

UNIVERSIDAD DE VALLADOLID  
MÁSTER UNIVERSITARIO  
**Ingeniería Informática**



TRABAJO FIN DE MÁSTER

**Desarrollo de modelos de Deep Learning  
con herramientas de Inteligencia Artificial  
Explicable (XAI) para el diagnóstico  
automático de la apnea del sueño a partir  
del movimiento respiratorio**

Realizado por **Marta Fernández Poyatos**



**Universidad de Valladolid**

**25 de junio de 2024**

Tutores: Benjamín Sahelices Fernández

Verónica Barroso García



## Resumen

La apnea del sueño es una patología con una alta prevalencia en la población adulta, caracterizada por pausas respiratorias durante el sueño, que pueden ser totales (apnea) o parciales (hipopnea). Estas pausas tienen una duración mínima de 10 segundos y, repetidas durante la noche, provocan una ventilación inadecuada y un sueño fragmentado y no reparador. Está relacionada con múltiples patologías de tipo cardiovascular, somnolencia diurna, trastornos metabólicos, cáncer y enfermedades neurodegenerativas. A pesar de estos riesgos, sigue siendo una enfermedad infradiagnosticada. La principal razón es la disponibilidad limitada de la prueba diagnóstica estándar: la polisomnografía nocturna. Esta prueba se realiza en unidades del sueño especializadas, ubicadas en hospitales, y se monitorizan hasta 32 señales fisiológicas, lo que la convierte en una prueba costosa y que consume mucho tiempo, provocando largas listas de espera y retrasos en el diagnóstico y en el tratamiento. En busca de soluciones más accesibles y eficientes, en este trabajo se va a evaluar la capacidad de modelos de *deep learning* para estimar la severidad de la apnea del sueño de los sujetos de la base de datos del Sleep Heart Health Study (SHHS) empleando únicamente las señales de esfuerzo respiratorio torácico y abdominal, transformadas a una representación avanzada llamada bispectrum. Se ha tratado de estimar tanto el índice de apnea-hipopnea para todos los eventos apneicos, como específicamente para los eventos centrales, en los que es el cerebro el que no envía las señales adecuadas a los músculos que controlan la respiración. Para todos los eventos apneicos, se ha obtenido un índice kappa de 0.169 y una tasa de acierto del 44.8 % en el conjunto de test. En el caso específico de los eventos apneicos centrales, el índice kappa fue de 0.297 y la tasa de acierto alcanzó el 97.8 %. Empleando técnicas de inteligencia artificial explicable se ha podido comprobar que los modelos de redes convolucionales no han logrado capturar eficazmente los patrones presentes en las señales de esfuerzo respiratorio transformadas a bispectrums. Esto sugiere que, a diferencia de otras señales cardiorrespiratorias, los eventos de apnea e hipopnea no provocan cambios, o estos son demasiado sutiles, en las interacciones no lineales, el acoplamiento o la gaussianidad de las señales de esfuerzo respiratorio torácico y abdominal. Como resultado, las alteraciones producidas por estos eventos no se reflejan adecuadamente en los bispectrum de estas señales.

## Descriptores

Apnea del sueño, SHHS, señales de esfuerzo respiratorio, señal abdominal, señal torácica, bispectrum, inteligencia artificial explicable, Gradient-weighted Class Activation Mapping, *deep learning*, redes neuronales convolucionales.



### Abstract

Sleep apnea is a prevalent disorder in the adult population, characterized by respiratory pauses during sleep, which can be either complete (apnea) or partial (hypopnea). These pauses have a minimum duration of 10 seconds and, when occur repeatedly throughout the night, lead to inadequate ventilation and fragmented, unrefreshing sleep. This condition is associated with numerous major health problems, including cardiovascular diseases, daytime sleepiness, metabolic disorders, cancer, and neurodegenerative diseases. Despite these risks, sleep apnea remains underdiagnosed. The primary reason for this is the limited availability of the standard diagnostic test: overnight polysomnography. This test is conducted in specialized sleep units located in hospitals, monitoring up to 32 physiological signals, making it expensive and time-consuming, leading to long waiting lists and delays in both diagnosis and treatment. In search of more accessible and efficient solutions, this study evaluates the capability of deep learning models to estimate the severity of sleep apnea in subjects from the Sleep Heart Health Study (SHHS) database using only thoracic and abdominal respiratory effort signals, transformed into an advanced representation called bispectrum. The study attempts to estimate both the apnea-hypopnea index for all apneic events and specifically for central apneic events, where inadequate signaling from the brain to the respiratory muscles occurs. For all apneic events, a kappa index of 0.169 and an accuracy of 44.8% were obtained in the test set. For central apneic events specifically, the kappa index was 0.297 and the accuracy reached 97.8%. Using explainable artificial intelligence techniques, it was found that convolutional neural network models have not effectively captured the patterns present in bispectrum-transformed respiratory effort signals. This suggests that unlike other cardiorespiratory signals, apnea and hypopnea events do not induce significant changes, or these changes are too subtle, in the nonlinear interactions, coupling, or gaussianity of thoracic and abdominal respiratory effort signals. Consequently, the alterations caused by these events are not adequately reflected in the bispectrum of these signals.

### Keywords

Sleep apnea, SHHS, respiratory effort signals, abdominal signal, thoracic signal, bispectrum, explainable artificial intelligence, Gradient-weighted Class Activation Mapping, deep learning, convolutional neural network.



## **Agradecimientos**

A mis tutores Benjamín y Verónica por su constante dedicación, por permitirme colaborar con ellos y seguir las investigaciones que comenzaron durante el trabajo de fin de grado y por introducirme en el mundo de la investigación científica. Y a los grupos de Modelado, Biomecánica y Visualización Avanzada del Patrimonio (MoViVAP), de Caracterización de Materiales y Dispositivos Electrónicos (GCME) y de Ingeniería Biomédica (GIB) de la Universidad de Valladolid, por proporcionarme todos los recursos necesarios para el desarrollo del proyecto.

A mis padres, por su apoyo incondicional y los valores que me han transmitido desde pequeña. Y a mi pareja, por creer siempre en mí, ayudarme a crecer e inspirarme a seguir adelante.





---

# Índice general

---

|                                                                  |           |
|------------------------------------------------------------------|-----------|
| Índice general                                                   | VII       |
| Índice de figuras                                                | IX        |
| Índice de tablas                                                 | XI        |
| Siglas                                                           | XIII      |
| <b>1. Introducción</b>                                           | <b>1</b>  |
| 1.1. Hipótesis y objetivos . . . . .                             | 3         |
| 1.2. Agenda 2030 . . . . .                                       | 4         |
| 1.3. Tecnologías empleadas . . . . .                             | 6         |
| 1.4. Estructura del documento . . . . .                          | 7         |
| <b>2. Planificación</b>                                          | <b>8</b>  |
| 2.1. Metodología SCORE . . . . .                                 | 8         |
| 2.2. Alcance del proyecto . . . . .                              | 9         |
| 2.3. Gestión del cronograma . . . . .                            | 10        |
| 2.4. Gestión de los recursos y del coste . . . . .               | 11        |
| 2.5. Gestión de los riesgos . . . . .                            | 13        |
| <b>3. Contexto médico</b>                                        | <b>16</b> |
| 3.1. Apnea del sueño . . . . .                                   | 16        |
| 3.2. Polisomnografía . . . . .                                   | 22        |
| <b>4. Contexto tecnológico</b>                                   | <b>26</b> |
| 4.1. Redes neuronales convolucionales . . . . .                  | 29        |
| 4.2. Arquitecturas de redes neuronales convolucionales . . . . . | 35        |
| 4.3. Inteligencia Artificial Explicable . . . . .                | 41        |
| <b>5. Sujetos y señales</b>                                      | <b>46</b> |

|                                                                           |           |
|---------------------------------------------------------------------------|-----------|
| 5.1. Sleep Heart Health Study . . . . .                                   | 46        |
| 5.2. Sujetos y señales a estudiar . . . . .                               | 49        |
| 5.3. Procesamiento de señales . . . . .                                   | 51        |
| 5.4. Construcción de bispectrums . . . . .                                | 54        |
| <b>6. Metodología experimental</b>                                        | <b>56</b> |
| 6.1. Variantes del problema . . . . .                                     | 56        |
| 6.2. Entrada de datos . . . . .                                           | 58        |
| 6.3. Arquitecturas . . . . .                                              | 58        |
| 6.4. Estrategia de aprendizaje . . . . .                                  | 62        |
| 6.5. Estrategia de validación . . . . .                                   | 65        |
| <b>7. Resultados</b>                                                      | <b>67</b> |
| 7.1. Problema clásico de regresión . . . . .                              | 67        |
| 7.2. Problema de clasificación . . . . .                                  | 68        |
| 7.3. Problema de regresión junto con una regresión lineal final . . . . . | 71        |
| 7.4. Problema de regresión con nuevo mapa de color . . . . .              | 73        |
| 7.5. Problema de regresión con señal completa . . . . .                   | 74        |
| 7.6. Análisis de los mejores resultados . . . . .                         | 75        |
| <b>8. Conclusiones y líneas de trabajo futuras</b>                        | <b>85</b> |
| 8.1. Conclusiones . . . . .                                               | 85        |
| 8.2. Limitaciones . . . . .                                               | 86        |
| 8.3. Líneas futuras . . . . .                                             | 87        |
| <b>Bibliografía</b>                                                       | <b>88</b> |

---

# Índice de figuras

---

|                                                                                                                                                                                               |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1. Estructura de Descomposición de Trabajo . . . . .                                                                                                                                        | 12 |
| 2.2. Cronograma . . . . .                                                                                                                                                                     | 13 |
| 3.3. Sistema 10-20 de posicionamiento de electrodos en EEG (adaptado de [70]) . .                                                                                                             | 24 |
| 4.4. Diagrama de Venn que muestra la relación entre la IA y los campos fundamen-<br>tales de la IA . . . . .                                                                                  | 27 |
| 4.5. Comparación entre <i>machine learning</i> y <i>deep learning</i> . . . . .                                                                                                               | 28 |
| 4.6. Arquitectura LeNet-5 propuesta en [39] para el reconocimiento de dígitos . . .                                                                                                           | 30 |
| 4.7. Operación de convolución . . . . .                                                                                                                                                       | 31 |
| 4.8. Aplicación de diferentes <i>kernels</i> predefinidos a una imagen . . . . .                                                                                                              | 32 |
| 4.9. Aplicación de un <i>padding</i> de ceros a una matriz de datos . . . . .                                                                                                                 | 33 |
| 4.10. Aplicación de una operación de <i>pooling</i> sobre una matriz de datos . . . . .                                                                                                       | 34 |
| 4.11. Aplanado y capa <i>fully-connected</i> en una CNN . . . . .                                                                                                                             | 35 |
| 4.12. Comparación de funciones anidadas y no anidadas en busca de $f^*$ . . . . .                                                                                                             | 37 |
| 4.13. Bloque de redes neuronales común (izquierda) y bloque residual (derecha) . .                                                                                                            | 38 |
| 4.14. Comparación de escalado de dimensiones (ancho, profundidad y resolución)<br>en CNN convencionales (izquierda), y escalado compuesto en arquitectura<br>EfficientNet (derecha) . . . . . | 39 |
| 4.15. Comparación del rendimiento de las distintas variantes de modelo EfficientNet<br>con otras arquitecturas reconocidas del estado del arte (extraído de [80]) . . . .                     | 40 |
| 4.16. Conexiones en modelo DenseNet . . . . .                                                                                                                                                 | 40 |
| 4.17. Comparación ResNet (izquierda) y DenseNet (derecha) . . . . .                                                                                                                           | 41 |
| 4.18. Aplicación de la técnica de Grad-Cam a un bispectrum . . . . .                                                                                                                          | 43 |
| 4.19. Aplicación de ablaciones sobre un bispectrum sin eventos apneicos . . . . .                                                                                                             | 44 |
| 5.20. Fragmentos de señal con un evento central, un evento obstructivo y sin eventos                                                                                                          | 50 |
| 5.21. Distribución de los datos de entrenamiento en los conjuntos de entrenamiento,<br>validación y test . . . . .                                                                            | 51 |
| 5.22. Comparación de una señal observada y esta misma señal tras aplicar un filtro<br>de paso bajo, con FIR como método de diseño y ventana de Hamming . . . . .                              | 53 |
| 6.23. Comparación de los mapas de color empleados en las distintas variantes . . . .                                                                                                          | 57 |

|                                                                                                                                       |    |
|---------------------------------------------------------------------------------------------------------------------------------------|----|
| 6.24. Distribución de los bispetrums ABDO y THOR como datos de entradas de los modelos . . . . .                                      | 59 |
| 6.25. Arquitectura Double ResNet 18 . . . . .                                                                                         | 60 |
| 6.26. Arquitectura de red CNN de diseño propio . . . . .                                                                              | 61 |
| 7.27. Evolución del <i>loss</i> para los mejores modelos de cada variante . . . . .                                                   | 77 |
| 7.28. Matrices de confusión calculadas sobre el conjunto de test . . . . .                                                            | 79 |
| 7.29. Gradcam sobre un sujeto con apnea severa y una predicción correcta de severidad del IAH con modelo EfficientNet b1 . . . . .    | 81 |
| 7.30. Ablaciones sobre un sujeto con apnea severa y una predicción correcta de severidad del IAH con modelo EfficientNet b1 . . . . . | 82 |
| 7.31. Gradcam sobre un sujeto con apnea severa y una predicción correcta de severidad del IAC con modelo ResNet RS . . . . .          | 83 |
| 7.32. Ablaciones sobre un sujeto con apnea severa y una predicción correcta de severidad IAC con modelo ResNet RS . . . . .           | 84 |

