Physiology-Heart and %
% Circulatory Physiology, 278(6), H2039-H2049. %
% ----- %

```

---

```

function value = f_calculaSampEn(signal, m, r, dist_type)

% Error detection and defaults
if nargin < 3, error('Not enough parameters.');
```

```

end
if nargin < 4
    dist_type = 'chebychev';
    fprintf('[WARNING] Using default distance method: chebychev.\n');
```

```

end
if ~isvector(signal)
    error('The signal parameter must be a vector.');
```

```

end
if ~ischar(dist_type)
    error('Distance must be a string.');
```

```

end
if m > length(signal)
    error('Embedding dimension must be smaller than the signal length
(m<N).');
```

```

end

% Useful parameters
signal = signal(:)';
N = length(signal);    % Signal length
sigma = std(signal);   % Standard deviation

% Create the matrix of matches
matches = NaN(m+1,N);
for i = 1:1:m+1
    matches(i,1:N+1-i) = signal(i:end);
end
matches = matches';

% Check the matches for m
d_m = pdist(matches(:,1:m), dist_type);
if isempty(d_m)
    % If B = 0, SampEn is not defined: no regularity detected
    % Note: Upper bound is returned
    value = Inf;
else
    % Check the matches for m+1
    d_m1 = pdist(matches(:,1:m+1), dist_type);

    % Compute A and B
    % Note: logical operations over NaN values are always 0
    B = sum(d_m <= r*sigma);
    A = sum(d_m1 <= r*sigma);

    % Sample entropy value
    % Note: norm. comes from [nchoosek(N-m+1,2)/nchoosek(N-m,2)]
    value = -log((A/B)*((N-m+1)/(N-m-1)));
end

% If A=0 or B=0, SampEn would return an infinite value. However, the

```

```

% lowest non-zero conditional probability that SampEn should
% report is A/B = 2/[(N-m-1)(N-m)]
if isinf(value)
    % Note: SampEn has the following limits:
    %     - Lower bound: 0
    %     - Upper bound: log(N-m)+log(N-m-1)-log(2)
    value = -log(2/((N-m-1)*(N-m)));
end
end

```

### **Función: f\_calculaCTM**

```

function CTM = f_calculaCTM(registro, radio)
% Función que calcula la Medida de la Tendencia Central (CTM). La CTM es
% un parámetro que mide el grado de variabilidad de una serie temporal,
% haciendo uso para su cálculo de diagramas de dispersión los cuales
% representan diferencias de segundo orden. Estas están formadas a partir
% de del desplazamiento temporal de la señal original.
% La CTM se calcula seleccionando una región circular de radio r,
% alrededor del origen, contando el número de puntos que caen dentro del
% radio y dividiendo por el número total de puntos. Esto implica que el
% valor se encuentre siempre dentro del rango de 0 a 1.
%
% Argumentos de entrada:
% - registro: serie de datos de entrada de la que estimaremos su CTM.
% - radio: valor del radio que definirá la región circular en la que se
% van a contar los puntos que caen dentro.
%
% Argumentos de salida:
% - CTM: valor de la CTM para el registro introducido.
%
% Programado por: María Fernández Vaquerizo
%
% Última actualización: 30/03/2023

x = zscore(registro);
sum = 0;

for i = 1:length(x)-2
    if ((x(i+2)-x(i+1))^2 + (x(i+1)-x(i))^2)^(1/2) < radio
        sum = sum + 1;
    end
end

CTM = 1/(length(x)-2)*sum;
end

```



**Función: lzcomplexity tramas**

```

function LZcomplexity = lzcomplexity_tramas(registro, umbral_db,
num_simbolos,tam)
% Función que calcula la complejidad de Lempel-Ziv (Lempel-Ziv
% complexity, LZC) mediante el algoritmo propuesto en A. Lempel and J.
% Ziv, "On the complexity of finite sequences," IEEE Transactions on
% Information Theory, vol. IT-22, pp. 75-81, 1976.
%
% El LZC es una medida no paramétrica de cuantificación de la complejidad
% en series de datos finitas. LZC está directamente relacionada con el
% número de cadenas (subsecuencias) diferentes y su repetición a lo largo
% de una secuencia de datos. Para detectar las diferentes subsecuencias,
% se convierte la señal original en una secuencia binaria de símbolos
% mediante la comparación con un umbral, habitualmente la media o la
% mediana de la serie de datos a analizar.
% La serie de entrada es dividida en segmentos, de forma que se obtiene
% un valor de LZC para cada segmento. Finalmente se realiza un
% promediado para obtener un único valor de LZC para la serie.
%
% Argumentos de entrada:
%
% - registro: Serie de datos de entrada de la que estimaremos su LZC.
% - umbral_db: Umbral empleado en la conversión binaria de la serie.
% Tomará los valores 'media' o 'mediana'.
% - num_simbolos: Número de símbolos empleados en la conversión binaria
% de la serie de datos original.
% - tam: Número de muestras de los segmentos en que se dividirá
% la serie de datos original. Si tam = 0, se aplicará el
% algoritmo sobre la serie original sin segmentar.
%
% Variables de salida:
% - LZcomplexity: Vector de valores de LZC para cada segmento
% del registro de SpO2. Este vector tiene tamaño 1xk,
% donde k es el número de segmentos de tamaño 'tam'
% muestras en que se posible dividir el registro de
% SpO2
%
% Grupo de Ingeniería Biomédica
% http://www.gib.tel.uva.es
% Universidad de Valladolid
%
% Programado por: Daniel Abásolo Baz.
% Modificado por: Daniel Álvarez González.
%
% Última actualización: 29/11/2007
%
% LZC se puede calcular sobre el registro completo (tam=0) o
% sobre el registro dividido en tramas de tamaño 'tam' muestras.
% Calculamos el número de veces a aplicar el algoritmo en la señal
% correspondiente viendo en cuántas tramas no solapadas de longitud
% 'tam' es posible dividir el registro de SpO2 original

```

---

```

if tam==0 % Si tam = 0 el algoritmo se calcula sobre el registro completo
    n_tramas=1; % El algoritmo se aplica una vez n_tramas=1
    tam=length(registro);
else % El algoritmo se aplica n_tramas veces
    n_tramas=floor(length(registro)/tam);
end
% Calculamos LZC para cada trama
n_tramas
for l=1:n_tramas
    l
    % Extraemos la trama correspondiente del registro original.
    trama=registro((1+(l-1)*tam):(l*tam));

    % Número de muestras de la trama.
    n=length(trama);

    % Transformación de la trama en una secuencia binaria
    trama=transforma_binaria(trama,num_simbolos,umbral_db);

    if num_simbolos == 2
        b=n/log2(n);
    else
        b=n/(log(n)/log(3));
    end

    % Inicializamos las variables empleadas en el cálculo la complejidad LZ.
    c = 1; % Valor inicial del contador de complejidad.
    S = trama(1); % Inicialización de la subsecuencia S.
    Q = trama(2); % Inicialización de la subsecuencia Q.

    for i = 2:tam
        % Concatenamos ambas subsecuencias.
        SQ = [S,Q];
        % Eliminamos el último carácter de la subsecuencia resultado de la
        % concatenación.
        SQ_pi = [SQ(1:(length(SQ)-1))];

        % Comprobamos si la subsecuencia Q se encuentra contenida en SQ_pi
        % Para encontrar coincidencias dentro de la secuencia de símbolos
        % bajo estudio, se considera que el contenido de la secuencia es
        % una cadena de caracteres y se utiliza una función típica de
        % comparación de cadenas, muy común en las librerías básicas de
        % muchos lenguajes de programación.
        indice = findstr(Q,SQ_pi); % Nos da los índices en los que Q empieza
        dentro de SQ_pi

        if length(indice)==0
            % Q no se encuentra al inspeccionar SQ_pi: Q es una secuencia
            nueva.

            c = c+1; % Incrementamos el contador de complejidad.
            if (i+1)>tam % Si llegamos al final de la serie hemos
            terminado.

```

```

        break;
    else % Si no hemos llegado al final concatenamos
las subsecuencias S y Q.
        S = [S,Q]; % Formamos una nueva subsecuencia S.
        Q = trama(i+1); % Actualizamos la subsecuencia Q.
    end
    else
        % Q forma parte de SQ_pi.
        if (i+1)>tam % Si llegamos al final de la serie hemos
terminado.
            break;
        else
            Q = [Q,trama(i+1)]; % Extendemos la subsecuencia Q.
        end
    end
end
end
% Normalizamos el contador de complejidad c de tal forma que  $0 \leq c/b$ .
%  $\leq 1$ 
LZcomplexity(1) = c/b;
end

```

### **Función: transforma binaria**

```

function s = transforma_binaria(serie, num_simbolos, umbral)
% Esta función transforma una señal temporal en una serie compuesta por un
% número de símbolos determinado de naturaleza binaria.
%
% Argumentos de entrada:
% - serie: Señal temporal a transformar
% - num_simbolos: Número de símbolos a utilizar en la transformación (2 o 3)
% - umbral: Umbral de decisión a aplicar en la transformación.
%           Será una cadena de caracteres que indique si el
%           umbral es la 'media' o la 'mediana' de la serie
%           temporal bajo estudio.
%
% Argumentos de salida:
% - s: Señal transformada. Es la secuencia de símbolos
%      binarios
%
% Grupo de Ingeniería Biomédica
% http://www.gib.tel.uva.es
% Universidad de Valladolid
%
% Programado por: Alicia Rodrigo de Diego y Jose Victor Marcos Martin.
% Modificado por: Daniel Álvarez González.
%
% Última actualización: 11/12/2007
if num_simbolos==2
    if strcmp(umbral,'mediana')
        mediana=median(serie);
        for i=1:1:length(serie)

```

---

```

        if serie(i)<mediana
            s(i)=0;
        else
            s(i)=1;
        end
    end
else
    media=mean(serie);
    for i=1:1:length(serie)
        if serie(i)<media
            s(i)=0;
        else
            s(i)=1;
        end
    end
end

else
    mediana=median(serie);
    mx=abs(max(serie));
    mn=abs(min(serie));

    td1=mediana-mn/16;
    td2=mediana+mx/16;

    for i=1:1:length(serie)
        if serie(i)<=td1
            s(i)=0;
        else if serie(i)<td2
            s(i)=1;
        else
            s(i)=2;
        end
    end
end
end
end

```

## ANEXO 2. Código para el preprocesado y extracción de variables de la señal de variabilidad de la frecuencia cardiaca