---

## Índice de tablas

---

|                                                                                                                                                         |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1. Requisito R-01 . . . . .                                                                                                                           | 9  |
| 2.2. Requisito R-02 . . . . .                                                                                                                           | 10 |
| 2.3. Requisito R-03 . . . . .                                                                                                                           | 10 |
| 2.4. Requisito R-04 . . . . .                                                                                                                           | 10 |
| 2.5. Requisito R-05 . . . . .                                                                                                                           | 11 |
| 2.6. Requisito R-06 . . . . .                                                                                                                           | 11 |
| 2.7. Requisito R-07 . . . . .                                                                                                                           | 11 |
| 2.8. Requisito R-08 . . . . .                                                                                                                           | 12 |
| 2.9. Hitos y fecha estimada . . . . .                                                                                                                   | 12 |
| 2.10. Riesgos del proyecto . . . . .                                                                                                                    | 14 |
| 5.11. Tabla resumen de los datos sociodemográficos de los participantes en el estudio                                                                   | 52 |
| 5.12. Representación de bispectrum sin (a) y con (b) eventos apneicos para las señales<br>de respiración torácica y abdominal . . . . .                 | 55 |
| 6.13. Hiperparámetros explorados . . . . .                                                                                                              | 63 |
| 7.14. Resultados problema clásico de regresión (todos los eventos apneicos) . . . . .                                                                   | 67 |
| 7.15. Resultados problema clásico de regresión (eventos centrales) . . . . .                                                                            | 68 |
| 7.16. Resultados clasificación con umbral 5 e/h (todos los eventos apneicos) . . . . .                                                                  | 69 |
| 7.17. Resultados clasificación con umbral 5 e/h (eventos centrales) . . . . .                                                                           | 69 |
| 7.18. Resultados clasificación modelo ResNet 18 con exploración de grano grueso de<br>diferentes hiperparámetros (todos los eventos apneicos) . . . . . | 70 |
| 7.19. Resultados clasificación modelo ResNet 18 con exploración de grano grueso de<br>diferentes hiperparámetros (eventos centrales) . . . . .          | 71 |
| 7.20. Resultados regresión junto con una regresión lineal (todos los eventos apneicos)                                                                  | 72 |
| 7.21. Resultados regresión junto con una regresión lineal (eventos centrales) . . . . .                                                                 | 72 |
| 7.22. Resultados regresión junto con una regresión lineal (todos los eventos apneicos)                                                                  | 72 |
| 7.23. Resultados regresión junto con una regresión lineal (eventos centrales) . . . . .                                                                 | 73 |
| 7.24. Resultados regresión con nuevo mapa de color (todos los eventos apneicos) . . . . .                                                               | 73 |
| 7.25. Resultados regresión con nuevo mapa de color (eventos centrales) . . . . .                                                                        | 74 |
| 7.26. Resultados regresión con señal completa (todos los eventos apneicos) . . . . .                                                                    | 74 |
| 7.27. Resultados regresión con señal completa con exploración de grano grueso de<br>diferentes hiperparámetros (todos los eventos apneicos) . . . . .   | 75 |

|                                                                                                                                               |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------|----|
| 7.28. Resultados regresión con señal completa (eventos centrales) . . . . .                                                                   | 75 |
| 7.29. Resultados regresión con señal completa con exploración de grano grueso de<br>diferentes hiperparámetros (eventos centrales) . . . . .  | 76 |
| 7.30. Resumen de resultados (todos los eventos apneicos) . . . . .                                                                            | 76 |
| 7.31. Resumen de resultados (eventos centrales) . . . . .                                                                                     | 76 |
| 7.32. Resultados sobre test para el mejor resultado de estimación del IAH y del IAC<br>(* No existen sujetos con $IAC \geq 30$ e/h) . . . . . | 78 |

---

# Siglas

---

$pCO_2$  Presión parcial de dióxido de carbono en sangre

**AASM** *American Academy of Sleep Medicine*

**ABDO** Movimiento abdominal

**ACS** Apnea Central del Sueño

**Adam** *Adaptive Moment Estimation*

**AdamW** *Adaptive Moment Estimation Weight Decay*

**AOS** Apnea Obstructiva del Sueño

**CAM** *Class Activation Mapping*

**CNN** Redes Neuronales Convolucionales

**CPAP** *Continuous positive airway pressure*

**ECG** Electrocardiograma

**EDT** Estructura de Descomposición de Trabajo

**EEG** Electroencefalograma

**EMG** Electromiograma

**EOG** Electrooculograma

**FA** Flujo Aéreo (*airflow*)

**FC** *fully-connected*

**HOS** *Higher Order Spectrum*

**HSAT** *Home Sleep Apnea Testing*

**IA** Inteligencia Artificial

**IAC** Índice de apnea central

**IAH** Índice de apnea-hipopnea

**ICSAS** *Idiopathic Central Sleep Apnea Syndrome*

**ILSVRC** *ImageNet Large Scale Visual Recognition Challenge*

**IMC** Índice de Masa Corporal

**MSE** *Mean Squared Error*

**NREM** Sueño sin movimientos oculares rápidos

**NSRR** *National Sleep Research Resource*

**ODS** Objetivos de Desarrollo Sostenible

**OLS** Mínimos Cuadrados Ordinarios (*Ordinary Least Squares*)

**PMBOK** *Project Management Body of Knowledge*

**PSG** Polisomnografía

**REM** Sueño con movimientos oculares rápidos (*Rapid Eye Movement*)

**RIP** Pletismografía inductiva respiratoria (*Respiratory Inductance Plethysmography*)

**RNN** Redes Neuronales Recurrentes

**SHHS** *Sleep Heart Health Study*

**THOR** Movimiento torácico



**VGG** *Visual Geometry Group*

**VPN** Valor Predictivo Negativo

**VPP** Valor Predictivo Positivo

**XAI** *Explainable Artificial Intelligence*

**IRPF** Impuesto sobre la Renta de las Personas Físicas

---

# 1: Introducción

---

La apnea del sueño es un trastorno respiratorio caracterizado por presentar repetidas pausas totales (apnea) o parciales (hipopnea) de la respiración. Estas pausas pueden durar desde 10 segundos hasta varios minutos y producirse en repetidas ocasiones durante una noche de sueño [9]. Estos eventos respiratorios se pueden clasificar en obstructivos, cuando los músculos de la garganta se relajan en exceso y bloquean las vías respiratorias; centrales, cuando es el cerebro el que no envía las señales adecuadas a los músculos que controlan la respiración; y mixtos, que son una combinación de los anteriores. La apnea del sueño está relacionada con múltiples patologías de tipo cardiovascular, pudiendo aumentar la mortalidad en pacientes con enfermedades pulmonares, producir complicaciones en procesos postoperatorios [56], somnolencia diurna [87] o aumentar el riesgo de aparición de diabetes de tipo 2 [87]. Además, estas pausas respiratorias, producidas de forma recurrente, provocan una ventilación inadecuada y un sueño fragmentado y no reparador.

A pesar de las múltiples complicaciones que puede provocar esta enfermedad, la apnea del sueño a menudo está infradiagnosticada [91], en gran medida debido a la limitada disponibilidad y el alto coste de la prueba diagnóstica estándar: la polisomnografía nocturna [70]. Esta prueba consiste en un estudio detallado de una noche de sueño de un paciente que se realiza en las unidades del sueño en el hospital. En la polisomnografía se monitorizan múltiples señales fisiológicas de los pacientes, como la actividad cerebral, los niveles de oxígeno en sangre, la frecuencia cardíaca, la respiración y los movimientos oculares y musculares. A pesar de la fiabilidad y la precisión de esta prueba diagnóstica, tiene una larga lista de inconvenientes. En primer lugar, requiere la estancia del paciente durante una noche completa en el hospital, acompañado en todo momento de personal especializado [8] [22]. Además, las grabaciones nocturnas se deben analizar posteriormente de manera visual para determinar el diagnóstico del paciente. El análisis de estas señales es complejo, lento y puede dar lugar a diagnósticos subjetivos. Todo esto hace que la polisomnografía sea una prueba cara y requiera mucho tiempo. Esto, unido a la alta prevalencia de la enfermedad, provoca sobrecargas en las unidades del sueño, largas listas de espera y retrasos en el diagnóstico y en el tratamiento de la enfermedad [17].

Por este motivo, la comunidad científica busca alternativas diagnósticas automáticas que simplifiquen la detección de esta enfermedad. En esta búsqueda, surge la aplicación

de algoritmos de *deep learning*, que permiten descubrir patrones y características de los datos de forma automática, evitando procesos costosos de procesamiento. La motivación de este trabajo es justo esta, tratar de estimar la severidad de la apnea del sueño a partir de las señales de esfuerzo respiratorio torácico y abdominal. Para ello se van a emplear algoritmos de *deep learning*, en concreto, redes neuronales convolucionales, con el fin de obtener diagnósticos objetivos, llegando a un grupo de la población más numeroso, con un menor coste y en un menor tiempo.

Algunos estudios del estado del arte analizan las características de las señales de esfuerzo respiratorio torácico y abdominal, como la amplitud, la frecuencia o la correlación entre ambas mediante técnicas de *feature engineering*, para clasificar los eventos apneicos en obstructivos, centrales o respiración normal [40] [58]. Otros estudios realizaron esta misma clasificación, pero empleando la diferencia de fase en el dominio del tiempo [86], la fase instantánea y el valor de acoplamiento de fase de las dos señales [2]. En cambio, todos estos estudios, además de requerir un gran conocimiento de las señales para realizar las transformaciones pertinentes, emplean conjuntos de datos con un tamaño muy reducido, haciendo muy difícil utilizar técnicas más avanzadas de aprendizaje profundo. En este trabajo se va a emplear una base de datos mucho más grande, permitiendo reducir el sesgo y aplicar técnicas avanzadas de *deep learning* para identificar de forma automática las características de las señales.

Posteriormente, se han desarrollado estudios en los que se emplean técnicas de aprendizaje profundo como los modelos de redes convolucionales, modelos de redes recurrentes o combinaciones de ambos sobre las señales de esfuerzo torácico y abdominal unidas a las señales de flujo aéreo [25] [88]. Sin embargo, estos estudios se centran en la detección o diferenciación de eventos apneicos, y no en la estimación de la severidad de la enfermedad. Por otro lado, algunos estudios han explorado el uso de las señales de esfuerzo respiratorio torácico y abdominal junto con la señal de oximetría para clasificar a los sujetos estudiados en diferentes grupos de severidad [68]. También se han desarrollado estudios que se enfrentan al mismo problema que el abordado en este trabajo, pero empleando otras señales fisiológicas, como la saturación de oxígeno [19] [69], el flujo aéreo y la saturación de oxígeno [34] o el electrocardiograma [23] [61]. En cambio, el uso de estas señales no permite distinguir entre los eventos obstructivos y centrales. Como la propia AASM recomienda [9], las señales ideales para la identificación de los eventos apneicos centrales son las señales de esfuerzo respiratorio torácico y abdominal. Es importante destacar que la distinción entre estos dos tipos de eventos apneicos resulta fundamental para proporcionar un tratamiento adecuado y efectivo a cada paciente.

En etapas previas a esta investigación, se ha estudiado el problema de estimación de la severidad de la apnea del sueño a partir de las señales de esfuerzo respiratorio torácico y abdominal [4], obteniendo buenos resultados, especialmente en la estimación de la severidad de la apnea central. En este proyecto se trata de ir un paso más allá, y en lugar de estudiar las señales en crudo, se va a utilizar un tipo de representación especial de las señales llamado bispectrum, que destaca por su capacidad para reflejar cambios en la linealidad, gaussianidad y acoplamiento de la señal [15] [97], proporcionando una visión

más profunda de las interacciones entre las diferentes componentes de frecuencia. El uso de bispectrums ha demostrado ser de gran utilidad en el contexto de la apnea del sueño con otras señales como el flujo aéreo, la saturación de oxígeno en sangre, la variabilidad de la frecuencia cardíaca o el electrocardiograma [5] [47]. Sin embargo, hasta la fecha, los estudios han empleado principalmente técnicas de *feature engineering* y no han explorado plenamente el potencial del bispectrum en el ámbito del *deep learning*. A partir de los bispectrums se pretende obtener una nueva forma de analizar las señales y ver si a partir de las frecuencias presentadas en las imágenes, se obtiene una mejora en la estimación de la severidad de la apnea del sueño, tanto central como obstructiva. Además, se emplearán técnicas de inteligencia artificial explicable con el objetivo de que los resultados puedan ser interpretados por los médicos y se mejore la confianza en el uso de la inteligencia artificial en el ámbito médico. Hasta donde se conoce, ésta es la primera vez que se aborda el problema de estimación de la severidad de la apnea del sueño empleando los bispectrums de las señales de esfuerzo respiratorio torácico y abdominal y también es la primera vez que se aplican técnicas de explicabilidad en este contexto específico y con estas señales.