```

% Primero se cargan todos los registros de la señal de HRV
path_origin = 'C:\Users\Neumologia\Documents\TFG MARIA\REGISTROS ECG\pacientes_HRV';
files = dir('*.csv');

```

```
Identificadores = [];
```

```
tiempo = [];
frecuencia = [];
```

---

```

no_lineal = [];

for i = 1:length(files)
    fprintf('...File: %s (%i/%i)\n', files(i).name(2:end-4), i, size(files,2))

    Identificadores1(i) = (string(files(i).name(2:end-4))).';
    Identificadores = Identificadores1.';

    paciente = readmatrix(files(i).name, 'Delimiter',';', 'DecimalSeparator','.');

    %% ELIMINACIÓN DE ARTEFACTOS Y DISMINUCIÓN DE LA LONGITUD DE LA SEÑAL %%
    registro = eliminar_muestras_perdidas(paciente, i);

    %% ANÁLISIS EN EL DOMINIO DEL TIEMPO %%
    [tiempo(i,1) tiempo(i,2) tiempo(i,3)] = analisis_tiempo(registro);
    AVNN = tiempo(:,1); SDNN = tiempo(:,2); RMSSD = tiempo(:,3);

    %% ANÁLISIS EN EL DOMINIO DE LA FRECUENCIA %%
    window = 30*3.41;
    noverlap = 0.5*window;
    nfft = 128;
    fs = 3.41;

    registro_interpolado = interp1(registro, 1:(1/3.41):length(registro), 'spline');
    [pxx, f] = pwelch(registro_interpolado,window,noverlap,nfft,fs);

    figure()
    plot(f,pxx)
    title('Método de Welch')
    xlabel('Frecuencia (Hz)')
    ylabel('Amplitud (mV)')

    [frecuencia(i,1) frecuencia(i,2) frecuencia(i,3) frecuencia(i,4)] =
    analisis_frecuencia(pxx,f);
    VLF = frecuencia(:,1); LF = frecuencia(:,2); HF = frecuencia(:,3); rel_LH
    = frecuencia(:,4);

    %% ANÁLISIS EN EL DOMINIO NO LINEAL %%
    tam = 0;
    radio_ctm = 0.122418;
    m = 2;
    radio_se = 0.15*std(registro);

    [no_lineal(i,1), no_lineal(i,2), no_lineal(i,3)] =
    analisis_no_lineal(registro, tam, m, radio_se, radio_ctm);
    SampEn = no_lineal(:,1); CTM = no_lineal(:,2); LZC = no_lineal(:,3);
end

% Una vez extraídas todas las variables, estas son guardadas en una tabla de
Excel.
variables = table(Identificadores, AVNN, SDNN, RMSSD, VLF, LF, HF, rel_LH,
SampEn, CTM, LZC);

```

```
filename = 'variablesHRV.xlsx';
writetable(variables,filename,'Sheet',1,'Range','A1')
```

### **Función: eliminar muestras perdidas**

```
function HRV = eliminar_muestras_perdidas(registro, num_reg)
% Se acorta la señal para que corresponda a 5 minutos de registros y se
preprocesa para eliminar cualquier irregularidad que pueda haber que no
sea compatible fisiológicamente.

HRV = registro(:,2);
t = registro(:,1);
HRV_original = HRV;
t_original = t;

%%% DISIMINUCIÓN DE LA LONGITUD DEL REGISTRO A 300 MUESTRAS %%%
t_original(1:60, :) = [];
lim = find(t_original > 360);
t_original(lim(:,1):length(t_original), :) = [];

HRV_original(1:60, :) = [];
HRV_original(lim(:,1):length(HRV_original), :) = [];

t(1:60, :) = [];
lim = find(t > 360);
t(lim(:,1):length(t), :) = [];

HRV(1:60, :) = [];
HRV(lim(:,1):length(HRV), :) = [];

%%% ELIMINACIÓN DE MUESTRAS PERDIDAS %%%
% 1. EL VALOR DE LAS MUESTRAS TIENE QUE SER MAYOR QUE 0.33
pos_033 = find(HRV < 0.33);

while ~isempty(pos_033)
    HRV_new = [(HRV(1:pos_033(1)-1))' (HRV(pos_033(1)+1:end))']';
    HRV_new(pos_033(1)) = HRV(pos_033(1)) + HRV(pos_033(1)+1);
    t_new = [(t(1:pos_033(1)-1))' (t(pos_033(1)+1:end))']';

    HRV = HRV_new;
    t = t_new;
    pos_033 = find(HRV < 0.33);
end

% 2. EL VALOR DE LAS MUESTRAS TIENE QUE SER MENOR QUE 1.50
pos_150 = find(HRV > 1.50);

while ~isempty(pos_150)
    pendiente=(HRV(pos_150(1)+1)-HRV(pos_150(1)-1))/(t(pos_150(1)+1)-t(pos_150(1)-1));
    HRV(pos_150(1)) = HRV(pos_150(1)-1)+pendiente*(t(pos_150(1))-t(pos_150(1)-1));
```

---

```

t_new = (t(pos_150(1)-1) + t(pos_150(1)))/2;
HRV_new = HRV(pos_150(1)-1) + pendiente*(t_new - t(pos_150(1)-1));

HRV = [HRV(1:(pos_150(1)-1))' HRV_new HRV(pos_150(1):end)']';
t = [t(1:(pos_150(1)-1))' t_new t(pos_150(1):end)']';
pos_150 = find(HRV > 1.50);
end

% 3. LA DISTANCIA DE LA DIFERENCIA TIENE QUE SER MENOR DE 0.66
vector_dif = abs(diff(HRV));
pos_066 = find(vector_dif > 0.66);

valores_intrp = interp1(t, HRV, pos_066, 'linear');

for i = 1:length(HRV)
    for j = 1:length(pos_066)
        if i == pos_066(j)
            HRV(i,1) = valores_intrp(j);
        end
    end
end

% 4. ELIMINAR MUESTRAS PERDIDAS NO DETECTADAS CON LAS CONDICIONES
ANTERIORES
% 4.1. Valores cercanos a 0.33:
puntos_inf = [];
desv = std(HRV);

for k = 1:length(HRV)
    umbral_inf = HRV(k) - 0.33;

    if umbral_inf < desv
        puntos_inf = [puntos_inf' k]';
    end
end

while ~isempty(puntos_inf)
    HRV_new = [(HRV(1:puntos_inf(1)-1))' (HRV(puntos_inf(1)+1:end))']';
    HRV_new(puntos_inf(1)) = HRV(puntos_inf(1)) + HRV(puntos_inf(1)+1);
    t_new = [(t(1:puntos_inf(1)-1))' (t(puntos_inf(1)+1:end))']';

    HRV = HRV_new;
    t = t_new;
    puntos_inf = [];
    desv = std(HRV);

    for k = 1:length(HRV)
        umbral_inf = HRV(k) - 0.33;
        if umbral_inf < desv
            puntos_inf = [puntos_inf' k]';
        end
    end
end

```

```

        end
    end
end

% 4.2. Valores cercanos a 1.50:
puntos_sup = [];
desv2 = std(HRV);

for k = 1:length(HRV)
    umbral_sup = 1.50 - HRV(k);

    if umbral_sup < desv2
        puntos_sup = [puntos_sup' k]';
    end
end

while ~isempty(puntos_sup)
    pendiente = (HRV(puntos_sup(1)+1) - HRV(puntos_sup(1)-1))/(t(puntos_sup(1)+1) - t(puntos_sup(1)-1));
    HRV(puntos_sup(1)) = HRV(puntos_sup(1)-1) + pendiente*(t(puntos_sup(1)) - t(puntos_sup(1)-1));
    t_new = (t(puntos_sup(1)-1) + t(puntos_sup(1)))/2;
    HRV_new = HRV(puntos_sup(1)-1) + pendiente*(t_new - t(puntos_sup(1)-1));

    HRV = [HRV(1:(puntos_sup(1)-1))' HRV_new HRV(puntos_sup(1):end)]';
    t = [t(1:(puntos_sup(1)-1))' t_new t(puntos_sup(1):end)]';

    puntos_sup = [];
    desv2 = std(HRV);

    for k = 1:length(HRV)
        umbral_sup = 1.50 - HRV(k);

        if umbral_sup < desv2
            puntos_sup = [puntos_sup' k]';
        end
    end
end

%%% REPRESENTACIÓN DE LA SEÑAL HRV ORIGINAL Y DE LA SEÑAL HRV PROCESADA
%%%
figure(1)
plot(t, HRV, 'LineWidth', 3)
hold on
plot(t_original, HRV_original, 'LineWidth', 1)
plot(t, 1.5*ones(1, length(t)), '-', t, 0.33*ones(1, length(t)))
title('Señal HRV', num_reg)
legend('Señal HRV sin muestras perdidas', 'Señal HRV original')
xlabel('Tiempo (s)')
ylabel('Tiempo')

```



```

hold off
end

```

### **Función: análisis tiempo**

```

function [AVNN, SDNN, RMSSD] = analisis_tiempo(datos)
% Cálculo de todas las métricas relevantes para el análisis TEMPORAL.

% AVNN: el promedio del intervalo de pulso a pulso, siendo una estimación
% global del período entre latidos (inverso de la frecuencia del pulso).
AVNN = mean(datos);

% SDNN: la desviación estándar de los intervalos de pulso a pulso, que
% cuantifica el grado de variabilidad.
SDNN = std(datos);

% RMSSD: la raíz cuadrada media de las diferencias sucesivas de intervalos
% de pulso a pulso (RMSSD), que representa la actividad vagal.
diferencias = diff(datos);
RMSSD = rms(diferencias);
end

```

### **Función: análisis frecuencia**

```

function [VLFn, LFn, HFn, rel_LHn] = analisis_frecuencia(pxx, f)
% Cálculo de todas las métricas relevantes para el análisis ESPECTRAL.

% VLF: potencia en la banda de muy bajas frecuencias (0.003-0.04 Hz).
VLF = sum(pxx(find(f>=0.003 & f<0.04)));

% LF: la potencia en la banda de bajas frecuencias (0.04-0.15 Hz).
LF = sum(pxx(find(f>=0.04 & f<0.15)));

% HF: la potencia en la banda de altas frecuencias (0.15-0.40 Hz).
HF = sum(pxx(find(f>=0.15 & f<0.40)));

% Normalización de las variables
VLFn = VLF/(LF+HF);
LFn = LF/(LF+HF);
HFn = HF/(LF+HF);

% LF/HF: la relación a baja y alta frecuencia.
rel_LHn = LF/HF;
end

```

### **Función: análisis no lineal**

```

function [SampEn, CTM, LZC] = analisis_no_lineal(datos, tam, m, radio_se,
radio_ctm)

```

```

% Cálculo de todas las métricas relevantes para el análisis NO LINEAL.

% SampEn: análisis no lineal para medir la irregularidad de la serie
temporal.
SampEn = f_calculaSampEn(datos, m, radio_se, 'chebychev');

% CTM: análisis no lineal para medir la variabilidad de la serie temporal.
CTM = f_calculaCTM(datos, radio_ctm);

% Lempel-Ziv: análisis no lineal para cuantificar la complejidad de la
serie temporal.
umbral_db = median(datos);
num_simbolos = 2;
LZC = lzcomplexity_tramas(datos,umbral_db,num_simbolos,tam);
end

```

### ANEXO 3. Código para el preprocesado y extracción de variables de la señal de EDR para la cuantificación del acoplamiento cardiorrespiratorio

```

% Primero se cargan todos los registros de la señal de HRV y del EG
cd      'C:\Users\maria\OneDrive\Documentos\INGENIERIA      BIOMEDICA\TFG
MARIA\REGISTROS ECG\pacientes_HRV'

files_HRV = dir('*.csv');
ident_HRV = [];

cd      'C:\Users\maria\OneDrive\Documentos\INGENIERIA      BIOMEDICA\TFG
MARIA\REGISTROS ECG\pacientes_ECG'

files = dir('*.txt');
Identificadores = [];

addpath 'C:\Users\maria\OneDrive\Documentos\INGENIERIA BIOMEDICA\TFG MARIA\REGISTROS
ECG'

% Cargar vector con la frecuencia de pulso de cada paciente.
fp = load('FP.txt');

variables_sup = []; variables_inf = [];
correlacion_sup = []; correlacion_inf = [];
coherencia_sup = []; coherencia_inf = [];
infom_sup = []; infom_inf = [];

for i = 1:length(files)
    %% IDENTIFICAR ARCHIVOS EN MATLAB DE LAS SEÑALES DE HRV %%
    cd      'C:\Users\maria\OneDrive\Documentos\INGENIERIA      BIOMEDICA\TFG
MARIA\REGISTROS ECG\pacientes_HRV'
    fprintf('...File: %s (%i/%i)\n', files_HRV(i).name(2:end-4), i, size(files_HRV,2))

```

```

ident_HRV1(i) = (string(files_HRV(i).name(2:end-4))).';
ident_HRV = ident_HRV1.';
paciente_hrv = readmatrix(files_HRV(i).name, 'Delimiter',';', 'DecimalSeparator','.');

% Procesado de la señal HRV
s_HRV = eliminar_muestras_perdidas(paciente_hrv, i);
HRV_interp = interp1(s_HRV, 1:(1/25):length(s_HRV), 'spline');

%%% IDENTIFICAR ARCHIVOS EN MATLAB DE LAS SEÑALES DEL ECG %%%
cd      'C:\Users\maria\OneDrive\Documentos\INGENIERIA      BIOMEDICA\TFG
MARIA\REGISTROS ECG\pacientes_ECG'
fprintf('...File: %s (%i/%i)\n', files(i).name(2:end-4), i, size(files,2))
paciente = fopen(files(i).name, "r");
datos = textscan(paciente, '%f%f%f', 'HeaderLines', 3);
col1 = transpose(datos{: ,1})';
col3 = transpose(datos{: ,3})';
tabla = [col1 col3];
fclose(paciente)

Identificadores1(i) = (string(files(i).name(2:end-4))).';
Identificadores = Identificadores1.';

%%% ACORTAR Y GIRAR LA SEÑAL DE ECG (PARA QUE LOS PICOS R QUEDEN HACIA
ARRIBA) %%%
eje_x = tabla(:,1);
eje_y = tabla(:,2);

registro_ECG = tabla(:,2)*-1; % Para dar la vuelta al registro.
eje_xx = eje_x;

eje_xx(1:6000, :) = [];
lim = find(eje_xx > 360000);
eje_xx(lim(:,1):length(eje_xx), :) = [];

registro_ECG(1:60000, :) = [];
lim2 = find(eje_xx > 360000);
registro_ECG(lim2(:,1):length(registro_ECG), :) = [];

%%% OBTENCIÓN DE LA SEÑAL RESPIRATORIA DERIVADA DEL ECG (EDR) %%%
% "np" es la separación entre picos R, la duración del ciclo: 0.25 Hz, 4s.
latidos_segundo = fp(i)/60; % Latidos por segundo.
np = round((latidos_segundo*1000)*0.90); % Muestras por segundo mínimas entre latidos.
[yupper, ylower] = envelope(registro_ECG, np, 'peak');

fs_original = 1000;
fs_objetivo = 25;
p = fs_objetivo/gcd(fs_objetivo, fs_original);
q = fs_original/gcd(fs_objetivo, fs_original);

yupper_interp = resample(yupper, p, q)';

```

---

```

tyupper = (0:length(yupper_interp)-1)/fs_objetivo;
ylower_interp = resample(ylower, p, q)';
tylower = (0:length(ylower_interp)-1)/fs_objetivo;

if length(yupper_interp) > length(HRV_interp)
    yupper_interp(length(HRV_interp)+1:end) = [];
    ylower_interp(length(HRV_interp)+1:end) = [];
else
    HRV_interp(length(yupper_interp)+1:end) = [];
end

figure(3)
plot(eje_xx, registro_ECG)
hold on
plot(eje_xx, yupper, 'linewidth', 1)
hold on
plot(eje_xx, ylower, 'linewidth', 1)
hold off
legend('Registro ECG', 'Envolvente superior', 'Envolvente inferior')

%%% ANÁLISIS EN EL DOMINIO DEL TIEMPO DE LA SEÑAL DE ACR %%%
% Primero se buscan los picos R del registro de ECG.
[picos_envolvente_sup, loc_sup] = findpeaks(yupper);

figure(1)
findpeaks(yupper)
title('Picos en la envolvente superior del ECG')
dist_picos_sup = abs(diff(picos_envolvente_sup));

[picos_envolvente_inf, loc_inf] = findpeaks(ylower);

figure(2)
findpeaks(ylower)
title('Picos en la envolvente inferior del ECG')
dist_picos_inf = abs(diff(picos_envolvente_inf));

% RESPIRATORY SINUS ARRHYTHMIA (RSA)
[variables_sup(i,1), variables_sup(i,2)] = variables_ECG(s_HRV, loc_sup);
[variables_inf(i,1), variables_inf(i,2)] = variables_ECG(s_HRV, loc_inf);

RR_sup = variables_sup(:,1); RR_inf = variables_inf(:,1);
FC_FR_sup = variables_sup(:,2); FC_FR_inf = variables_inf(:,2);

% CORRELACIÓN CRUZADA: medida de la similitud entre un vector x y las
copias desplazadas (desfasadas) de un vector y como función del desfase.
corr_cruzada_sup = corrcoef(yupper_interp, HRV_interp);
corr_cruzada_inf = corrcoef(ylower_interp, HRV_interp);

corr_cruzada_sup = corr_cruzada_sup(1, 2);
corr_cruzada_inf = corr_cruzada_inf(1, 2);

```

```

correlacion_sup = [correlacion_sup corr_cruzada_sup];
correlacion_inf = [correlacion_inf corr_cruzada_inf];

%%% ANÁLISIS EN EL DOMINIO DE LA FRECUENCIA DE LA SEÑAL ACR %%%
% COHERENCIA ESPECTRAL: permite cuantificar el grado de correlación entre
las señales para cada componente frecuencia individual de frecuencia
comunes a las señales.
fs = 25;
window = 30*fs;
noverlap = 0.5*window;
nfft = 128;

[Cxy_sup, f_sup] = mscohere(yupper_interp, HRV_interp, window, noverlap, nfft, fs);
[Cxy_inf, f_inf] = mscohere(ylower_interp, HRV_interp, window, noverlap, nfft, fs);

coherencia_media_sup = mean(Cxy_sup);
coherencia_media_inf = mean(Cxy_inf);

coherencia_sup = [coherencia_sup coherencia_media_sup];
coherencia_inf = [coherencia_inf coherencia_media_inf];

%%% ANÁLISIS EN EL DOMINIO NO LINEAL DE LA SEÑAL ACR %%%
% INFORMACIÓN MUTUA: cuantifica las inter-dependencias lineales y no
lineales entre dos series temporales.
im_sup = mi_cont_cont(yupper_interp, HRV_interp, 5);
im_inf = mi_cont_cont(ylower_interp, HRV_interp, 5);

infom_sup = [infom_sup im_sup];
infom_inf = [infom_inf im_inf];
end

% Una vez extraídas todas las variables, estas son guardadas en una tabla de
Excel.
variables = table(Identificadores, RR_sup, RR_inf, FC_FR_sup, FC_FR_inf,
correlacion_sup', correlacion_inf', coherencia_sup', coherencia_inf',
infom_sup', infom_inf');

filename = 'variablesACOP.xlsx';
writetable(variables,filename,'Sheet',1,'Range','A1')

```

### **Función: variables\_ECG**

```

function [RR, FC_FR] = variables_ECG(registro_HRV, vector_picos)
% Cálculo de la RSA (variable temporal) desde dos enfoques distintos.

vector_pos = round(vector_picos/1000);
inicio = vector_pos(2);
fin = vector_pos(3);
vector_diff_RR = [];

```

---

```

vector_diff_FC_FR = [];

for i = 3:length(vector_pos)-1
    if fin > length(registro_HRV)
        break
    else
        segmento = registro_HRV(inicio:fin);

        %% PROMEDIO DE LA DIFERENCIA ENTRE EL INTERVALO RR MÁS LARGO Y EL
        MÁS CORTO PARA CADA RESPIRACIÓN %%
        vector_dif = diff(segmento);
        max_RR = max(vector_dif);
        min_RR = min(vector_dif);

        diferencia_RR = max_RR - min_RR;
        vector_diff_RR = [vector_diff_RR diferencia_RR];