## 1.1. Hipótesis y objetivos

Para la definición de los objetivos de este trabajo, se puede formular una hipótesis para indicar el punto de partida del estudio: la representación en forma de bispectrums de las señales de esfuerzo respiratorio torácico y abdominal contiene información suficiente para estimar automáticamente la severidad de la apnea del sueño para los eventos apneicos en general, y los eventos centrales en particular. Específicamente, se espera que las redes neuronales convolucionales sean capaces de extraer de estos bispectrums las características esenciales necesarias para una estimación precisa de la severidad de la apnea. Además, se espera que las técnicas de inteligencia artificial explicable sean capaces de determinar las regiones específicas de los bispectrums en las que los modelos de *deep learning* se basan para elaborar sus predicciones.

El objetivo principal de este trabajo es el desarrollo de técnicas avanzadas de *deep learning* basadas en redes neuronales convolucionales para estimar la severidad de la apnea del sueño. Mientras que los métodos tradicionales de *feature engineering* requieren una caracterización detallada de las señales para extraer características relevantes, los modelos de *deep learning* automatizan la extracción de la información más significativa de las señales [54]. Esto se vuelve especialmente beneficioso en el análisis de señales biomédicas, ya que las técnicas de *deep learning* ofrecen una gran flexibilidad y un buen rendimiento tanto en la interpretación de señales unidimensionales como en representaciones bidimensionales, tales como espectrogramas, *bispectrums* o *recurrence plots*. Además, se pretende adaptar estas redes para poder procesar varias señales biomédicas de forma simultánea.

Resulta fundamental que los modelos de *deep learning* desarrollados actúen como herramientas de apoyo complementarias al criterio médico. Por lo tanto, se debe priorizar la creación de métodos que sean fácilmente explicables y comprensibles para los profesionales de la salud, con el fin de que puedan integrar estas tecnologías en su trabajo diario de

manera efectiva. Además, se aspira a que estas técnicas ofrezcan una alternativa viable y más barata a los métodos de diagnóstico tradicionales. Para lograr esto, se debe poner un énfasis especial en la interpretabilidad de los resultados obtenidos mediante estos métodos.

Tras completar este trabajo se espera haber alcanzado los siguientes objetivos secundarios:

- Adquirir los conocimientos básicos sobre el contexto médico de la apnea del sueño.
- Aprender los conceptos básicos de teoría de la señales necesarios para trabajar con señales fisiológicas.
- Comprender el procesamiento de señales biomédicas y su conversión en representaciones bidimensionales específicas, como el *bispectrum*.
- Aprender los conceptos básicos relacionados con la explicabilidad de los modelos de *deep learning*.
- Desarrollar tecnología de *deep learning* explicable para el diagnóstico de la apnea del sueño a partir de señales fisiológicas.
- Evaluar los resultados obtenidos tras la ejecución de diversos experimentos a partir de un modelo base.
- Documentar exhaustivamente todo el trabajo realizado durante la investigación.

## 1.2. Agenda 2030

Con el título “Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible” [83] la Asamblea General de las Naciones Unidas, celebrada en Nueva York el 25 de septiembre de 2015, aprobó la Resolución A/RES/70/1. Este documento constituye un marco referencial compuesto por 17 objetivos y 169 metas, cuya finalidad es concretar un plan de acción en favor de las personas, el planeta y la prosperidad, que busca erradicar la pobreza, proteger el planeta, así como hacer proteger los derechos humanos de todas las personas, y alcanzar la igualdad entre los géneros y el empoderamiento de todas las mujeres y niñas. Estos objetivos y metas abarca las esferas económica, social y ambiental; y se tratarán de alcanzar en los 15 años posteriores a su publicación.

Como ya se ha mencionado previamente, el objetivo del trabajo descrito en el presente documento es la estimación de la severidad de la apnea del sueño a través de modelos de *deep learning*, para lograr un diagnóstico de forma precisa, rápida y sencilla, ayudando a aliviar los altos costes económicos y temporales de la prueba diagnóstica de esta enfermedad y a reducir las largas listas de espera en el diagnóstico y en el tratamiento. Además, la posibilidad de ofrecer una estimación, no solo de los eventos apneicos en general, sino también de los eventos centrales en particular, permite proporcionar un tratamiento específico y adecuado para cada paciente. Este trabajo se alinea estrechamente con la

Agenda 2030 y puede contribuir al logro de los Objetivos de Desarrollo Sostenible (ODS). En particular se puede destacar su influencia en los siguientes objetivos y metas concretas:

- **ODS 1: fin de la pobreza, meta 1.3:** esta meta pretende “*implementar a nivel nacional sistemas y medidas apropiados de protección social para todos, incluidos niveles mínimos, y lograr una amplia cobertura de las personas pobres y vulnerables*” [83]. El desarrollo de tecnología *deep learning* para el diagnóstico de la apnea el sueño permite reducir los costes asociados tanto al diagnóstico como al tratamiento, reduciendo el impacto financiero sobre población vulnerable que hace que los sistemas de protección sean más inclusivos.
- **ODS 3: salud y bienestar, meta 3.4:** esta meta trata de “*reducir en un tercio la mortalidad prematura por enfermedades no transmisibles mediante su prevención y tratamiento, y promover la salud mental y el bienestar*” [83]. Un diagnóstico adecuado y temprano de la apnea del sueño evita complicaciones u otras enfermedades asociadas, como enfermedades cardiovasculares o accidentes cerebrovasculares; reduciendo la mortalidad prematura por dichas patologías y mejorando la calidad de vida.
- **ODS 3: salud y bienestar, meta 3.8:** con esta meta se pretende lograr “*la cobertura sanitaria universal, incluida la protección contra los riesgos financieros, el acceso a servicios de salud esenciales de calidad y el acceso a medicamentos y vacunas inocuos, eficaces, asequibles y de calidad para todos*” [83]. La alta reducción de costes que permite el diagnóstico automatizado de la apnea del sueño a partir de tecnología *deep learning* permite que esta herramienta sea accesible a toda la población, facilitando una cobertura sanitaria universal.
- **ODS 8: trabajo decente y crecimiento económico, meta 8.5:** esta meta pretende que se logre “*empleo pleno y productivo y el trabajo decente para todas las mujeres y los hombres, incluidos los jóvenes y las personas con discapacidad, así como la igualdad de remuneración por trabajo de igual valor*” [83]. Dada la condición de salud que supone la apnea y su influencia en patologías asociadas, se contribuye al logro de una fuerza laboral más saludable y productiva, así como una disminución de las tasas de absentismo laboral por incapacidades, sin olvidar que el desarrollo de tecnologías de inteligencia artificial generará nuevos empleos e industrias.
- **ODS 9: industria, innovación e infraestructura, meta 9.5:** esta meta trata de “*aumentar la investigación científica y mejorar la capacidad tecnológica de los sectores industriales de todos los países, en particular los países en desarrollo*” [83]. Este trabajo en sí mismo es un proyecto de investigación científica. El uso de inteligencia artificial y tecnología *deep learning* en cualquier ámbito de la salud es innovación, que además permite el desarrollo de nuevas tecnologías, nuevos dispositivos, aplicaciones y sistemas de monitorización que mejoran la infraestructura médica.
- **ODS 10: reducción de las desigualdades, meta 10.2:** esta meta trata de “*potenciar y promover la inclusión social, económica y política de todas las personas,*

*independientemente de su edad, sexo, discapacidad, raza, etnia, origen, religión o situación económica u otra condición” [83]. El abaratamiento de los costes en el diagnóstico y la reducción de las largas listas de espera hacen accesible a toda la población el diagnóstico de esta enfermedad, contribuyendo de manera efectiva a reducir las desigualdades en el acceso a la atención sanitaria, especialmente en el caso de comunidades con recursos limitados.*

### 1.3. Tecnologías empleadas

Durante el desarrollo de todo el proyecto serán necesarias las siguientes herramientas *software*:

- **MATLAB (MATrix LABoratory)**: es un entorno de desarrollo y un lenguaje de programación de alto nivel (M) ampliamente utilizado en el ámbito científico y de ingeniería para el análisis de datos, desarrollo de algoritmos y creación de modelos. También cuenta con un formato de almacenamiento de datos binarios (`.mat`). En este proyecto se va a emplear para el procesamiento de las señales de esfuerzo respiratorio, así como para la generación de los bispectrums a partir de éstas.
- **Python**: es un lenguaje de programación de alto nivel, interpretado y muy conocido por su sencillez y amplia gama de aplicaciones, como la programación de sistemas, desarrollo web o desarrollo de modelos de inteligencia artificial. En el contexto de este proyecto se va a emplear para el desarrollo de los experimentos de *deep learning*. Este lenguaje de programación permite el uso de múltiples bibliotecas, entre las que se van a utilizar: Matplotlib, NumPy, pickle, Seaborn, Scikit-learn y PyTorch.
- **PyTorch**: es una biblioteca de código abierto desarrollada por el AI Research Lab de Facebook para el desarrollo de aplicaciones de aprendizaje automático. Permite la construcción y entrenamiento de modelos de redes neuronales de forma intuitiva y flexible en una gran variedad de tareas como visión artificial, procesamiento de lenguaje natural o [IA generativa](#). En este proyecto se va a emplear para el desarrollo y entrenamiento de los modelos de redes neuronales convolucionales para la estimación de la severidad de la apnea del sueño.
- **Comet ML**: es una plataforma de gestión de experimentos diseñada para proyectos de *machine learning* y *deep learning*. Permite monitorizar el entrenamiento de modelos y visualizar métricas y resultados a través de *dashboards* interactivos, además de facilitar la gestión de experimentos. En este proyecto, Comet ML se utilizará para almacenar las métricas de validación de cada uno de los experimentos que se realicen.
- **Jupyter Notebook**: es una aplicación web de código abierto que permite crear y compartir documentos que contienen código ejecutable, texto, imágenes o ecuaciones matemáticas. Es una aplicación de gran popularidad en el mundo de la ciencia de datos, dado que permite desarrollar, documentar y visualizar análisis de datos en un

formato interactivo. Además, permite el uso de diferentes lenguajes de programación como Python, R o Julia. En el contexto de este proyecto se va a emplear para el desarrollo del código Python empleado para los experimentos.

## 1.4. Estructura del documento

En este primer capítulo se realiza una breve introducción al problema que se ha desarrollado a lo largo del documento, así como la motivación, los objetivos y las tecnologías empleadas. En el segundo capítulo se presenta la planificación y la metodología empleada que se pretende seguir para cumplir con los objetivos marcados para completar con éxito el proyecto. En el tercer capítulo se realiza una descripción del contexto médico de la apnea del sueño, tanto obstructiva, como central y mixta, y de la prueba diagnóstica por excelencia de esta enfermedad: la polisomnografía nocturna. En el cuarto capítulo se introducen los conceptos de *deep learning*, redes neuronales convolucionales e inteligencia artificial explicable que se van a desarrollar en etapas posteriores de la investigación. En el quinto capítulo se presenta la base de datos empleada durante todo el proyecto, y se describen las señales de esfuerzo respiratorio torácico y abdominal que se van a estudiar, así como el procesamiento al que se someten y su transformación en bispectrums. En el sexto capítulo se describe la metodología experimental desarrollada, las variantes del problema a estudiar, las arquitecturas y las estrategias de entrenamiento y validación empleadas. En el séptimo capítulo se presentan los resultados obtenidos tras la experimentación realizada y en el octavo y último capítulo se analizan las conclusiones extraídas tras el proyecto, así como las limitaciones del estudio y las futuras líneas de investigación.



---

## 2: Planificación

---

En este capítulo se va a describir la planificación seguida para el desarrollo del proyecto, tomando como referencia la metodología descrita en el *Project Management Body of Knowledge* (PMBOK) [31]. En primer lugar, se va a describir la metodología empleada, el alcance del proyecto, junto con una recopilación de requisitos y Estructura de Descomposición de Trabajo, el cronograma, los recursos y costes asociados al proyecto y los posibles riesgos que se pueden materializar a lo largo de la duración del proyecto.

### 2.1. Metodología SCORE

Durante el desarrollo de todo el proyecto se ha empleado una metodología de tipo ágil llamada SCORE (SCrum fOr REsearch) [28]. El uso de una metodología ágil implica dividir el proyecto en pequeñas tareas y evaluar constantemente los requisitos, planificación y resultados. Esto permite responder rápidamente a cualquier cambio que pueda surgir, proporcionando un enfoque iterativo en la gestión de proyectos y en el desarrollo de software. Las metodologías ágiles son conocidas por su flexibilidad y capacidad para adaptarse a entornos de trabajo dinámicos, y SCORE lleva estos principios al ámbito de la investigación.

Esta metodología, basada en los principios de Scrum pero adaptada para los proyectos de investigación, se enfoca en la definición de objetivos claros y medibles. A diferencia de Scrum, que se centra en el desarrollo de software, SCORE se adapta mejor a la naturaleza exploratoria y a menudo incierta de los proyectos de investigación. Permite una mayor flexibilidad en la planificación y facilita la colaboración efectiva entre los miembros del equipo.