        %% PROMEDIO DE LAS DIFERENCIAS ENTRE LA FRECUENCIA CARDIACA Y LA
        FRECUENCIA RESPIRATORIA PARA CADA CICLO RESPIRATORIO %%
        FC = 1/mean(segmento);
        FR = length(segmento)/1000;

        diferencia_FC_FR = FC - FR;
        vector_diff_FC_FR = [vector_diff_FC_FR diferencia_FC_FR];

        inicio = vector_pos(i);
        fin = vector_pos(i + 1);
    end
end

RR = mean(vector_diff_RR);
FC_FR = mean(vector_diff_FC_FR);
end

```

### **Función: mi\_cont\_cont**

```

%-----
%---
% mi_cont_cont: mutual information using the first method in [1]
%
% Syntax: mi = mi_cont_cont(x, y, k)
%
% Inputs:
%   x - input vector (length-N)
%   y - input vector (length-N)
%   k - number of nearest neighbours (default 3)
%
% Outputs:
%   mi - mutual information estimate (scalar)
%

```

```

% Example:
%   % set the number of nearest neighbours:
%   k = 3;
%
%   % generate low-frequency random signals:
%   x = randn(1, 1000);
%   y = randn(1, 1000);
%
%   % calculate MI between continuous x and discrete y:
%   mi = mi_cont_cont(x, y, k);
%   fprintf('MI = %g\n', mi);
%
%
% Requires:
%   'knnsearch' from the statistics toolbox
%
%
% [1] Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual
%       information. Physical Review E, 69(6), 16.
% https://doi.org/10.1103/PhysRevE.69.066138
% John M. O' Toole, University College Cork
% Started: 05-08-2020
%
% last update: Time-stamp: <2020-08-12 13:34:38 (otoolej)>
%-----
---
function mi = mi_cont_cont(x, y, k)
    if(nargin < 3 || isempty(k)), k = 5; end
    if(length(x) ~= length(y))
        error('x and y should be the same length');
    end
    N = length(x);
    x = x(:);
    y = y(:);
    xy = [x y];
    %-----
    % 1. calculate the k-th nearest neighbour:
    %-----
    nn_md1x = createns(xy, 'Distance', 'chebychev');
    [~, dist_kth] = knnsearch(nn_md1x, xy, 'k', k + 1);
    dist_kth = dist_kth(:, k + 1);
    %-----
    % 2. find all points within distance:
    % find all the points (all classes) within the k-th distance :
    %-----
    % sort vectors first:
    [xs, idx] = sort(x);
    [ys, idy] = sort(y);
    [~, idx] = sort(idx);
    [~, idy] = sort(idy);
    nx_range = rangesearch_var_radius(xs, idx, dist_kth);
    ny_range = rangesearch_var_radius(ys, idy, dist_kth);

```

```

%-----
% 3. calculate MI (see [1])
%-----
mi = psi(k) + psi(N) - mean(psi(nx_range + 1) + psi(ny_range + 1));
% correct if <0
mi(mi < 0) = 0;
end

```

## ANEXO 4. Código para la selección de variables mediante algoritmos genéticos

### Función: fun\_ga

```

Function [x, score] = fun_ga(MD, Y, Psize, PIni, Pc, Pm, NGen)
% Se aplica la función de algoritmo genético llamando a las funciones y
opciones necesarias para poder ejecutarla. A la salida mostrará las
variables optimas a utilizar.

nvars = size(MD, 2);

% Número de pacientes de la población total.
PopulationSize = Psize;

% Población semilla inicial. Especifica con 1 que variables se escogen para
cada iteración. Máximo de 5 variables por iteración (los 1's en cada fila).
InitialPopulationMatrix = PIni;

% EXPLICACIÓN DE MUTATIONUNIFORM:
% La mutación uniforme es un proceso de dos pasos. Primero, el algoritmo
selecciona una fracción de las entradas de vector de un individuo para la
mutación, donde cada entrada tiene una probabilidad 'rate' de ser mutada.
El valor predeterminado de 'rate' es 0.01. En el 2º paso, el algoritmo
reemplaza cada entrada seleccionada por un número aleatorio seleccionado
uniformemente del rango para esa entrada.
rate_mutation = Pm; % Esta sería la probabilidad de mutación.

% NOTA:
% 1. Según MATLAB cuando el número de filas de la 'InitialPopulationMatrix'
es menor que 'PopulationSize' el algoritmo de 'ga' completará las filas
restantes llamando a CreationFcn.
% 2. 'CrossoverFraction': Fija que porcentaje de la población va a ser
sustituido de antemano. NO ES la probabilidad de cruce perse.

handleplot = @(options, state, flag)myfunplot(options, state, flag);

options = optimoptions('ga', ...
'PopulationSize', PopulationSize, ...
'InitialPopulationMatrix', InitialPopulationMatrix,...
'MutationFcn',{@mutationuniform, rate_mutation}, ...
'MaxGenerations', NGen, ...

```



```

'CrossoverFraction', Pc,...
'PlotFcn', handleplot, ...
'EliteCount', 1);

% Hay que generar las matrices de entrenamiento y test en proporciones de
75% entrenamiento y 25% test Para ello, en primer lugar se usa la función
'cvpartition'.
c = cvpartition(Y, 'HoldOut', 0.25);
idc2 = training(c);

% Se generan las matrices de entrenamiento.
MDtrain2 = MD(idc2 == 1, :);
Ytrain2 = Y(idc2 == 1);

% Se estandarizan los valores de test con el valor de la media y de la
desviación típica de las columnas de la matriz de train.
media = [];
desv = [];

for i = 1:size(MDtrain2, 2)
    media(i) = mean(MDtrain2(:, i));
    desv(i) = std(MDtrain2(:, i));
end

% Se generan y estandarizan las matrices de test.
MDtest = MD(idc2 == 0, :);
MDtest2 = (MDtest - repmat(media, size(MDtest, 1), 1)) ./ (repmat(desv,
size(MDtest, 1), 1));
Ytest2 = Y(idc2 == 0);

% Utilizo el 75% de train para hacer validación cruzada 10 veces.
c2 = cvpartition(Ytrain2, 'KFold', 10);

% Función target a optimizar (acierto del clasificador).
ffun = @(x)fun_fitness(x, MDtrain2, Ytrain2, MDtest2, Ytest2, c2);

A = []; b = []; Aeq = []; beq = []; lb = []; ub = []; nonlcon = [];
[x, score, exitflag, output, pop, scores] = ga(ffun, nvars, A, b, Aeq, beq,
lb, ub, nonlcon, options);
end

```

### **Función: myfunplot**

```

function state = myfunplot(options, state, flag)
% Función para representar el valor de precisión obtenido en cada iteración,
indicando el mejor fitness y la media de las precisiones por cada iteración.

plot(state.Generation, -100*mean(state.Score), 'b*', 'MarkerSize', 5)
hold on
plot(state.Generation, -100*min(state.Score), 'r*', 'MarkerSize', 5)
axis([0 100 0 100])

```

```

titleText = sprintf('Best fitness (%%): %s Mean (%%): %s', num2str(-
100*min(state.Score)), num2str(-100*mean(state.Score)));
title(titleText);

legend('Mean', 'Best fitness')
xlabel('Generaciones')
ylabel('Precisión (%)')
end

```

### **Función: fun\_fitness**

```

function Acc = fun_fitness(x, MDtrain, Ytrain, MDtest, Ytest, c)
% Programación de una función de fitness que devuelva una precisión para
la selección de variables óptimas. Esta función elegirá los hiperparámetros
óptimos.

MDtrain = MDtrain(:, x == 1);
MDtest = MDtest(:, x == 1);

Ytrain_mlp = zeros(length(Ytrain), 3);
Ytrain_mlp(Ytrain == 0, 1) = 1;
Ytrain_mlp(Ytrain == 1, 2) = 1;
Ytrain_mlp(Ytrain == 2, 3) = 1;

% Defino los parámetros de configuración para encontrar los
hiperparámetros.
lambda = [0.01, 0.1, 1]; % Valores para el parámetro de regularización
Nh = [1, 2, 3, 5, 10, 15]; % Valores para el número de capas ocultas

% Realizo la validación cruzada, buscando la mejor configuración para la
RN.
acc_opt = 0;
lambda_opt = 0;
Nh_opt = 0;

for i = 1:length(lambda)
    for j = 1:length(Nh)
        accuracy = zeros(10, 1); % Precisión para cada iteración de
validación cruzada

        for k = 1:10
            % Índices de train y test.
            idc_train = c.training(k);
            idc_test = c.test(k);

            % Conjuntos de train y test.
            MDtrain3 = MDtrain(idc_train, :);
            MDtrain3_std = zscore(MDtrain3);

```

---

```

Ytrain3 = Ytrain(idc_train, :);
Ytrain3_mlp = zeros(length(Ytrain3), 3);
Ytrain3_mlp(Ytrain3 == 0, 1) = 1;
Ytrain3_mlp(Ytrain3 == 1, 2) = 1;
Ytrain3_mlp(Ytrain3 == 2, 3) = 1;

% Se estandarizan los valores de test con el valor de la media
% y de la desviación típica de las columnas de la matriz de train.
media = [];
desv = [];

for m = 1:size(MDtrain3, 2)
    media(m) = mean(MDtrain3(:, m));
    desv(m) = std(MDtrain3(:, m));
end

% Se generan las matrices de test.
MDtest3 = MDtrain(idc_test, :);
size(MDtest3)
MDtest3_std = (MDtest3 - repmat(media, size(MDtest3, 1), 1))
./ (repmat(desv, size(MDtest3, 1), 1));
Ytest3 = Ytrain(idc_test, :);

% Entrenamiento de la red neuronal con la configuración de esta
% interacción.
RN = mlp(size(MDtrain3_std, 2), Nh(j), 3, 'softmax', lambda(i));
RN = mlptrain(RN, MDtrain3_std, Ytrain3_mlp, 100);

% Predicción en el conjunto de prueba.
predict = mlpfwd(RN, MDtest3_std);

% Precisión en el conjunto de prueba.
[~, predict_labels] = max(predict, [], 2);
predict_labels = predict_labels - 1;
accuracy(k) = sum(predict_labels == Ytest3) / length(Ytest3);
end

% Precisión media de la validación cruzada para la configuración
% actual.
mean_acc = mean(accuracy);

% Compruebo si la configuración es mejor que la configuración
% anterior.
if mean_acc > acc_opt
    acc_opt = mean_acc;
    Nh_opt = Nh(j);
    lambda_opt = lambda(i);
end
end
end
end

```

```

% A continuación, entreno la red neuronal utilizando los parámetros óptimos
encontrados.
Acc_v = zeros(1,5);

for p = 1:5
    RN_final = mlp(size(MDtrain, 2), Nh_opt, 3, 'softmax', lambda_opt);
    RN_final = mlptrain(RN_final, zscore(MDtrain), Ytrain_mlp, 100);

    pred = mlpfwd(RN_final, MDtest);

    [~, predict_label] = max(pred, [], 2);
    predict_label = predict_label -1;

    % Calcular la precisión en el conjunto de test.
    Acc_v(p) = -1*sum(predict_label == Ytest) / length(Ytest);
end

Acc = mean(Acc_v);
end

```

## ANEXO 5. Código para la clasificación multiclase mediante una red neuronal multicapa

```

% Primeramente se carga la base de datos donde se encuentran las variables y
el vector con las etiquetas.

load('BBDDRN.mat');
datos = BBDDRN(:, [1:2, 4:50]);
label = BBDDRN(:,3);

% Defino los parámetros de configuración para encontrar los hiperparámetros.
alpha = [0.01, 0.1, 1]; % Valores para el parámetro de regularización
Nh = [1, 2, 3, 5, 10, 15]; % Valores para el número de capas ocultas

% Número de pacientes de cada clase.
n_pacientes_0 = sum(label==0);
n_pacientes_1 = sum(label==1);
n_pacientes_2 = sum(label==2);

disp(['El número de pacientes asociado a la clase 0 es: ',
num2str(n_pacientes_0)]); %25
disp(['El número de pacientes asociado a la clase 1 es: ',
num2str(n_pacientes_1)]); %25
disp(['El número de pacientes asociado a la clase 2 es: ',
num2str(n_pacientes_2)]); %33

% Divido los datos en conjuntos de entrenamiento (train) y de prueba (test)
mediante la estrategia HoldOut.

```

---

```

val_cruz = cvpartition(label, 'HoldOut', 0.25);
idc = training(val_cruz);

% Se generan las matrices de entrenamiento y las de test, con sus
correspondientes etiquetas.
MDtrain1 = datos(idc == 1, :);
Ytrain1 = label(idc == 1);

Ytrain1_mlp = zeros(length(Ytrain1), 3);
Ytrain1_mlp(Ytrain1 == 0, 1) = 1;
Ytrain1_mlp(Ytrain1 == 1, 2) = 1;
Ytrain1_mlp(Ytrain1 == 2, 3) = 1;

MDtest1 = datos(idc == 0, :);
Ytest1 = label(idc == 0);

% Se estandarizan los valores de test con el valor de la media y de la desviación
típica de las columnas de la matriz de train.
media = [];
desv = [];

for i = 1:size(MDtrain1, 2)
    media(i) = mean(MDtrain1(:, i));
    desv(i) = std(MDtrain1(:, i));
end

MDtest1_std = (MDtest1 - repmat(media, size(MDtest1, 1), 1)) ./ (repmat(desv,
size(MDtest1, 1), 1));

% Para la población escojo un valor aleatorio entre 25 y 30.
Psize = ceil(25 + rand*5);

% Establezco el número de variables independientes del modelo.
GenomeLength = size(datos, 2);

% Establezco la matriz semilla para el modelo de optimización.
PIni = mi_pob_inic(GenomeLength, [], [], Psize);

% El número máximo de generaciones por defecto es 100.
% - La opción 'MaxGenerations' determina el número máximo de generaciones que
toma el algoritmo genético. Incrementar 'MaxGenerations' puede mejorar.
% el resultado final.
% - La opción 'MaxStallGenerations' controla el número de pasos que 'ga'
revisa para ver si está progresando. El aumento 'MaxStallGenerations' puede
permitir a 'ga' continuar cuando el algoritmo necesita más evaluaciones de
funciones para encontrar una mejor solución. Por defecto este último valor vale
50.
NGen = 100;
[x, score] = fun_ga(MDtrain1, Ytrain1, Psize, PIni, 0.8, 0.01, NGen);

% Estas serían las variables seleccionadas como óptimas por el algoritmo
% genético:

```

```

find(x)

% Se seleccionan en las matrices de train y de test solo las variables óptimas.
MDtrain_opt = MDtrain1(:, x == 1);
MDtest_std_opt = MDtest1_std(:, x == 1);

% Determino los hiperparámetros óptimos que se van a utilizar. Para ello se
deja fijo el valor de lambda y se van variando el número de nodos de la capa
oculta. Para ello utilizo el 75% de mis instancias (es decir, mi matriz de
train). Se utiliza la estrategia Leave-one-out.
ch = cvpartition(Ytrain1, 'LeaveOut');
num_iteraciones = ch.NumTestSets;

% Se define la matriz que va a recoger los valores de precisión para combinación
de Nh y de alpha.
Acc_matrix = zeros(length(alpha), length(Nh));
% Realizo la validación cruzada, buscando la mejor configuración para la RN.
for i = 1:length(alpha)
    for j = 1:length(Nh)
        predict = zeros(num_iteraciones, 3); % Precisión para cada iteración
                                                de validación cruzada.

        for k = 1:num_iteraciones
            % Índices de train y test
            idc2 = training(ch, k);
            % Conjuntos de train y test
            MDtrain4 = MDtrain1(idc2 == 1, :);
            MDtrain4_std = zscore(MDtrain4);
            Ytrain4 = Ytrain1(idc2 == 1, :);
            Ytrain4_mlp = zeros(length(Ytrain4), 3);
            Ytrain4_mlp(Ytrain4 == 0, 1) = 1;
            Ytrain4_mlp(Ytrain4 == 1, 2) = 1;
            Ytrain4_mlp(Ytrain4 == 2, 3) = 1;

            % Se estandarizan los valores de test con el valor de la media y
            de la desviación típica de las columnas de la matriz de train.
            media = [];
            desv = [];

            for m = 1:size(MDtrain4, 2)
                media(m) = mean(MDtrain4(:, m));
                desv(m) = std(MDtrain4(:, m));
            end

            % Se generan las matrices de test.
            MDtest4 = MDtrain1(idc2 == 0, :);
            MDtest4_std = (MDtest4 - repmat(media, size(MDtest4, 1), 1)) ./
            (repmat(desv, size(MDtest4, 1), 1));

            Ytest4 = Ytrain1(idc2 == 0, :);

            % Entrenamiento de la red neuronal con la configuración de esta
            interacción.

```

---

```

RN = mlp(size(MDtrain4_std, 2), Nh(j), 3, 'softmax', lambda(i));
RN = mlptrain(RN, MDtrain4_std, Ytrain4_mlp, 100);

% Predicción en el conjunto de prueba
predict(k, :) = mlpfwd(RN, MDtest4_std);
end

[~, predict_label] = max(predict, [], 2);
predict_label = predict_label -1;

% Precisión media de la validación cruzada para la configuración actual
Acc_matrix(i, j) = sum(Ytest4 == predict_label) /
length(predict_label);
end
end

Acc_matrix

[M1, I1] = max(Acc_matrix);
[M2, I2] = max(M1);

i_alpha = I1(I2);
j_Nh = I2;

% A continuación, entreno la red neuronal utilizando los parámetros óptimos
encontrados.
RN_final = mlp(size(MDtrain_opt, 2), Nh(j_Nh), 3, 'softmax', alpha(i_alpha));
RN_final = mlptrain(RN_final, zscore(MDtrain_opt), Ytrain1_mlp, 100);

nombreArchivo = 'redNeuronalEntrenada.mat';

% Guardar la red neuronal en el archivo.
save(nombreArchivo, 'RN_final');

pred = mlpfwd(RN_final, MDtest_std_opt);
[~, predict_label] = max(pred, [], 2);
predict_label = predict_label -1;

% Mostrar los resultados.
disp('Etiquetas predichas:'); disp(predict_label);

% Matriz de confusión multiclase.
MC = confusionmat(Ytest1, predict_label);
labels = {'Clase 0', 'Clase 1', 'Clase2'};

figure(2)
confusionchart(MC, labels);

% Matriz confusión de 2 clases (No COVID vs. (COVID + SPC-CP))
MC1 = [MC(1,1) (MC(1,2) + MC(1,3))];

```

```

        (MC(2,1) + MC(3,1)) (MC(2,2) + MC(2,3) + MC(3,2) + MC(3,3))];
labels = {'Clase 0', 'Clase 1'};

figure(3)
confusionchart(MC1, labels);
Se1 = MC1(1,1) / sum(MC1(1,:)); % sensibilidad
Sp1 = MC1(2,2) / sum(MC1(2,:)); % especificidad
PPV1 = MC1(1,1) / sum(MC1(:,1)); % valor predictivo positivo
NPV1 = MC1(2,2) / sum(MC1(:,2)); % valor predictivo negativo
PLR1 = Se1 / (1 - Sp1); % razón de verosimilitud positiva
NLR1 = (1 - Se1) / Sp1; % razón de verosimilitud positiva
Acc1 = (MC1(1,1) + MC1(2,2))/sum(sum(MC1));

valores1 = [Se1 Sp1 PPV1 NPV1 PLR1 NLR1 Acc1]

% Matriz confusión de 2 clases ((No COVID + COVID) vs. SPC-CP)
MC2 = [(MC(1,1) + MC(1,2) + MC(2,1) + MC(2,2)) (MC(1,3) + MC(2,3));
        (MC(3,1) + MC(3,2)) MC(3,3)];
labels = {'Clase 0', 'Clase 1'};

figure(4)
confusionchart(MC2, labels);

Se2 = MC2(1,1) / sum(MC2(1,:)); % sensibilidad
Sp2 = MC2(2,2) / sum(MC2(2,:)); % especificidad
PPV2 = MC2(1,1) / sum(MC2(:,1)); % valor predictivo positivo
NPV2 = MC2(2,2) / sum(MC2(:,2)); % valor predictivo negativo
PLR2 = Se2 / (1 - Sp2); % razón de verosimilitud positiva
NLR2 = (1 - Se2) / Sp2; % razón de verosimilitud positiva
Acc2 = (MC2(1,1) + MC2(2,2))/sum(sum(MC2));
valores2 = [Se2 Sp2 PPV2 NPV2 PLR2 NLR2 Acc2]