La metodología SCORE tiene dos elementos fundamentales: *status meetings*, reuniones celebradas dos o tres veces por semana para monitorizar el avance del proyecto, y las *on-demand technical meetings*, reuniones bajo demanda y sin una frecuencia específica celebradas para tratar en mayor profundidad ciertos temas surgidos en las *status meetings*. En el marco del desarrollo de este proyecto, la comunicación entre los tutores y la alumna se ha realizado principalmente a través de reuniones semanales. Estas reuniones

han permitido una evaluación continua del progreso y una discusión abierta sobre cualquier problema o duda que haya surgido. Además, se han utilizado correos electrónicos cuando ha sido necesario para resolver dudas o coordinar aspectos específicos fuera de las reuniones programadas.

## 2.2. Alcance del proyecto

El objetivo fundamental del proyecto es el desarrollo de modelos de *deep learning* y la aplicación de técnicas de explicabilidad a los modelos desarrollados para la estimación de la severidad de la apnea del sueño, a partir de las señales de esfuerzo respiratorio de los pacientes recogidas durante una noche de sueño. El desarrollo de modelos de *deep learning* permite obtener diagnósticos precisos y en un tiempo breve, agilizando el tratamiento de la enfermedad de la apnea del sueño. Pero al mismo tiempo estos modelos se pueden considerar “cajas negras” dado que se desconoce la justificación de las predicciones de los modelos. Por este motivo se emplean técnicas de explicabilidad que permiten hacer que los resultados obtenidos sean interpretados por los médicos, con el fin de mejorar la precisión y confianza en el uso de inteligencia artificial para el diagnóstico de esta enfermedad.

Para la definición del alcance del proyecto se deben tener en cuenta las restricciones de la investigación: la duración del proyecto debe ser de 150 horas, comenzando el día 20 de mayo de 2024 y terminando el día 23 de junio de 2024, y los recursos *hardware* estarán limitados a los proporcionados por los tutores. También se debe tener en cuenta que el presente trabajo fin de máster se ha realizado tras la estancia en el grupo de investigación MoBiVAP durante la asignatura de I+D+i en Informática, por lo que las tareas realizadas en este trabajo representan la continuación de los trabajos realizados previamente en dicha estancia.

## Recopilación de requisitos

Los requisitos más importantes de este trabajo que serán necesarios para cumplir con los objetivos planteados al inicio del proyecto se muestran en las tablas 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7 y 2.8. Para cada requisito se presenta un identificador, nombre, una breve descripción, la identificación de las partes interesadas y su prioridad.

|                    |                                                                        |
|--------------------|------------------------------------------------------------------------|
| <b>ID</b>          | <b>R-01</b>                                                            |
| <b>Nombre</b>      | Inicio del trabajo                                                     |
| <b>Descripción</b> | El inicio del trabajo de fin de máster será el día 20 de mayo de 2024. |
| <b>Interesados</b> | Alumna, tutores                                                        |
| <b>Prioridad</b>   | Alta                                                                   |

Tabla 2.1: Requisito R-01

|                    |                                                                      |
|--------------------|----------------------------------------------------------------------|
| <b>ID</b>          | <b>R-02</b>                                                          |
| <b>Nombre</b>      | Fin estancia                                                         |
| <b>Descripción</b> | El fin del trabajo de fin de máster será el día 23 de junio de 2024. |
| <b>Interesados</b> | Alumna, tutores                                                      |
| <b>Prioridad</b>   | Alta                                                                 |

Tabla 2.2: Requisito R-02

|                    |                                                                                                                                                                                   |
|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ID</b>          | <b>R-03</b>                                                                                                                                                                       |
| <b>Nombre</b>      | <i>Software</i>                                                                                                                                                                   |
| <b>Descripción</b> | Se utilizará el lenguaje de programación Python y la biblioteca PyTorch para la creación de modelos de <i>deep learning</i> y metodologías de inteligencia artificial explicable. |
| <b>Interesados</b> | Alumna, tutores                                                                                                                                                                   |
| <b>Prioridad</b>   | Alta                                                                                                                                                                              |

Tabla 2.3: Requisito R-03

|                    |                                                                                                                                                |
|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ID</b>          | <b>R-04</b>                                                                                                                                    |
| <b>Nombre</b>      | Diagnóstico de la severidad de la apnea del sueño                                                                                              |
| <b>Descripción</b> | La temática sobre la que se trabajará será el diagnóstico de la enfermedad de la apnea del sueño a partir de señales de esfuerzo respiratorio. |
| <b>Interesados</b> | Alumna, tutores                                                                                                                                |
| <b>Prioridad</b>   | Alta                                                                                                                                           |

Tabla 2.4: Requisito R-04

## Estructura de Descomposición de Trabajo

La Estructura de Descomposición de Trabajo ([EDT](#)) es una herramienta fundamental para la gestión de proyectos, que permite visualizar las tareas y actividades necesarias para completar con éxito un proyecto. En la figura [2.1](#) se presenta la [EDT](#) que se llevará a cabo en este proyecto. El trabajo se ha dividido en cuatro grandes bloques: estudio del estado del arte, diseño de experimentos, entrenamiento y validación de los experimentos y documentación del trabajo.

### 2.3. Gestión del cronograma

A partir de la [EDT](#), se ha calendarizado cada tarea del proyecto, que se desarrollará en un total de 150 horas, distribuidas en cuatro bloques principales: estudio del estado del arte, diseño de experimentos, entrenamiento y validación de los modelos, y documentación del proyecto. Los tres primeros bloques se llevarán a cabo de forma secuencial, mientras

|                    |                                                                                                                                                |
|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ID</b>          | <b>R-05</b>                                                                                                                                    |
| <b>Nombre</b>      | Base de datos                                                                                                                                  |
| <b>Descripción</b> | Se analizarán las señales recogidas mediante polisomnografías nocturnas de la base de datos Sleep Heart Health Study ( <a href="#">SHHS</a> ). |
| <b>Interesados</b> | Alumna, tutores                                                                                                                                |
| <b>Prioridad</b>   | Alta                                                                                                                                           |

Tabla 2.5: Requisito R-05

|                    |                                                                                                                            |
|--------------------|----------------------------------------------------------------------------------------------------------------------------|
| <b>ID</b>          | <b>R-06</b>                                                                                                                |
| <b>Nombre</b>      | Señales a estudiar                                                                                                         |
| <b>Descripción</b> | Se analizarán las señales de esfuerzo respiratorio, en concreto las señales de esfuerzo respiratorio torácico y abdominal. |
| <b>Interesados</b> | Alumna, tutores                                                                                                            |
| <b>Prioridad</b>   | Alta                                                                                                                       |

Tabla 2.6: Requisito R-06

|                    |                                                                                                                                                |
|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ID</b>          | <b>R-07</b>                                                                                                                                    |
| <b>Nombre</b>      | Estudio del estado del arte                                                                                                                    |
| <b>Descripción</b> | Será necesario un estudio previo del estado del arte, centrado en la aplicación de técnicas de explicabilidad a modelos <i>deep learning</i> . |
| <b>Interesados</b> | Alumna, tutores                                                                                                                                |
| <b>Prioridad</b>   | Alta                                                                                                                                           |

Tabla 2.7: Requisito R-07

que el último bloque de documentación del proyecto será una actividad continua a lo largo de todo el proyecto. El cronograma, representado en el diagrama de Gantt de la figura 2.2, detalla las actividades con su duración estimada, y sus fechas de inicio y fin, abarcando las cuatro semanas desde el 20 de mayo de 2024 hasta el 23 de junio de 2024. Además, se han identificado cinco hitos, presentados en la tabla 2.9, que deben cumplirse para conseguir los objetivos del proyecto.

## 2.4. Gestión de los recursos y del coste

En la planificación de un proyecto de investigación en *deep learning* como el que se va a desarrollar, resulta necesario contar con una estrategia clara para la gestión de recursos y costes. Los recursos necesarios para el desarrollo de este proyecto se pueden clasificar como recursos materiales y recursos humanos.

Para estimar el coste de los recursos humanos en este proyecto, se va a considerar exclusivamente el salario de un investigador contratado. Este investigador debe poseer

|                    |                                                                                                                                                                                              |
|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ID</b>          | <b>R-08</b>                                                                                                                                                                                  |
| <b>Nombre</b>      | Uso de alguna técnica de XAI aplicada a la predicción de la severidad de la apnea del sueño                                                                                                  |
| <b>Descripción</b> | Se diseñará un modelo de <i>deep learning</i> y se le aplicará una técnica explicable para la predicción de la severidad de la apnea del sueño a partir de señales de esfuerzo respiratorio. |
| <b>Interesados</b> | Alumna, tutores                                                                                                                                                                              |
| <b>Prioridad</b>   | Alta                                                                                                                                                                                         |

Tabla 2.8: Requisito R-08

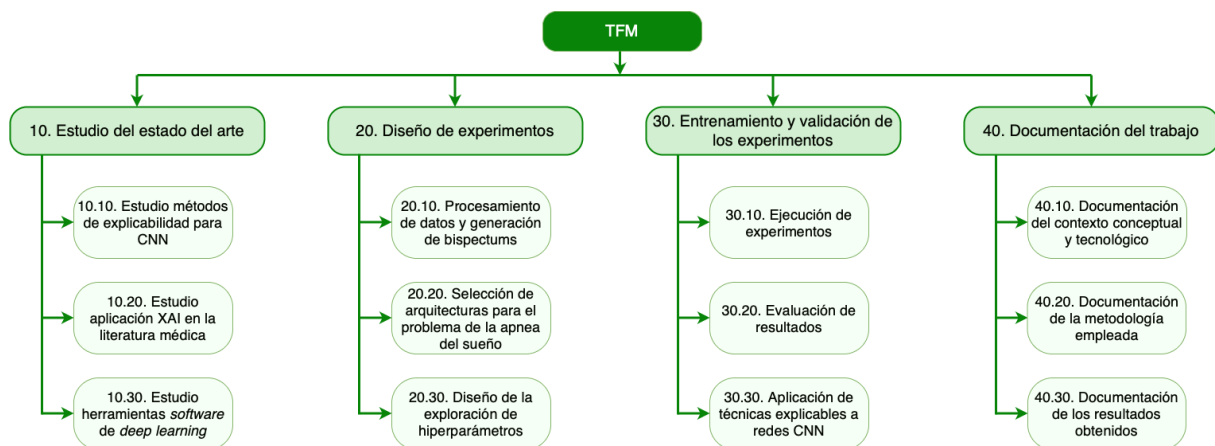


Figura 2.1: Estructura de Descomposición de Trabajo

| Nombre                                                    | Fecha        |
|-----------------------------------------------------------|--------------|
| <b>Inicio del trabajo</b>                                 | lun 20/05/24 |
| <b>Fin del estudio del estado del arte</b>                | vie 24/05/24 |
| <b>Fin del diseño de experimentos</b>                     | mar 04/06/24 |
| <b>Fin del entrenamiento y validación de experimentos</b> | mie 19/06/24 |
| <b>Fin de la documentación del trabajo</b>                | dom 23/06/24 |

Tabla 2.9: Hitos y fecha estimada

conocimientos avanzados en inteligencia artificial y *deep learning*, con una salario mínimo de 60 € por hora. Este importe cubre no solo el salario base, sino también todas las cotizaciones a la seguridad social, que incluyen tanto contingencias comunes como profesionales: Accidentes de Trabajo y Enfermedades Profesionales (AT y EP), desempleo, formación y Fondo de Garantía Salarial. Dado que la duración prevista del proyecto es de 150 horas y el coste por hora se ha establecido en 60 €, el coste total de los recursos humanos ascendería a 9000 €. Por simplicidad, en este cálculo no se van a considerar los costes asociados a los responsables del proyecto.

Respecto a los recursos materiales, se debe considerar en primer lugar el coste del ordenador personal del trabajador, un MacBook Air de 13 pulgadas (M1, 2020), con un

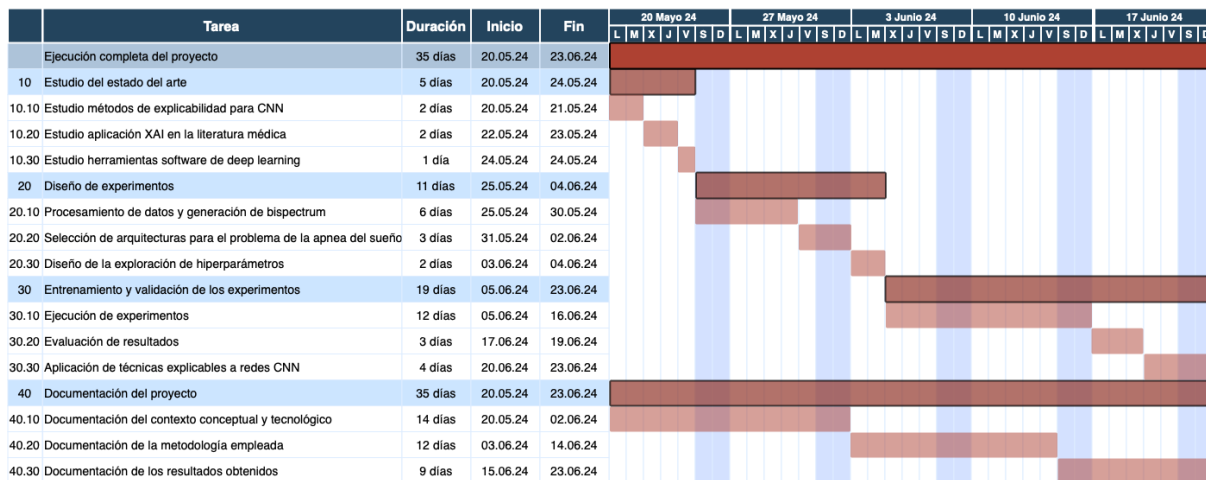


Figura 2.2: Cronograma

precio de compra de 1399 €. También se ha empleado una GPU NVIDIA GeForce RTX 3060 con 12 GB de RAM, proporcionada por el grupo de investigación, con un coste de 500 €. Teniendo en cuenta el coste de amortización para los equipos de tratamiento y sistemas y programas informáticos según la Agencia Tributaria, la amortización lineal máxima es del 26 % [81], lo que implica 363.74 € anuales por el ordenador personal del investigador y 130 € anuales por la GPU. El coste estimado para la duración del proyecto (un mes) sería de 30.31 € y 10.83 €, respectivamente.