% Parámetros necesarios para calcular el MACRO F1 SCORE (MF1).
F1_0 = 2*((MC(1,1)/sum(MC(1,:)))*(MC(1,1)/sum(MC(:,1)))) /
((MC(1,1)/sum(MC(1,:)) + (MC(1,1)/sum(MC(:,1)))));
F1_1 = 2*((MC(2,2)/sum(MC(2,:)))*(MC(2,2)/sum(MC(:,2)))) /
((MC(2,2)/sum(MC(2,:)) + (MC(2,2)/sum(MC(:,2)))));
F1_2 = 2*((MC(3,3)/sum(MC(3,:)))*(MC(3,3)/sum(MC(:,3)))) /
((MC(3,3)/sum(MC(3,:)) + (MC(3,3)/sum(MC(:,3)))));

% Calcular la precisión en el conjunto de test mediante distintos índices.
Acc = sum(predict_label == Ytest1) / length(Ytest1);
Kappa = kappa(MC);
MF1 = (1/3)*(F1_0 + F1_1 + F1_2);

```



**Función: mi\_pob\_inic**

```
function PI = mi_pob_inic(GenomeLength, FitnessFcn, options, Psize)
    % Se establece el tamaño de la población inicial, indicando el número de
    % posibles variables en cada potencias solución.
    PI = zeros(Psize, GenomeLength); % Psize: filas

    vector_sem=[ones(1,5) zeros(1,44)];

    for i = 1:Psize
        vector = vector_sem(randperm(length(vector_sem)));
        PI(i,:) = vector;
    end
end
```

**Función: kappa**

```
function k=kappa(varargin)

% KAPPA: This function computes the Cohen's kappa coefficient.
% Cohen's kappa coefficient is a statistical measure of inter-rater
% reliability. It is generally thought to be a more robust measure than
% simple percent agreement calculation since k takes into account the
% agreement occurring by chance.
% Kappa provides a measure of the degree to which two judges, A and B,
% concur in their respective sortings of N items into k mutually exclusive
% categories. A 'judge' in this context can be an individual human being, a
% set of individuals who sort the N items collectively, or some non-human
% agency, such as a computer program or diagnostic test, that performs a
% sorting on the basis of specified criteria.
% The original and simplest version of kappa is the unweighted kappa
% coefficient introduced by J. Cohen in 1960. When the categories are
% merely nominal, Cohen's simple unweighted coefficient is the only form of
% kappa that can meaningfully be used. If the categories are ordinal and if
% it is the case that category 2 represents more of something than category
% 1, that category 3 represents more of that same something than category
% 2, and so on, then it is potentially meaningful to take this into
% account, weighting each cell of the matrix in accordance with how near it
% is to the cell in that row that includes the absolutely concordant items.
% This function can compute a linear weights or a quadratic weights.
%
% Syntax:    kappa(X,W,ALPHA)
%
% Inputs:
%     X - square data matrix
%     W - Weight (0 = unweighted; 1 = linear weighted; 2 = quadratic
%         weighted; -1 = display all. Default=0)
%     ALPHA - default=0.05.
%
% Outputs:
%     - Observed agreement percentage
```

---

```

%           - Random agreement percentage
%           - Agreement percentage due to true concordance
%           - Residual not random agreement percentage
%           - Cohen's kappa
%           - kappa error
%           - kappa confidence interval
%           - Maximum possible kappa
%           - k observed as proportion of maximum possible
%           - k benchmarks by Landis and Koch
%           - z test results
%
% Example:
%
%           x=[88 14 18; 10 40 10; 2 6 12];
%
%           Calling on Matlab the function: kappa(x)
%
%           Answer is:
%
% UNWEIGHTED COHEN'S KAPPA
% -----
% -----
% Observed agreement (po) = 0.7000
% Random agreement (pe) = 0.4100
% Agreement due to true concordance (po-pe) = 0.2900
% Residual not random agreement (1-pe) = 0.5900
% Cohen's kappa = 0.4915
% kappa error = 0.0549
% kappa C.I. (alpha = 0.0500) = 0.3839      0.5992
% Maximum possible kappa, given the observed marginal frequencies = 0.8305
% k observed as proportion of maximum possible = 0.5918
% Moderate agreement
% Variance = 0.0031      z (k/sqrt(var)) = 8.8347      p = 0.0000
% Reject null hypothesis: observed agreement is not accidental
%
%           Created by Giuseppe Cardillo
%           giuseppe.cardillo-edta@poste.it
%
% To cite this file, this would be an appropriate format:
% Cardillo G. (2007) Cohens kappa: compute the Cohen's kappa ratio on a 2x2
matrix.
% http://www.mathworks.com/matlabcentral/fileexchange/15365

%Input Error handling
global m f x w alpha

args=cell(varargin);
nu=numel(args);
if isempty(nu)
    error('Warning: Matrix of data is missed...')
elseif nu>3
    error('Warning: Max three input data are required')

```

---

```

end
default.values = {[],0,0.05};
default.values(1:nu) = args;
[x w alpha] = deal(default.values{:});

if isempty(x)
    error('Warning: X matrix is empty...')
end
if isvector(x)
    error('Warning: X must be a matrix not a vector')
end
if ~all(isfinite(x(:))) || ~all(isnumeric(x(:)))
    error('Warning: all X values must be numeric and finite')
end
if ~isequal(x(:),round(x(:)))
    error('Warning: X data matrix values must be whole numbers')
end

m=size(x);
if ~isequal(m(1),m(2))
    error('Input matrix must be a square matrix')
end
if nu>1 %eventually check weight
    if ~isscalar(w) || ~isfinite(w) || ~isnumeric(w) || isempty(w)
        error('Warning: it is required a scalar, numeric and finite Weight
value.')
    end
    a=-1:1:2;
    if isempty(a(a==w))%check if w is -1 0 1 or 2
        error('Warning: Weight must be -1 0 1 or 2.')
    end
end
if nu>2 %eventually check alpha
    if ~isscalar(alpha) || ~isnumeric(alpha) || ~isfinite(alpha) ||
isempty(alpha)
        error('Warning: it is required a numeric, finite and scalar ALPHA
value.');
    end
    if alpha <= 0 || alpha >= 1 %check if alpha is between 0 and 1
        error('Warning: ALPHA must be comprised between 0 and 1.')
    end
end
clear args default nu

m(2)=[];
tr= repmat('-',1,80);
if w==0 || w==-1
    f=diag(ones(1,m)); %unweighted
    disp('UNWEIGHTED COHEN'S KAPPA')
    disp(tr)
    k=kcomp;
    disp(' ')

```

```

end
if w==1 || w==-1
    J= repmat((1:1:m),m,1);
    I= flipud(rot90(J));
    f= 1-abs(I-J)./(m-1); %linear weight
    disp('LINEAR WEIGHTED COHEN'S KAPPA')
    disp(tr)
    k=kcomp;
    disp(' ')
end
if w==2 || w==-1
    J= repmat((1:1:m),m,1);
    I= flipud(rot90(J));
    f= 1-((I-J)./(m-1)).^2; %quadratic weight
    disp('QUADRATIC WEIGHTED COHEN'S KAPPA')
    disp(tr)
    k=kcomp;
end
return
end

function k=kcomp
global m f x alpha
n=sum(x(:)); %Sum of Matrix elements
x=x./n; %proportion
r=sum(x,2); %rows sum
s=sum(x); %columns sum
Ex=r*s; %expected proportion for random agree
pom=sum(min([r';s]));
po=sum(sum(x.*f));
pe=sum(sum(Ex.*f));
k=(po-pe)/(1-pe);
km=(pom-pe)/(1-pe); %maximum possible kappa, given the observed marginal
frequencies
ratio=k/km; %observed as proportion of maximum possible
sek=sqrt((po*(1-po))/(n*(1-pe)^2)); %kappa standard error for confidence
interval
ci=k+([-1 1].*(abs(-realsqrt(2)*erfcinv(alpha))*sek)); %k confidence interval
wbari=r'*f;
wbarj=s*f;
wbar= repmat(wbari',1,m)+repmat(wbarj,m,1);
a=Ex.*((f-wbar).^2);
var=(sum(a(:))-pe^2)/(n*((1-pe)^2)); %variance
z=k/sqrt(var); %normalized kappa
p=(1-0.5*erfc(-abs(z)/realsqrt(2)))*2;
%display results
fprintf('Observed agreement (po) = %0.4f\n',po)
fprintf('Random agreement (pe) = %0.4f\n',pe)
fprintf('Agreement due to true concordance (po-pe) = %0.4f\n',po-pe)
fprintf('Residual not random agreement (1-pe) = %0.4f\n',1-pe)
fprintf('Cohen's kappa = %0.4f\n',k)
fprintf('kappa error = %0.4f\n',sek)

```

---