A los costes materiales se debe sumar el coste del *software* que se va a emplear para el desarrollo del proyecto. La mayoría del *software* es libre, salvo Comet.ml y Overleaf. En el caso de Comet.ml no será necesaria la compra de una licencia, dado que dispone de licencias gratuitas. Por otro lado, sí será necesaria la compra de la licencia de Overleaf, por los cortos tiempos de compilación de los que dispone, insuficientes para la escritura de un documento de las dimensiones de un Trabajo de Fin de Máster. El precio mensual para los estudiantes es de 8 €. Por último, se debe añadir el precio de la luz e internet empleado por el investigador, que se estima en 20 €.

Teniendo en cuenta los recursos materiales y los recursos humanos, el coste total final del uso de los recursos para el desarrollo del proyecto, con una duración estimada de 1 mes, es de: 9000 €+ 30.31 €+ 10.83 €+ 8 €+ 20 € = 9069.14 €.

## 2.5. Gestión de los riesgos

En la planificación de un proyecto de investigación como el que se va a desarrollar, la gestión de riesgos resulta fundamental para completarlo con éxito. En la tabla 2.10 se muestran los riesgos que podrían materializarse, su clasificación en amenaza u oportunidad, así como la probabilidad de ocurrencia, el impacto que puede tener la materialización de estos riesgos y la prioridad que se le debe asignar a este riesgo.

Tanto la probabilidad de ocurrencia como el impacto de la materialización de los riesgos pueden tomar valores entre 0 y 1, siendo 0 algo nada probable o con ningún impacto, y 1 algo que ocurrirá muy probablemente o que muy probablemente tenga graves consecuencias. La prioridad se calculará como el producto de la probabilidad por el impacto.

| ID  | Riesgo                                                                     | Tipo        | Probabilidad | Impacto | Prioridad (P·I) |
|-----|----------------------------------------------------------------------------|-------------|--------------|---------|-----------------|
| R01 | Mala comprensión inicial del proyecto                                      | Amenaza     | 0.1          | 0.9     | 0.009           |
| R02 | Mala estimación de tiempos y problemas de organización                     | Amenaza     | 0.3          | 0.7     | 0.21            |
| R03 | Desarrollo técnico demasiado complejo                                      | Amenaza     | 0.4          | 0.7     | 0.28            |
| R04 | Falta de recursos <i>hardware</i> para la realización de experimentos      | Amenaza     | 0.2          | 0.8     | 0.16            |
| R05 | Dificultad en el tratamiento de datos                                      | Amenaza     | 0.1          | 0.7     | 0.007           |
| R06 | Construcción de modelos con resultados de mala calidad                     | Amenaza     | 0.5          | 0.8     | 0.40            |
| R07 | Escasez de recursos bibliográficos                                         | Amenaza     | 0.1          | 0.9     | 0.009           |
| R08 | Enfermedad/circunstancia personal que impida terminar a tiempo el proyecto | Amenaza     | 0.1          | 1       | 0.1             |
| R09 | Cambios en el alcance del proyecto                                         | Oportunidad | 0.4          | 0.6     | 0.24            |
| R10 | Problemas de escalabilidad                                                 | Oportunidad | 0.6          | 0.5     | 0.30            |

Tabla 2.10: Riesgos del proyecto

Ante la posibilidad de la materialización de los riesgos planteados en la tabla 2.10, se debe preparar un plan de acción para continuar con el desarrollo del proyecto y conseguir completarlo en el tiempo establecido. Es por esto que se ha preparado un plan de acción para cada uno de los riesgos planteados:

- **R01:** estudio inicial exhaustivo del estado del arte.
- **R02:** definición clara de todas las actividades a desarrollar a lo largo del trabajo, y ajustarse a los tiempos marcados en el cronograma.

- **R03:** estudio previo de la viabilidad de las ideas del proyecto.
- **R04:** préstamo por parte de los tutores de una GPU potente para el entrenamiento de los modelos.
- **R05:** aprendizaje de técnicas necesarias para la transformación de las señales fisiológicas, centrándose específicamente en el uso del bispectrum.
- **R06:** búsqueda de otros modelos similares que mejoren la calidad y limpieza de posible ruido en los datos.
- **R07:** búsqueda de referencias bibliográficas en libros especializados en las áreas a estudiar.
- **R08:** modificación de la planificación inicial.
- **R09:** un cambio en el alcance puede ser una oportunidad para ampliar el trabajo previsto en la planificación.
- **R010:** explorar nuevas arquitecturas y técnicas de optimización que permitan el uso eficiente de los recursos computacionales.



---

## 3: Contexto médico

---

En este capítulo exploran los conceptos fundamentales de la apnea del sueño y se analizan con detalle los distintos tipos de eventos que caracterizan esta enfermedad. Asimismo, se describe el funcionamiento de la polisomnografía nocturna, la prueba diagnóstica estándar para la apnea del sueño, y se explican las diversas señales fisiológicas monitorizadas durante esta prueba.

### 3.1. Apnea del sueño

La apnea del sueño es un trastorno respiratorio producido durante el sueño y caracterizado por la repetición de episodios de ausencia total (apnea) o parcial (hipopnea) del flujo aéreo en las vías respiratorias, con una duración mínima de 10 segundos en adultos [74]. La presencia de este tipo de eventos respiratorios provoca un sueño fragmentado y no reparador. Se estima que entre el 9 % y el 38 % de la población adulta [74] y entre el 1 % y el 5 % de los niños de todo el mundo [18] padecen algún grado de apnea del sueño. Esta prevalencia puede fluctuar según diversos factores como el sexo, la edad, la obesidad u otros trastornos médicos coexistentes. Además, esta enfermedad provoca cambios en los ciclos de sueño, fluctuaciones de la presión intratorácica, hipoxemia e hipercapnia [9], aumentando el riesgo de desarrollar enfermedades graves como problemas cardíacos, trastornos metabólicos, cáncer y enfermedades neurodegenerativas [44], [79], [92].

Dependiendo de la causa y origen de los eventos apneicos, se pueden distinguir tres tipos fundamentales: obstructivos, centrales y mixtos [9]. Los eventos apneicos obstructivos son los más habituales, se producen por un colapso en las vías respiratorias superiores, y están ocasionados por un tono muscular inadecuado en la lengua y en los músculos dilatadores de las vías respiratorias. Los eventos apneicos centrales están originados por un fallo en el sistema nervioso central, donde el cerebro deja de enviar las señales adecuadas a los pulmones para controlar el movimiento respiratorio. Por último, los eventos apneicos mixtos son una combinación tanto de eventos apneicos obstructivos, como centrales.

## Apnea obstructiva del sueño

La apnea obstructiva del sueño ([AOS](#)) es un trastorno respiratorio del sueño caracterizado por episodios repetidos de obstrucción total (apnea) o parcial (hipopnea) de las vías respiratorias durante el sueño [9]. Esta obstrucción provoca una respuesta del cuerpo en forma de excitación en el cerebro, activación simpática y desaturación de oxígeno en la sangre para reactivar los músculos de la garganta y recuperar el flujo aéreo. Esta respuesta del cuerpo provoca la aparición de microdespertares, haciendo que el sueño de las personas que lo sufren sea fragmentado y no reparador. Como consecuencia, estas personas pueden experimentar fatiga y cansancio, somnolencia diurna excesiva, insomnio y dolores de cabeza por la mañana, aunque muchas pueden ser asintomáticas [74].

Según la Academia Americana de Medicina del Sueño ([AASM](#), por sus siglas en inglés de *American Academy of Sleep Medicine*), una apnea se define como una reducción superior al 90 % de la señal de flujo aéreo ([FA](#)) durante al menos dos ciclos respiratorios, mientras que una hipopnea se define como una reducción del [FA](#) de entre el 30 % y el 90 % durante dos ciclos respiratorios, seguido de una caída de la saturación de oxígeno en sangre ( $SpO_2$ ) de al menos un 3 % o de una excitación electroencefalográfica [9].

La métrica fundamental para evaluar la severidad de la [AOS](#) es el Índice de Apnea-Hipopnea ([IAH](#)), que representa el número medio de eventos apnéicos obstructivos por hora de sueño. La [AASM](#) define cuatro grupos de severidad en los que se pueden clasificar los individuos adultos según su [IAH](#): no apnea ( $IAH < 5$  e/h), leve ( $5 \text{ e/h} \leq IAH < 15$  e/h), moderada ( $15 \text{ e/h} \leq IAH < 30$  e/h) y grave ( $IAH \geq 30$  e/h).

Se han realizado múltiples estudios acerca de la prevalencia de esta enfermedad en la población, puesto que, cada vez más, se reconoce como una causa importante de morbilidad y mortalidad, con numerosas consecuencias para la salud de los pacientes. Es importante destacar que la estimación de la prevalencia de esta enfermedad es compleja, debido a los diferentes esquemas de muestreo empleados, las disparidades de las técnicas empleadas para la monitorización del sueño y la respiración de los pacientes y la variabilidad en las definiciones y criterios médicos para categorizar y detectar los eventos apnéicos [74].

En algunos estudios como el realizado por Senaratna *et al.* [74] se estima que la prevalencia de la [AOS](#) en la población adulta es de entre el 9 % y el 38 %, resultando ser más elevado este porcentaje en hombres que en mujeres e incrementándose con la edad. Además, esta proporción es superior en hombres y mujeres obesos, en comparación con hombres y mujeres con sobrepeso. La prevalencia en la población con una edad avanzada es muy alta. En el caso de un  $IAH \geq 5$  e/h, en hombres entre 65 y 69 años, es del 88 %, mientras que para hombres entre 60 y 85 años, es del 90 %. En el caso de las mujeres, para los mismos rangos de edad, la prevalencia fue del 66 % y del 78 %, respectivamente. Para un nivel de severidad moderado (más de 15 e/h), la prevalencia en la población adulta se encuentra entre el 6 % y el 17 %, mientras para que los grupos de edad más avanzados llega a situarse alrededor del 49 %.

En otros estudios como el llevado a cabo por Young *et al.* [92], se analiza el caso concreto de la prevalencia de la [AOS](#) en la población española. Para un rango de edad de

entre 30 y 70 años, un 26 % de los hombres y un 28 % de las mujeres estudiadas presentaron un  $\text{IAH} \geq 5$  e/h y un 14 % de los hombres y un 7 % de las mujeres presentaron un  $\text{IAH} \geq 15$  e/h. En este estudio también se remarca la influencia de la edad en la prevalencia de esta enfermedad, en concreto se destaca la ciudad de Vitoria-Gasteiz, en la que la diferencia de la prevalencia entre la población de mediana y de mayor edad fue más grande. Para edades comprendidas entre 71 y 100 años, un 81 % de los hombres y un 80 % de las mujeres presentan un  $\text{IAH} \geq 5$  e/h, y un 57 % de hombres y un 49 % de mujeres, un  $\text{IAH} \geq 15$  e/h.

Otros múltiples estudios han analizado la prevalencia de la  $\text{AOS}$  en diversos grupos poblacionales, en los que se han podido descubrir cuáles son los factores de riesgo más importantes para el desarrollo de esta enfermedad, como pueden ser: la edad, en la que la prevalencia crece a medida que avanza la edad; el sexo, siendo más frecuente en hombres que en mujeres; exceso de peso, las personas con un índice de masa corporal más alto ( $\text{IMC}$ ) son más propensos a padecer esta enfermedad; anatomía craneofacial, algunas estructuras esqueléticas y tejidos blandos pueden estar relacionados con una mayor propensión a sufrir  $\text{AOS}$ ; predisposición genética; consumo de alcohol y tabaco; presencia de otras enfermedades, principalmente cardiovasculares como la hipertensión, enfermedades de las arterias coronarias, insuficiencia cardíaca congestiva o accidentes cerebrovasculares [66].

La  $\text{AOS}$  no tratada adecuadamente constituye un factor de riesgo para múltiples patologías, ampliamente reconocidas como principales causas de mortalidad en adultos, entre las que se incluyen enfermedades cardiovasculares, como los eventos cardiovasculares fatales y no fatales, infarto agudo de miocardio, así como un mayor riesgo de taquiarritmias complejas, fibrilación auricular y taquicardia ventricular no sostenida [63] [92].

Además, la  $\text{AOS}$  está relacionada con alteraciones en el sistema nervioso central, lo que puede provocar deterioro cognitivo, hipersomnolencia y fatiga, así como accidentes cerebrovasculares con riesgo de muerte prematura. También se ha vinculado la  $\text{AOS}$  con enfermedades como la epilepsia, en la que un tratamiento adecuado de la  $\text{AOS}$  podría mejorar el control de las convulsiones en estos pacientes.

Por otro lado, la  $\text{AOS}$  puede ser un factor de riesgo para la resistencia a la insulina y el desarrollo de diabetes, ya que los pacientes con  $\text{AOS}$  moderada o grave tienen más probabilidades de presentar niveles elevados de glucosa en ayunas y una peor tolerancia a la glucosa a las dos horas. Por último, la  $\text{AOS}$  puede provocar complicaciones perioperatorias, incluyendo problemas en las intubaciones, depresión respiratoria excesiva inducida por anestésicos y analgésicos, mayor probabilidad de reintubaciones postoperatorias, arritmias cardíacas y estancias hospitalarias más prolongadas [63].

El tratamiento estándar más comúnmente utilizado para la  $\text{AOS}$  es la terapia con la máquina de presión positiva continua en las vías respiratorias ( $\text{CPAP}$ , por sus siglas en inglés: *Continuous Positive Airway Pressure*) [64]. La máquina de  $\text{CPAP}$  mantiene abiertas las vías respiratorias superiores aplicando una presión de aire constante a través de una mascarilla que se coloca sobre la nariz o la nariz y la boca, previniendo los bloqueos en la respiración durante el sueño. La  $\text{CPAP}$  permite reducir de forma significativa el  $\text{IAH}$  de los pacientes y la somnolencia diurna, especialmente en los casos más severos [64]. Además,

algunos estudios han demostrado que el uso de CPAP puede ayudar a mejorar el control de la presión arterial y a reducir la fibrilación auricular, además de disminuir la morbilidad y mortalidad cardiovascular en menores de 65 años [63].

En cambio, el uso de CPAP no siempre funciona correctamente para todos los pacientes. Algunos rechazan su uso por la incomodidad de la mascarilla, sensación de claustrofobia o intolerancia a la presión, que les genera sequedad e irritación en las membranas nasales y faríngeas, congestión nasal o irritación ocular [64]. Es por esto que existen otros tratamientos a la AOS, como puede ser la BiPAP (presión positiva en las vías respiratorias a dos niveles), con el que se administra al paciente presiones variables dependiendo de si inhala o exhala [64]. Es una modalidad de ventilación no invasiva, que está indicada para pacientes con hipoventilación por obesidad, trastornos pulmonares obstructivos o hipoventilación alveolar en sujetos con trastornos como la esclerosis lateral amiotrófica (ELA). Otros posibles tratamientos son la máquina de presión positiva automática (APAP), la terapia con aparatos bucales de retención lingual y de avance mandibular e incluso las intervenciones quirúrgicas como la extracción de las amígdalas (adenomigdalectomía), cirugías nasales, del paladar, de la lengua o de la mandíbula (avance maxilomandibular) [64].

Todos estos tratamientos van acompañados de consejos de pérdida de peso para pacientes con sobrepeso u obesidad, con el objetivo de reducir el IMC a valores inferiores a  $25 \text{ kg/m}^2$ . También se recomienda a los pacientes dormir en una posición no supina, para aumentar la permeabilidad de las vías respiratorias. Por ello, habitualmente se recomienda el uso de un dispositivo para corregir la postura del paciente y evitar dormir en posición supina [64].

## Apnea central del sueño

La apnea central del sueño (ACS) es un trastorno respiratorio del sueño caracterizado por una ausencia de esfuerzo ventilatorio, que puede estar causado por varios factores, como el estado de sueño, bajos niveles temporales de dióxido de carbono en la sangre (hipocapnia transitoria), un sistema de control respiratorio inestable, reflejos débiles en las vías respiratorias superiores y la disminución de la capacidad del cerebro para generar los impulsos necesarios para controlar el ritmo de la respiración de manera regular y adecuada [3]. Al igual que la AOS, está relacionada con algunos problemas como los despertares nocturnos frecuentes, la somnolencia diurna excesiva y un mayor riesgo de problemas cardiovasculares [21].

La ACS puede estar causada por hipoventilación o postventilación, es decir, no tomar suficiente aire al respirar, o exhalar demasiado [3]. Cuando este impulso respiratorio se produce durante el sueño, especialmente en personas con problemas respiratorios, como enfermedades neuromusculares, puede producirse hipoventilación o eventos apnéicos. Además, algunos analgésicos opioides pueden contribuir a esta hipoventilación. La ACS es más frecuente durante la fase de sueño de movimientos oculares no rápidos (NREM), puesto que la respiración se vuelve más dependiente de señales químicas en lugar de impulsos cerebrales, es decir, si los niveles de presión parcial de dióxido de carbono en sangre ( $p\text{CO}_2$ )

se reducen por debajo de cierto umbral, la respiración puede llegar a detenerse de forma temporal [3].

Existen diversas manifestaciones de ACS, como la respiración periódica inducida por la altitud, la ACS idiopática, la ACS inducida por narcóticos, el síndrome de hipoventilación por obesidad y la respiración de Cheyne-Stokes [21]:

- **Respiración periódica inducida por la altitud:** es un patrón de respiración irregular durante el sueño en entornos de gran altitud, en los que la cantidad de oxígeno disponible es menor debido a la disminución de la presión barométrica, provocando hipoxia hipobárica [71]. Una exposición sostenida a esta hipoxia provoca diversos efectos adversos sobre el peso de las personas, su estructura muscular, el funcionamiento mental y la calidad del sueño. La hipoxia estimula la hiperventilación para incrementar la cantidad de oxígeno en el cuerpo, pero, al mismo tiempo, esta hiperventilación provoca un aumento de la concentración de dióxido de carbono en sangre, lo que puede interrumpir el impulso de respirar, y únicamente se reinicia la respiración cuando se produce una nueva caída en los niveles de oxígeno en sangre [71].
- **Síndrome de ACS idiopática (ICSAS):** es un trastorno poco común en el que las personas experimentan apneas centrales recurrentes durante el sueño sin tener insuficiencia ventilatoria, insuficiencia cardíaca o enfermedades neuromusculares [89]. Este síndrome se caracteriza por una hiperventilación crónica y aguda, lo que significa que las personas tienden a respirar más rápido y profundamente de lo normal, a veces llevando los niveles de dióxido de carbono por debajo de un umbral crítico, y provocando el desencadenamiento de los eventos apneicos centrales [89].
- **ACS inducida por narcóticos y otras terapias:** el uso a largo plazo de medicamentos a base de opioides puede conducir a una mayor propensión de padecer ACS hasta en un 50 % de los pacientes [21]. Una reducción de estas dosis de medicación a base de opioides pueden reducir la ACS inducida por altas dosis de narcóticos [21]. Además, como ya se ha comentado previamente, algunos pacientes de AOS no toleran el tratamiento con la máquina de CPAP, por diversos motivos. Recientes estudios sugieren que el uso de CPAP en algunos pacientes puede provocar la aparición de eventos apneicos de carácter central durante el sueño. Esto ocurre cuando la presión aplicada por la CPAP es demasiado elevada, provocando una expansión excesiva de los pulmones. Esta sobreexpansión activa los receptores de estiramiento en los pulmones, lo que a su vez estimula las fibras nerviosas del nervio vago, causando interrupciones en la respiración [41] [43].
- **Síndrome de hipoventilación por obesidad:** este trastorno se puede definir como una combinación de obesidad e hipercapnia arterial en estado de vigilia, en ausencia de otras causas conocidas de hipoventilación [62]. Quienes padecen esta enfermedad presentan síntomas como somnolencia diurna excesiva, fatiga o dolores de cabeza por las mañanas, síntomas similares a los que padecen las personas con

**AOS.** En cambio, los pacientes con hipoventilación por obesidad también presentan hipercapnia e hipoxemia durante el día, habitualmente asociado con la hipertensión pulmonar y la insuficiencia cardíaca congestiva del lado derecho (*cor pulmonale*) [62].

- **Respiración de Cheyne-Stokes:** la respiración de Cheyne-Stokes con apnea central del sueño es un tipo de respiración periódica caracterizada por patrones crecientes y decrecientes del volumen corriente respiratorio [93]. Este tipo de respiración es típico de algunas enfermedades como la insuficiencia cardíaca, y habitualmente está ocasionado por una inestabilidad del sistema de control respiratorio caracterizada por una tendencia a hiperventilar. Los pacientes con este tipo de respiración a menudo presentan síntomas de ortopnea, disnea paroxística nocturna y somnolencia diurna excesiva. Además, este tipo de respiración está asociado con un mayor riesgo de mortalidad [57] [93].

La prevalencia de **ACS** varía considerablemente entre los diferentes tipos de **ACS**. La mayoría de personas sanas, cuando ascienden a grandes altitudes, experimentan una respiración periódica si la magnitud del ascenso es suficiente para causar una hipoxia alveolar sustancial. La prevalencia de **ACS** para el síndrome de hipoventilación por obesidad va en aumento y se espera que siga creciendo, debido al aumento global de la obesidad. Por otro lado, la **ICSAS** es relativamente poco común, y cuenta con una prevalencia inferior al 5 % [21]. Resulta importante destacar que se considera que el diagnóstico principal es **ACS** cuando al menos el 50 % de los eventos apneicos se clasifican como de origen central; sino, el diagnóstico sería de **AOS** [9].

La prevalencia de la **ACS** también varía dependiendo del grupo de población al que nos referimos. En [32] se muestra que la prevalencia de la **ACS** en pacientes varones con insuficiencia cardíaca sistólica es del 37 %. Otros estudios como [10] demuestran que la prevalencia de la **ACS** aumenta con la edad, pasando del 0.4 % para grupos de mediana edad, al 1.1 % en grupos de edades más avanzadas. Por último, la prevalencia de la **ACS** y de la respiración de Cheyne-Stokes en la base de datos de Sleep Heart Health Study (**SHHS**) es del 0.9 % y del 0.4 %, respectivamente. Las personas que padecen **ACS** son más mayores, con un **IMC** más bajo, puntuaciones también bajas en la escala de somnolencia de Epworth y principalmente hombres. En el caso de personas con insuficiencia cardíaca, la prevalencia de la **ACS** es más alta, llegando hasta el 4.1 % [20].

Existen diferentes factores fisiológicos que pueden hacer que un paciente sea más susceptible de desarrollar apnea central, entre los que se encuentran las fases del sueño, la edad, el sexo y diversas afecciones médicas. La transición entre la vigilia y el sueño **NREM** puede estar asociada con la **ACS** [3]. Esto se debe a una excitación transitoria que provoca una breve hiperventilación, y la hipocapnia resultante desencadena un evento apneico central al reanudarse el sueño. La **ACS** es más prevalente en adultos en comparación con personas de mediana edad [3]. Esta mayor prevalencia también puede deberse a una mayor incidencia de afecciones comórbidas como hipotiroidismo, insuficiencia cardíaca congestiva y fibrilación auricular en adultos mayores. Además, los hombres son más susceptibles al desarrollo de **ACS** durante la ventilación mecánica en comparación con las mujeres. Por



último, las personas afectadas por diferentes enfermedades como la insuficiencia cardíaca, hipotiroidismo, insuficiencia renal o aterosclerosis carotídea asintomática, y personas con alguna lesión en la médula espinal o que recientemente han sufrido un accidente cerebrovascular, también son mucho más propensas a padecer [ACS](#) [3].

El tratamiento de la [ACS](#) depende de las patologías y del resultado de la prueba diagnóstica de forma individualizada para cada paciente, pero los más comunes incluyen las terapias de presión positiva, la estimulación del nervio frénico y la terapia farmacológica, entre la que se encuentra la administración de oxígeno suplementario de bajo flujo [3].

## Apnea mixta del sueño

La apnea mixta del sueño, también conocida como síndrome de la apnea del sueño compleja, es un trastorno respiratorio que ocurre durante el sueño. En este trastorno, se presentan eventos apneicos centrales después de que los eventos apneicos obstructivos han sido eliminados mediante terapia de presión positiva en las vías respiratorias [36]. Para que se considere apnea mixta del sueño, más de la mitad de los eventos apneicos residuales deben ser centrales o deben dar lugar a un patrón de respiración periódica. Con la terapia de presión positiva en las vías respiratorias, estos eventos centrales se vuelven predominantes, con más de 5 eventos centrales por hora. El uso de terapias con ventilación de dos niveles con presión positiva en las vías respiratorias también puede aumentar el riesgo de provocar eventos apneicos centrales [36].

La prevalencia de la apnea mixta del sueño oscila entre un 0.56 % y un 18 %, dependiendo de los centros en los que se haya realizado el estudio y las poblaciones heterogéneas que se han analizado y que se ven afectadas por diferentes factores como el uso de narcóticos, el [IMC](#) y otras comorbilidades [36].