```
fprintf('kappa C.I. (alpha = %0.4f) = %0.4f    %0.4f\n',alpha,ci)
fprintf('Maximum possible kappa, given the observed marginal frequencies =
%0.4f\n',km)
fprintf('k observed as proportion of maximum possible = %0.4f\n',ratio)
if k<0
    disp('Poor agreement')
elseif k>=0 && k<=0.2
    disp('Slight agreement')
elseif k>=0.21 && k<=0.4
    disp('Fair agreement')
elseif k>=0.41 && k<=0.6
    disp('Moderate agreement')
elseif k>=0.61 && k<=0.8
    disp('Substantial agreement')
elseif k>=0.81 && k<=1
    disp('Perfect agreement')
end
fprintf('Variance = %0.4f    z (k/sqrt(var)) = %0.4f    p = %0.4f\n',var,z,p)
if p<0.05
    disp('Reject null hypothesis: observed agreement is not accidental')
else
    disp('Accept null hypothesis: observed agreement is accidental')
end
return
end
```



---

# BIBLIOGRAFÍA

- [1] Chen Y, Klein SL, Garibaldi BT, Li H, Wu C, Osevala NM, et al. Aging in COVID-19: Vulnerability, immunity and intervention. *Ageing Res Rev* 2021;65:101205. <https://doi.org/10.1016/j.arr.2020.101205>.
- [2] Fung TS, Liu DX. Human Coronavirus: Host-pathogen interaction. *Annu Rev Microbiol* 2019;73:529–57. <https://doi.org/10.1146/annurev-micro-020518-115759>.
- [3] Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ. Nidovirales: evolving the largest RNA virus genome. *Virus Res* 2006;117:17–37. <https://doi.org/10.1016/j.virusres.2006.01.017>.
- [4] Makarov V, Riabova O, Ekins S, Pluzhnikov N, Chepur S. The past, present and future of RNA respiratory viruses: influenza and coronaviruses. *Pathog Dis* 2020;78:ftaa046. <https://doi.org/10.1093/femspd/ftaa046>.
- [5] Ramadan N, Shaib H. Middle East respiratory syndrome coronavirus (MERS-CoV): A review. *Germes* 2019;9:35–42. <https://doi.org/10.18683/germs.2019.1155>.
- [6] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020;382:727–33. <https://doi.org/10.1056/NEJMoa2001017>.
- [7] Pavli A, Theodoridou M, Maltezou HC. Post-COVID syndrome: Incidence, clinical spectrum, and challenges for primary healthcare professionals. *Arch Med Res* 2021;52:575–81. <https://doi.org/10.1016/j.arcmed.2021.03.010>.
- [8] Fernández-de-Las-Peñas C, Palacios-Ceña D, Gómez-Mayordomo V, Cuadrado ML, Florencio LL. Defining post-COVID symptoms (post-Acute COVID, long COVID, persistent post-COVID): An integrative classification. *Int J Environ Res Public Health* 2021;18:2621. <https://doi.org/10.3390/ijerph18052621>.
- [9] Salamanna F, Veronesi F, Martini L, Landini MP, Fini M. Post-COVID-19 Syndrome: The persistent symptoms at the post-viral stage of the disease. A systematic review of the current data. *Front Med* 2021;8:653516. <https://doi.org/10.3389/fmed.2021.653516>.

- 
- [10] Oronsky B, Larson C, Hammond TC, Oronsky A, Kesari S, Lybeck M, et al. A review of persistent post-COVID syndrome (PPCS). *Clin Rev Allergy Immunol* 2023;64:66–74. <https://doi.org/10.1007/s12016-021-08848-3>.
- [11] Asarcikli LD, Hayiroglu Mİ, Osken A, Keskin K, Kolak Z, Aksu T. Heart rate variability and cardiac autonomic functions in post-COVID period. *J Interv Card Electrophysiol Int J Arrhythm Pacing* 2022;63:715–21. <https://doi.org/10.1007/s10840-022-01138-8>.
- [12] Salem AM, Yar T, Al Eid M, Almahfoudh H, Alsaffar M, Al Ibrahim A, et al. Post-acute effect of SARS-CoV-2 infection on the cardiac autonomic function. *Int J Gen Med* 2022;15:7593–603. <https://doi.org/10.2147/IJGM.S382331>.
- [13] Kaliyaperumal D, Rk K, Alagesan M, Ramalingam S. Characterization of cardiac autonomic function in COVID-19 using heart rate variability: a hospital based preliminary observational study. *J Basic Clin Physiol Pharmacol* 2021;32:247–53. <https://doi.org/10.1515/jbcpp-2020-0378>.
- [14] Blitshteyn S, Whitelaw S. Postural orthostatic tachycardia syndrome (POTS) and other autonomic disorders after COVID-19 infection: a case series of 20 patients. *Immunol Res* 2021;69:205–11. <https://doi.org/10.1007/s12026-021-09185-5>.
- [15] Kramer O. Genetic Algorithms. In: Kramer O, editor. *Genet. Algorithm essent.*, Cham: Springer International Publishing; 2017, p. 11–9. [https://doi.org/10.1007/978-3-319-52156-5\\_2](https://doi.org/10.1007/978-3-319-52156-5_2).
- [16] Bishop C. *Neural networks for pattern recognition*, 1995.
- [17] Crook H, Raza S, Nowell J, Young M, Edison P. Long covid-mechanisms, risk factors, and management. *BMJ* 2021;374:n1648. <https://doi.org/10.1136/bmj.n1648>.
- [18] Lledó GM, Sellares J, Brotons C, Sans M, Antón JD, Blanco J, et al. Post-acute COVID-19 syndrome: a new tsunami requiring a universal case definition. *Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis* 2022;28:315–8. <https://doi.org/10.1016/j.cmi.2021.11.015>.
- [19] Carod-Artal FJ. Post-COVID-19 syndrome: epidemiology, diagnostic criteria and pathogenic mechanisms involved. *Rev Neurol* 2021;72:384–96. <https://doi.org/10.33588/rn.7211.2021230>.
- [20] Raveendran AV, Jayadevan R, Sashidharan S. Long COVID: An overview. *Diabetes Metab Syndr* 2021;15:869–75. <https://doi.org/10.1016/j.dsx.2021.04.007>.
- [21] Colafrancesco S, Alessandri C, Conti F, Priori R. COVID-19 gone bad: A new character in the spectrum of the hyperferritinemic syndrome? *Autoimmun Rev* 2020;19:102573. <https://doi.org/10.1016/j.autrev.2020.102573>.
- [22] Biehl M, Sese D. Post-intensive care syndrome and COVID-19 - Implications post pandemic. *Cleve Clin J Med* 2020. <https://doi.org/10.3949/ccjm.87a.ccc055>.



- 
- [23] Chippa V, Aleem A, Anjum F. Post acute coronavirus (COVID-19) syndrome. StatPearls, Treasure Island (FL): StatPearls Publishing; 2023.
- [24] Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and predictors of long COVID. *Nat Med* 2021;27:626–31. <https://doi.org/10.1038/s41591-021-01292-y>.
- [25] López-Sampalo A, Bernal-López MR, Gómez-Huelgas R. Síndrome de COVID-19 persistente. Una revisión narrativa. *Rev Clínica Esp* 2022;222:241–50. <https://doi.org/10.1016/j.rce.2021.10.003>.
- [26] González-Calle D, Eiros R, Sánchez PL. Corazón y SARS-CoV-2. *Med Clínica* 2022;159:440–6. <https://doi.org/10.1016/j.medcli.2022.07.002>.
- [27] Arnold DT, Milne A, Samms E, Staddon L, Maskell NA, Hamilton FW. Are vaccines safe in patients with Long COVID? A prospective observational study 2021:2021.03.11.21253225. <https://doi.org/10.1101/2021.03.11.21253225>.
- [28] Sykes DL, Holdsworth L, Jawad N, Gunasekera P, Morice AH, Crooks MG. Post-COVID-19 Symptom Burden: What is Long-COVID and how should we manage it? *Lung* 2021;199:113–9. <https://doi.org/10.1007/s00408-021-00423-z>.
- [29] Boutou AK, Asimakos A, Kortianou E, Vogiatzis I, Tzouveleakis A. Long COVID-19 pulmonary sequelae and management considerations. *J Pers Med* 2021;11:838. <https://doi.org/10.3390/jpm11090838>.
- [30] Aragón-Benedí C, Oliver-Forniés P, Galluccio F, Yamak Altinpulluk E, Ergonenc T, El Sayed Allam A, et al. Is the heart rate variability monitoring using the analgesia nociception index a predictor of illness severity and mortality in critically ill patients with COVID-19? A pilot study. *PloS One* 2021;16:e0249128. <https://doi.org/10.1371/journal.pone.0249128>.
- [31] Mol MBA, Strous MTA, van Osch FHM, Vogelaar FJ, Barten DG, Farchi M, et al. Heart-rate-variability (HRV), predicts outcomes in COVID-19. *PloS One* 2021;16:e0258841. <https://doi.org/10.1371/journal.pone.0258841>.
- [32] Marques KC, Silva CC, Trindade S da S, Santos MC de S, Rocha RSB, Vasconcelos PF da C, et al. Reduction of cardiac autonomic modulation and increased sympathetic activity by heart rate variability in patients with Long COVID. *Front Cardiovasc Med* 2022;9:862001. <https://doi.org/10.3389/fcvm.2022.862001>.
- [33] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet Lond Engl* 2020;395:507–13. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7).

- 
- [34] Serrano-Cumplido A, Trillo Calvo E, García Matarín L, del Río Herrero A, Gamir Ruiz FJ, Molina Escribano F, et al. Pulsioximetría: papel en el paciente COVID-19 domiciliario. *Semergen* 2022;48:70–7. <https://doi.org/10.1016/j.semerg.2021.03.004>.
- [35] Schulz S, Adochiei F-C, Edu I-R, Schroeder R, Costin H, Bär K-J, et al. Cardiovascular and cardiorespiratory coupling analyses: a review. *Philos Transact A Math Phys Eng Sci* 2013;371:20120191. <https://doi.org/10.1098/rsta.2012.0191>.
- [36] Bona Olexova L, Sekaninova N, Jurko A, Visnovcova Z, Grendar M, Jurko T, et al. Respiratory sinus arrhythmia as an Index of cardiac vagal Control in Mitral Valve Prolapse. *Physiol Res* 2020;69:S163–9. <https://doi.org/10.33549/physiolres.934402>.
- [37] Natori H, Kawayama T, Suetomo M, Kinoshita T, Matsuoka M, Matsunaga K, et al. Evaluation of the modified medical research council dyspnea scale for predicting hospitalization and exacerbation in Japanese patients with chronic obstructive pulmonary disease. *Intern Med Tokyo Jpn* 2016;55:15–24. <https://doi.org/10.2169/internalmedicine.55.4490>.
- [38] Jones PW, Harding G, Berry P, Wiklund I, Chen W-H, Kline Leidy N. Development and first validation of the COPD Assessment Test. *Eur Respir J* 2009;34:648–54. <https://doi.org/10.1183/09031936.00102509>.
- [39] Serhani MA, T El Kassabi H, Ismail H, Nujum Navaz A. ECG Monitoring Systems: Review, architecture, processes, and key challenges. *Sensors* 2020;20:1796. <https://doi.org/10.3390/s20061796>.
- [40] Shaffer F, Ginsberg JP. An Overview of heart rate variability metrics and norms. *Front Public Health* 2017;5:258. <https://doi.org/10.3389/fpubh.2017.00258>.
- [41] McCraty R, Shaffer F. Heart Rate Variability: New perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk. *Glob Adv Health Med* 2015;4:46–61. <https://doi.org/10.7453/gahmj.2014.073>.