Las características observables de los pacientes con apnea mixta son muy similares a los de un paciente con [AOS](#), y solo se pueden diferenciar cuando se aplica una terapia con [CPAP](#). Los pacientes con síndrome de apnea del sueño compleja son frecuentemente hombres, con un [IMC](#) más alto que el de la población general, y con una mayor incidencia de cardiopatías isquémicas o insuficiencia cardíaca en comparación con los pacientes con [AOS](#). El tratamiento más habitual para este tipo de apnea es el uso de dispositivos de presión positiva para reducir el número de eventos obstructivos, aunque, como ya se ha comentado previamente, en ocasiones el uso de [CPAP](#) también puede provocar el incremento del número de eventos apneicos de tipo central. Para estos pacientes, se puede tratar de reducir su número de eventos apneicos centrales a través de dispositivos que controlan la presión positiva de las vías respiratorias en dos niveles y la servoventilación adaptativa [36].

## 3.2. Polisomnografía

La prueba diagnóstica de referencia para la apnea del sueño es la polisomnografía nocturna ([PSG](#)). Esta prueba recopila múltiples parámetros sobre la respiración, la

actividad cerebral, la frecuencia cardíaca, la actividad eléctrica cardíaca y otras constantes vitales durante el sueño [70]. La **PSG** se usa frecuentemente para diagnosticar y cuantificar la gravedad de la apnea del sueño, y analizar el volumen de aire de la presión positiva en las vías aéreas en el tratamiento de los trastornos respiratorios relacionados con el sueño. También se puede emplear en la evaluación de otros trastornos del sueño como las hipersomnias del sistema nervioso central (narcolepsia e hipersomnia idiopática) y las parasomnias (el trastorno de la conducta durante el sueño con movimientos oculares rápidos, **REM**), así como en la evaluación de los movimientos periódicos de las extremidades durante el sueño [46].

Para asegurar que los resultados de las **PSG** sean consistentes en todos los laboratorios de sueño, la **AASM** ha creado el *Manual for Scoring of Sleep and Associated Events* [9]. Este manual establece criterios universales para registrar el sueño y puntuarlo. Divide el sueño en fases de 30 segundos con definiciones detalladas para cada una, como el estado de vigilia o diferentes fases del sueño [46]. Los datos fundamentales que se recogen durante la **PSG** nocturna incluyen las fases de sueño, flujo de aire y esfuerzo respiratorio, saturación de oxígeno, frecuencia y ritmo cardíaco, movimientos periódicos de las extremidades y la posición del cuerpo [70].

- **Fases del sueño:** las señales procedente de los electroencefalogramas (**EEG**), electro-oculogramas (**EOG**) y electromiogramas (**EMG**) permiten determinar las diferentes fases del sueño y vigilia analizando la actividad cerebral de las regiones frontal, central y occipital del cerebro, el movimiento rápido de ojos y el electromiograma submentoniano: despierto (W), etapa 1 (N1), etapa 2 (N2), etapa 3 (N3) y sueño **REM** [70]. En el **EEG**, las señales se recogen a través de electrodos que se sitúan en la cabeza del paciente siguiendo el Sistema Internacional 10-20, que se muestra en la figura 3.3. Se recomienda situar los electrodos en F4, C4, O2 y M1 para obtener las derivaciones F4-M1, C4-M1 y O2-M1, aunque también en las ubicaciones correspondientes del lado izquierdo para obtener las derivaciones F3-M2, C4-M2 y O1-M2, en caso de fallo de los electrodos durante el estudio. En el **EMG** se coloca un electrodo en el centro de la mandíbula, un centímetro por encima de su borde inferior, y dos electrodos colocados dos centímetros por debajo del borde de la mandíbula, a dos centímetros a la izquierda y a la derecha del centro. En el **EOG** se sitúa un electrodo un centímetro por debajo y un centímetro a la izquierda o derecha del canto lateral de cada ojo (electrodos E1 y E2) [9]. Cambios abruptos en las frecuencias del **EEG** o frecuencias mayores de 16 Hz durante al menos 3 segundos, permiten detectar despertares en las etapas N1, N2, N3 o **REM**. Además, para detectar un despertar en la fase **REM**, se debe producir durante al menos un segundo un aumento de la frecuencia en el **EMG** que acompañe al cambio de frecuencia en el **EEG**. Resulta de gran importancia conocer los despertares que sufre el paciente durante el sueño para poder evaluar posteriormente algunos eventos respiratorios.
- **Flujo de aire y esfuerzo respiratorio:** para medir el flujo de aire durante la **PSG**, se emplean dos sensores cualitativos del flujo de aire: el sensor termal oronasal



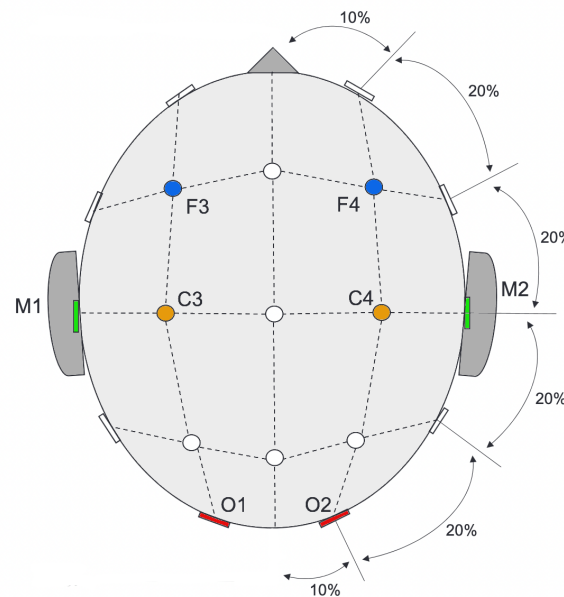


Figura 3.3: Sistema 10-20 de posicionamiento de electrodos en EEG (adaptado de [70])

(termistor), que detecta cambios de temperatura entre la respiración inspirada y espirada, y el transductor de presión nasal, que detecta cambios de presión durante la inspiración y espiración [70]. El primero es más efectivo para la identificación de los eventos apneicos, mientras que el segundo se emplea para identificar las hipopneas. Por otro lado, para determinar el esfuerzo respiratorio torácico y abdominal se emplean dos bandas de pletismografía inductiva respiratoria (RIP). Estas bandas o cinturones emiten cambios en el voltaje correspondientes al movimiento del tórax y el abdomen durante la respiración para determinar los diferentes tipos de eventos respiratorios, incluyendo apneas obstructivas, hipopneas obstructivas y apneas centrales.

- **Saturación de oxígeno en sangre:** la saturación de oxígeno en sangre se obtiene a través de un oxímetro de pulso, y permite puntuar las hipopneas cuando se produce una desaturación de oxígeno entre el 3 % y el 4 % asociada a un evento respiratorio [70]. Además, esta señal de oximetría permite determinar si se requiere administrar un suplemento nocturno de oxígeno.
- **Frecuencia y ritmo cardíaco:** para la monitorización de la frecuencia y el ritmo cardíaco se emplea un ECG, colocando electrodos en el torso del paciente. Con este ECG no siempre se pueden definir las arritmias complejas, en cambio, sí resultan fácilmente reconocibles enfermedades como la taquicardia, la fibrilación auricular, el bloqueo cardíaco significativo y la taquicardia o bradicardia sostenidas [70].
- **Movimientos periódicos de las extremidades:** los movimientos periódicos de las extremidades se pueden medir a través de un EMG, utilizando electrodos colocados en los músculos tibial anterior izquierdo y derecho para detectar el movimiento de las piernas [70]. Estos movimientos se puntúan en base a un aumento mínimo de la

amplitud de 8 mV por encima del voltaje habitual en reposo, con una duración de entre 0.5 y 10 segundos. Para clasificar un movimiento como periódico de las extremidades se requieren al menos cuatro movimientos consecutivos de las extremidades, con una separación de entre 5 y 90 segundos entre cada serie de movimientos de las extremidades.

- **Posición del cuerpo:** la posición del paciente durante el sueño se registra a través de un vídeo o de un monitor de posición conectado al paciente. Esta medición resulta muy importante, especialmente en pacientes con AOS, porque los eventos obstructivos tienden a ser más frecuentes con una posición supina [70].

A pesar de ser la prueba más adecuada para el diagnóstico de múltiples enfermedades del sueño, requiere que el paciente pase una noche entera en el hospital, un entorno fuera de lo habitual que puede alterar los patrones de sueño de los pacientes y, por lo tanto, también los resultados de la prueba. Además, durante toda la noche, el paciente debe estar acompañado por un especialista del sueño y todas las señales fisiológicas recogidas deben ser analizadas posteriormente por un médico especializado, haciendo que esta prueba sea altamente costosa, compleja y requiera mucho tiempo [8] [22]. Es por esto que la comunidad científica ha iniciado la búsqueda de nuevas técnicas de diagnóstico más sencillas y menos costosas. Una de ellas es la polisomnografía doméstica, una prueba en la que se recoge un menor número de señales fisiológicas durante la noche de sueño del paciente, y que se lleva a cabo en su propio domicilio, ofreciendo un entorno más cómodo y adecuado para obtener datos más representativos que en un entorno hospitalario. Además, esta prueba es más económica que la PSG habitual, es más accesible para un mayor número de pacientes y está demostrado que no genera peores resultados que una PSG [70].

La polisomnografía doméstica (HSAT, por sus siglas en inglés) es una alternativa a la PSG en el diagnóstico de la AOS, pero se debe emplear únicamente en pacientes con un riesgo moderado o alto de padecer AOS, y sin afecciones médicas comórbidas ni otros trastornos del sueño, como pueden ser: enfermedades cardiorrespiratorias significativas, debilidad en los músculos respiratorios debido a enfermedades neuromusculares, consumo crónico de opiáceos, alto riesgo de hipoventilación durante el sueño, reciente accidente cerebrovascular, insomnio severo o alguna razón personal o ambiental que pueda conducir a una mala interpretación de los resultados [70]. Tampoco se recomienda para pacientes mayores de 65 años o con un IMC  $> 40 \text{ kg/cm}^2$ . La tecnología empleada en la HSAT es similar y emplea el mismo tipo de sensores que una PSG, aunque no requiere tantos sensores ni registra el EEG. También incluye sensores de posición corporal, frecuencia cardíaca y evaluación del movimiento en sustitución del EEG [70].

---

## 4: Contexto tecnológico

---

Podemos definir la inteligencia como la capacidad de resolver un problema a partir de diversas habilidades y conocimientos. La inteligencia también se refiere al uso de las capacidades mentales de las personas para resolver, razonar y aprender en múltiples situaciones. Algunas de las funciones cognitivas que integran la inteligencia son: el lenguaje, la atención, la planificación, la memoria y la percepción. Por otro lado, la Inteligencia Artificial (IA) se conoce como un conjunto de algoritmos que tiene la capacidad de imitar la inteligencia humana y realizar diversas tareas que requieren pensar, aprender, resolver problemas y tomar decisiones [75].

En el año 1926, Alan Turing publicó un artículo llamado *On Computable Numbers, with an Application to the Entscheidungsproblem* [82], en que presenta la idea de una máquina teórica capaz de realizar cualquier tipo de cálculo que pudiera describirse en forma de algoritmo. Durante la II Guerra Mundial, desarrolló la máquina Enigma, una máquina capaz de descifrar las comunicaciones de los nazis. Esta máquina no solo demostró la capacidad de las máquinas para realizar tareas complejas de criptografía, sino que también sentó las bases para el desarrollo futuro de la IA.

En el año 1943, Warren McCulloch y Walter Pitts [51] crearon el primer modelo neuronal moderno, que se ha empleado como punto de partida para el desarrollo de múltiples modelos neuronales posteriores. Este modelo matemático de neurona trata de simular el comportamiento de una neurona cerebral, y se basó en la idea de que las neuronas biológicas pueden ser representadas como unidades computacionales simples que reciben entradas, las procesan y generan una salida en función de ciertas reglas de activación.

Posteriormente, en 1956, se celebró la Conferencia de Dartmouth impulsada principalmente por John McCarthy y Marvin Minsky. Esta fue la primera conferencia dedicada únicamente a la IA, y sentó las bases de la IA que conocemos hoy en día. El propio John McCarthy fue el responsable de introducir el término de *inteligencia artificial* en 1955 como “la ciencia y la ingeniería de fabricación de máquinas inteligentes especialmente programas informáticos inteligentes”, en el que la IA está relacionada con la tarea de

utilizar los ordenadores para comprender la inteligencia humana sin limitarse a métodos biológicamente observables [50].

A lo largo de los años y hasta la actualidad, la IA se ha ido desarrollando enormemente, y múltiples nuevos campos de investigación han surgido a su alrededor. En la figura 4.4 se muestra un diagrama de Venn con los tres campos fundamentales que se han desarrollado a partir de la inteligencia artificial: *machine learning*, redes neuronales artificiales y *deep learning*.