- [42] Belenguer Muncharaz A, Bisbal E, Reig Valero R, Mas Herrero S, Carregui R, Abizanda Campos R. Relación entre pulsioximetría y determinación de la saturación arterial de oxígeno. Influencia de los fármacos vasoactivos presores sobre la correlación SattcO2-SatO2. *Med Intensiva* 2001;25:333–9.
- [43] Hafen BB, Sharma S. Oxygen Saturation. StatPearls, Treasure Island (FL): StatPearls Publishing; 2023.
- [44] Netzer N, Eliasson A, Netzer C, Kristo D. Overnight pulse oximetry for sleep-disordered breathing in adults - A review. *Chest* 2001;120:625–33. <https://doi.org/10.1378/chest.120.2.625>.

- 
- [45] Barroso-García V, Gutiérrez-Tobal G, Gozal L, Alvarez D, Vaquerizo-Villar F, del Campo F, et al. Análisis del flujo aéreo nocturno mediante wavelets para la ayuda en el diagnóstico de la apnea del sueño infantil. 2020.
- [46] Penzel T, Kantelhardt JW, Grote L, Peter J-H, Bunde A. Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea. *IEEE Trans Biomed Eng* 2003;50:1143–51. <https://doi.org/10.1109/TBME.2003.817636>.
- [47] Montano N, Ruscone TG, Porta A, Lombardi F, Pagani M, Malliani A. Power spectrum analysis of heart rate variability to assess the changes in sympathovagal balance during graded orthostatic tilt. *Circulation* 1994;90:1826–31. <https://doi.org/10.1161/01.cir.90.4.1826>.
- [48] Léonard A, Clément S, Kuo C-D, Manto M. Changes in Heart Rate Variability During Heartfulness Meditation: A power spectral analysis including the residual spectrum. *Front Cardiovasc Med* 2019;6:62. <https://doi.org/10.3389/fcvm.2019.00062>.
- [49] Ubeyli ED, Güler I. Comparison of eigenvector methods with classical and model-based methods in analysis of internal carotid arterial Doppler signals. *Comput Biol Med* 2003;33:473–93. [https://doi.org/10.1016/s0010-4825\(03\)00021-0](https://doi.org/10.1016/s0010-4825(03)00021-0).
- [50] Welch P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans Audio Electroacoustics* 1967;15:70–3. <https://doi.org/10.1109/TAU.1967.1161901>.
- [51] Alvarez D, Hornero R, Marcos JV, Wessel N, Penzel T, Glos M, et al. Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of apnea diagnosis. *Int J Neural Syst* 2013;23:1350020. <https://doi.org/10.1142/S0129065713500202>.
- [52] Poza J, Hornero R, Abásolo D, Fernández A, García M. Extraction of spectral based measures from MEG background oscillations in Alzheimer’s disease. *Med Eng Phys* 2007;29:1073–83. <https://doi.org/10.1016/j.medengphy.2006.11.006>.
- [53] Alvarez D, Gutiérrez-Tobal G, Alonso-Alvarez ML, Terán J, del Campo F, Hornero R. Análisis espectral y no lineal de la señal de oximetría domiciliaria en la ayuda al diagnóstico de la apnea infantil. 2014.
- [54] Richman J, Moorman J. Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol* 2000;278:H2039-49. <https://doi.org/10.1152/ajpheart.2000.278.6.H2039>.
- [55] Alvarez D, Hornero R, Marcos JV, del Campo F. Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis. *IEEE Trans Biomed Eng* 2010;57:2816–24. <https://doi.org/10.1109/TBME.2010.2056924>.

- 
- [56] Cohen ME, Hudson DL, Deedwania PC. Applying continuous chaotic modeling to cardiac signal analysis. *IEEE Eng Med Biol Mag* 1996;15:97–102. <https://doi.org/10.1109/51.537065>.
- [57] Lempel A, Ziv J. On the complexity of finite sequences. *IEEE Trans Inf Theory* 1976;22:75–81. <https://doi.org/10.1109/TIT.1976.1055501>.
- [58] Aboy M, Hornero R, Abasolo D, Alvarez D. Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Trans Biomed Eng* 2006;53:2282–8. <https://doi.org/10.1109/TBME.2006.883696>.
- [59] Shumway RH, Stoffer DS. Characteristics of time series. In: Shumway RH, Stoffer DS, editors. *Time Ser. Anal. Its Appl. R Ex.*, Cham: Springer International Publishing; 2017, p. 1–44. [https://doi.org/10.1007/978-3-319-52452-8\\_1](https://doi.org/10.1007/978-3-319-52452-8_1).
- [60] Nolte G, Bai O, Wheaton L, Mari Z, Vorbach S, Hallett M. Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin Neurophysiol* 2004;115:2292–307. <https://doi.org/10.1016/j.clinph.2004.04.029>.
- [61] Montano N, Porta A, Cogliati C, Costantino G, Tobaldini E, Casali KR, et al. Heart rate variability explored in the frequency domain: A tool to investigate the link between heart and behavior. *Neurosci Biobehav Rev* 2009;33:71–80. <https://doi.org/10.1016/j.neubiorev.2008.07.006>.
- [62] Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Phys Rev E Stat Nonlin Soft Matter Phys* 2004;69:066138. <https://doi.org/10.1103/PhysRevE.69.066138>.
- [63] Stein PK, Pu Y. Heart rate variability, sleep and sleep disorders. *Sleep Med Rev* 2012;16:47–66. <https://doi.org/10.1016/j.smr.2011.02.005>.
- [64] Hornero R, Alvarez D, Abásolo D, del Campo F, Zamarrón C. Utility of approximate entropy from overnight pulse oximetry data in the diagnosis of the obstructive sleep apnea syndrome. *IEEE Trans Biomed Eng* 2007;54:107–13. <https://doi.org/10.1109/TBME.2006.883821>.
- [65] Garcia AJ, Koschnitzky JE, Dashevskiy T, Ramirez J-M. Cardiorespiratory coupling in health and disease. *Auton Neurosci Basic Clin* 2013;175:26–37. <https://doi.org/10.1016/j.autneu.2013.02.006>.
- [66] Vaquerizo-Villar F, Álvarez D, Kheirandish-Gozal L, Gutiérrez-Tobal GC, Barroso-García V, Crespo A, et al. Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings. *Comput Methods Programs Biomed* 2018;156:141–9. <https://doi.org/10.1016/j.cmpb.2017.12.020>.
- [67] Jacak W, Pröll K, Winkler S. Neural Networks Based Feature Selection in Biological Data Analysis. In: Klemm R, Nikodem J, Jacak W, Chaczko Z, editors. *Adv. Methods Appl.*

- Comput. Intell., Heidelberg: Springer International Publishing; 2014, p. 79–94. [https://doi.org/10.1007/978-3-319-01436-4\\_5](https://doi.org/10.1007/978-3-319-01436-4_5).
- [68] Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* 2003;3:1157–82.
- [69] Yang D, Yu Z, Yuan H, Cui Y. An improved genetic algorithm and its application in neural network adversarial attack. *PLoS ONE* 2022;17:e0267970. <https://doi.org/10.1371/journal.pone.0267970>.
- [70] Álvarez D, Hornero R, Marcos JV, del Campo F. Feature selection from nocturnal oximetry using genetic algorithms to assist in obstructive sleep apnoea diagnosis. *Med Eng Phys* 2012;34:1049–57. <https://doi.org/10.1016/j.medengphy.2011.11.009>.
- [71] Chan Y, Walmsley RP. Learning and understanding the Kruskal-Wallis one-way analysis-of-variance-by-ranks test for differences among three or more independent groups. *Phys Ther* 1997;77:1755–62. <https://doi.org/10.1093/ptj/77.12.1755>.
- [72] Rácz A, Bajusz D, Héberger K. Multi-Level Comparison of Machine Learning Classifiers and Their Performance Metrics. *Molecules* 2019;24:2811. <https://doi.org/10.3390/molecules24152811>.
- [73] Abraira V. El índice kappa. *SEMERGEN - Med Fam* 2001;27:247–9. [https://doi.org/10.1016/S1138-3593\(01\)73955-X](https://doi.org/10.1016/S1138-3593(01)73955-X).
- [74] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manag* 2009;45:427–37. <https://doi.org/10.1016/j.ipm.2009.03.002>.
- [75] Shah B, Kunal S, Bansal A, Jain J, Poundrik S, Shetty M, et al. Heart rate variability as a marker of cardiovascular dysautonomia in post-COVID-19 syndrome using artificial intelligence. *Indian Pacing Electrophysiol J* 2022;22. <https://doi.org/10.1016/j.ipej.2022.01.004>.
- [76] Al-Hadrawi DS, Al-Rubaye HT, Almulla AF, Al-Hakeim HK, Maes M. Lowered oxygen saturation and increased body temperature in acute COVID-19 largely predict chronic fatigue syndrome and affective symptoms due to Long COVID: A precision nomothetic approach. *Acta Neuropsychiatr* 2023;35:76–87. <https://doi.org/10.1017/neu.2022.21>.
- [77] Al-Hakeim HK, Al-Rubaye HT, Al-Hadrawi DS, Almulla AF, Maes M. Long-COVID post-viral chronic fatigue and affective symptoms are associated with oxidative damage, lowered antioxidant defenses and inflammation: a proof of concept and mechanism study. *Mol Psychiatry* 2023;28:564–78. <https://doi.org/10.1038/s41380-022-01836-9>.
- [78] Scarpelli S, De Santis A, Alfonsi V, Gorgoni M, Morin CM, Espie C, et al. The role of sleep and dreams in long-COVID. *J Sleep Res* 2023;32:e13789. <https://doi.org/10.1111/jsr.13789>.

- 
- [79] Alzueta E, Perrin PB, Yuksel D, Ramos-Usuga D, Kiss O, Iacovides S, et al. An international study of post-COVID sleep health. *Sleep Health* 2022;8:684–90. <https://doi.org/10.1016/j.sleh.2022.06.011>.
- [80] Aranyó J, Bazan V, Lladós G, Dominguez MJ, Bisbal F, Massanella M, et al. Inappropriate sinus tachycardia in post-COVID-19 syndrome. *Sci Rep* 2022;12:298. <https://doi.org/10.1038/s41598-021-03831-6>.
- [81] Freire APCF, Lira FS, Morano AE von A, Pereira T, Coelho-E-Silva M-J, Caseiro A, et al. Role of body mass and physical activity in autonomic function modulation on Post-COVID-19 condition: An observational subanalysis of fit-COVID study. *Int J Environ Res Public Health* 2022;19:2457. <https://doi.org/10.3390/ijerph19042457>.
- [82] Kurtoğlu E, Afsin A, Aktaş İ, Aktürk E, Kutlusoy E, Çağışar Ö. Altered cardiac autonomic function after recovery from COVID-19. *Ann Noninvasive Electrocardiol Off J Int Soc Holter Noninvasive Electrocardiol Inc* 2022;27:e12916. <https://doi.org/10.1111/anec.12916>.
- [83] Sirayder U, Inal-Ince D, Kepenek-Varol B, Acik C. Long-term characteristics of severe COVID-19: respiratory function, functional capacity, and quality of life. *Int J Environ Res Public Health* 2022;19:6304. <https://doi.org/10.3390/ijerph19106304>.