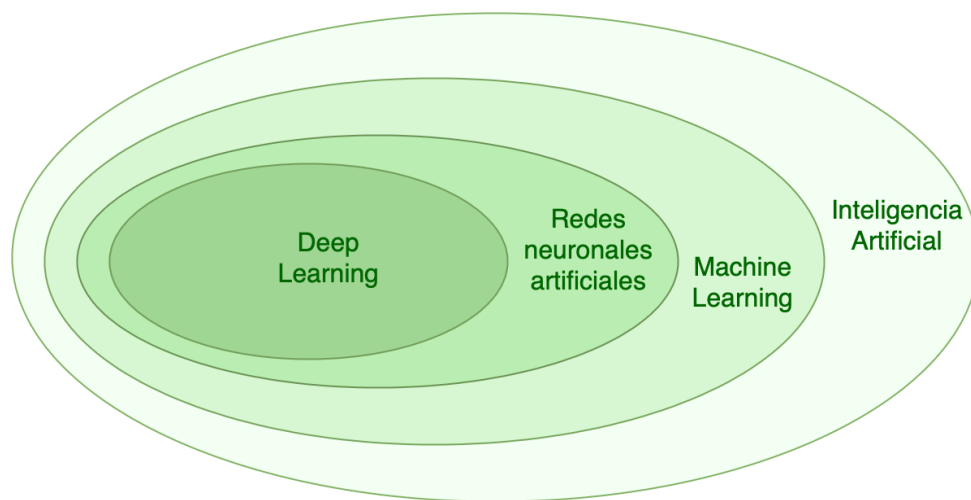


Figura 4.4: Diagrama de Venn que muestra la relación entre la IA y los campos fundamentales de la IA

El aprendizaje automático o *machine learning* se centra en el desarrollo de modelos y algoritmos que permitan a los ordenadores aprender patrones y realizar tareas específicas sin ser programados explícitamente para cada tarea, es decir, los algoritmos de *machine learning* son capaces de aprender de los datos disponibles para realizar predicciones o tomar decisiones basadas en ellos. En lugar de seguir instrucciones programadas, estos algoritmos analizan grandes conjuntos de datos e identifican los patrones más relevantes para tomar decisiones. Algunos de los algoritmos de aprendizaje automático más conocidos son los modelos de regresión lineal o regresión logística, los árboles de decisión o los *random forests* o los algoritmos de *clustering* como k-medias o el *clustering* jerárquico.

Las redes neuronales artificiales son modelos computacionales basados en el funcionamiento del cerebro humano. Constan de un conjunto de neuronas artificiales conectadas entre sí, que tratan de generar una predicción a partir de unos datos de entrada. Estas redes son capaces de aprender patrones y características complejas de los datos a través de un proceso de entrenamiento. El concepto de neurona en estas redes proviene del ya mencionado primer modelo neuronal moderno de McCulloch y Pitts. Cada neurona realiza un cálculo basado en una combinación lineal de las entradas ponderadas por unos pesos, seguido de una función de activación.

Dentro del conjunto de redes neuronales artificiales existe un subconjunto que se denomina redes neuronales profundas (o *deep learning*). Estos modelos son mucho más

profundos que las redes neuronales artificiales clásicas, lo que significa que tienen un gran número de capas ocultas entre la capa de entrada y la capa de salida de la red. Una mayor profundidad de estas redes permite aprender representaciones de los datos más complejas, haciendo que sean altamente efectivas en tareas de reconocimiento de patrones en conjuntos de datos de grandes dimensiones. Algunos de los ejemplos más conocidos de los modelos de *deep learning* son las redes convolucionales (CNN) y las redes recurrentes (RNN).

Una de las principales diferencias entre las redes de aprendizaje profundo y los modelos de aprendizaje automático es que los segundos a menudo requieren un procesamiento previo de los datos y una extracción manual de características relevantes para su entrenamiento, mientras que las redes de aprendizaje profundo tienen la capacidad de aprender automáticamente las características y patrones directamente de los datos de entrada, sin necesidad de un procesamiento previo. Esto permite trabajar con conjuntos de datos de una alta dimensionalidad de forma más eficaz que los modelos clásicos de aprendizaje automático. En la figura 4.5 se muestra de forma gráfica esta diferencia entre los modelos de aprendizaje automático y los de aprendizaje profundo.

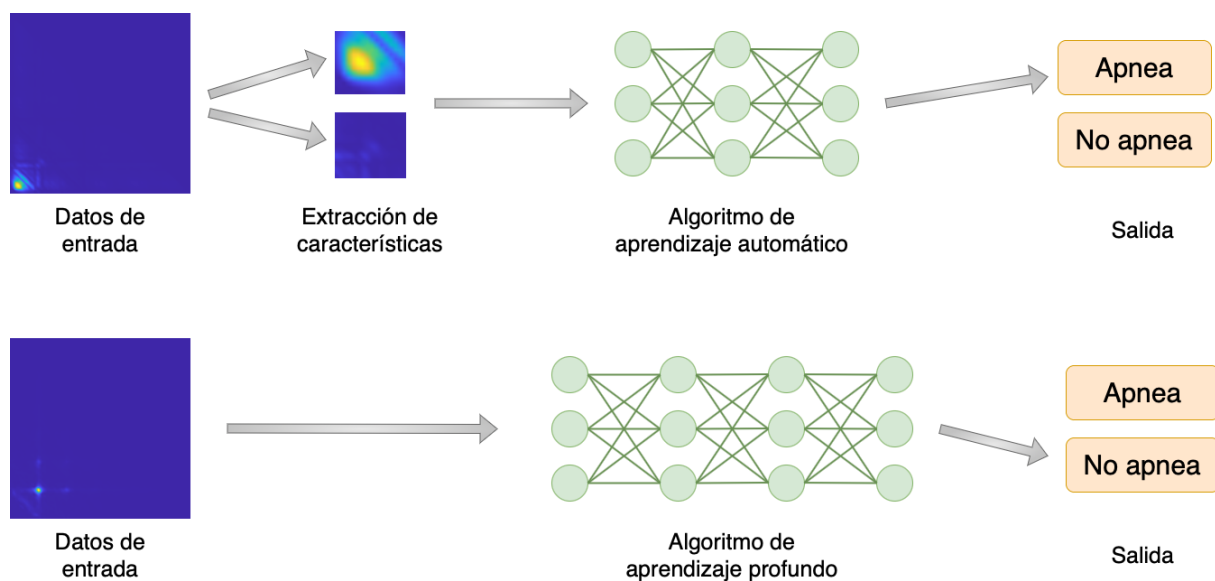


Figura 4.5: Comparación entre *machine learning* y *deep learning*

Dentro de los sistemas de aprendizaje automático, existen diversos métodos de aprendizaje. Uno de los métodos más comunes es el aprendizaje supervisado, en el que el modelo se entrena a partir de datos etiquetados, es decir, datos con entrada y salida esperada. Los métodos más comunes de aprendizaje supervisado son los de regresión y clasificación. En cambio, el aprendizaje no supervisado, entrena los modelos con datos no etiquetados, con el fin de encontrar patrones y estructuras subyacentes en los datos sin ningún tipo de ayuda para el modelo en forma de salida esperada. Por otro lado, el aprendizaje semi supervisado combina ambos enfoques, puesto que usa datos tanto etiquetados como no etiquetados para el entrenamiento de los modelos, normalmente una pequeña cantidad de

los datos están etiquetados, pero la mayoría no están etiquetados. Los datos no etiquetados ayudan a mejorar el rendimiento del modelo y la exactitud del aprendizaje. Por último, el aprendizaje por refuerzo consiste en un agente que interactúa con el entorno y toma decisiones para maximizar la recompensa obtenida [84]. Este tipo de modelos aprende a través de pruebas y errores, ajustando sus acciones en función de las recompensas obtenidas. Buscan alcanzar un objetivo específico sin haber sido programados explícitamente para ello y sin especificar cómo lograrlo.

En este trabajo, los modelos de IA que se van a emplear son los más complejos que se han descrito: los modelos de *deep learning*, en concreto, se van a emplear arquitecturas de redes neuronales convolucionales para la predicción de la severidad de la apnea del sueño. Para ello, se va a plantear un problema de aprendizaje supervisado, y se van a explorar tanto los problemas de regresión como los de clasificación.

## 4.1. Redes neuronales convolucionales

Las redes neuronales convolucionales (CNN) son un tipo de arquitectura de red neuronal diseñada para el procesamiento eficiente de datos matriciales, como series temporales (matrices 1D), imágenes o espectrogramas (matrices 2D) o, incluso, vídeos o imágenes volumétricas (matrices 3D). Su nombre proviene de la operación matemática que emplean, conocida como “convolución”: un tipo especializado de operación lineal [24]. Esta arquitectura se asemeja a las conexiones neuronales existentes en el cerebro humano, y se inspira en la organización de la corteza visual, lo que facilita la identificación de características relevantes en los datos.

Esta arquitectura de CNN fue introducida por Yann Le Cun en 1989 en [38]. El objetivo de la red presentada en este artículo era el reconocimiento de dígitos escritos a mano para diferenciar códigos postales escritos manualmente en las oficinas de correos de Estados Unidos. Para ello, empleó la técnica de retropropagación del error para el entrenamiento de una red neuronal. Esta red neuronal, conocida como LeNet-1 fue sufriendo modificaciones y mejoras hasta la versión LeNet-5, presentada en 1989 [39]. Esta arquitectura consta de siete capas: tres de ellas convolucionales, dos de submuestreo y dos capas *fully-connected* (FC). Las capas convolucionales emplean un *kernel* de tamaño 5x5 y *stride* 1 para extraer las características relevantes de las imágenes. Las capas de submuestreo reducen el tamaño de las imágenes con un *kernel* de tamaño 2x2 y *stride* 2. Por último, las capas *fully-connected* combinan las características extraídas en un formato adecuado para la clasificación final. En la figura 4.6 se muestra la arquitectura de la red LeNet-5 descrita.

### Operación de convolución

Una convolución es una operación realizada sobre unos datos de entrada a los que se aplica un filtro, también llamado *kernel*, que da como resultado una característica [24]. Es una operación realizada sobre dos funciones ( $x$  y  $w$ ) para generar una tercera ( $x * w$ ) y

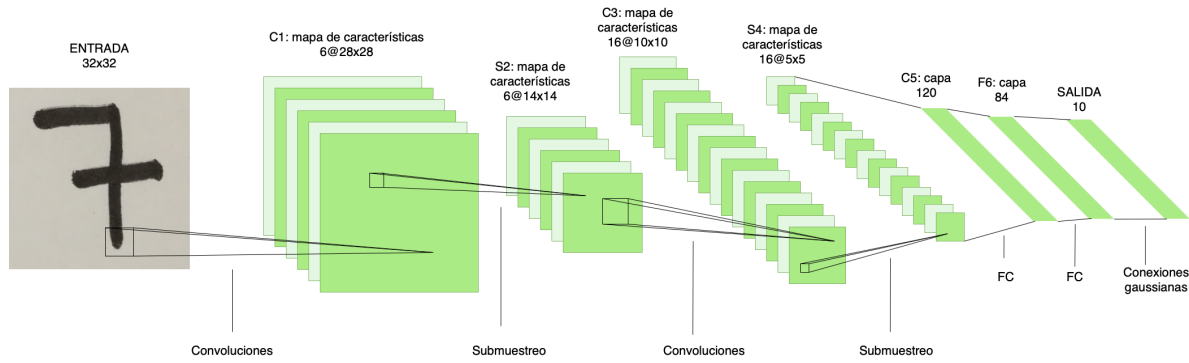


Figura 4.6: Arquitectura LeNet-5 propuesta en [39] para el reconocimiento de dígitos

se define como la integral del producto de las dos funciones tras invertir y desplazar a lo largo del eje  $t$  una de ellas. Matemáticamente, se expresa como [24]:

$$s(t) = (x * w)(t) = \int x(a)w(t - a)da \quad (4.1)$$

Donde,

- $a$ : es una variable de integración.
- $t$ : es la variable independiente que representa el punto en el que se evalúa la función resultante de la convolución.
- $x$ : es la matriz de entrada.
- $w$ : es el filtro o *kernel* aplicado a los datos de entrada, en las redes CNN es frecuentemente una matriz de pesos.
- $s(t)$ : es la salida obtenida, también llamada *feature map*.
- $x(a)$ : representa los datos de entrada  $x$  evaluados en  $a$ .
- $w(t - a)$ : es la función de entrada desplazada: representa cómo se desplaza y multiplica el *kernel*  $w$  a lo largo del eje  $t$  para realizar la convolución sobre los datos de entrada  $x$ .
- La integral representa el proceso de combinar las dos funciones  $x$  y  $w$  a lo largo de diferentes desplazamientos  $a$ , lo que resulta en la función convolucionada  $x * w$ .

Un ejemplo del funcionamiento de esta operación de convolución y el desplazamiento del *kernel* a través de los datos de entrada se muestra en la figura 4.7. El *kernel*, como ya se ha comentado, es un filtro o una matriz que se aplica sobre los datos de entrada para la extracción de características. El tamaño del *kernel* es variable, y se puede definir en función de las necesidades específicas del problema planteado, de forma que se puedan

capturar correctamente las características relevantes de los datos de entrada. Por ejemplo, en la figura 4.7, el tamaño del *kernel* es 3x3, y se aplica sobre una matriz de tamaño 5x5, obteniendo como resultado, una matriz de tamaño 3x3. El desplazamiento del *kernel* sobre la matriz de datos se realiza de izquierda a derecha y de arriba a abajo.

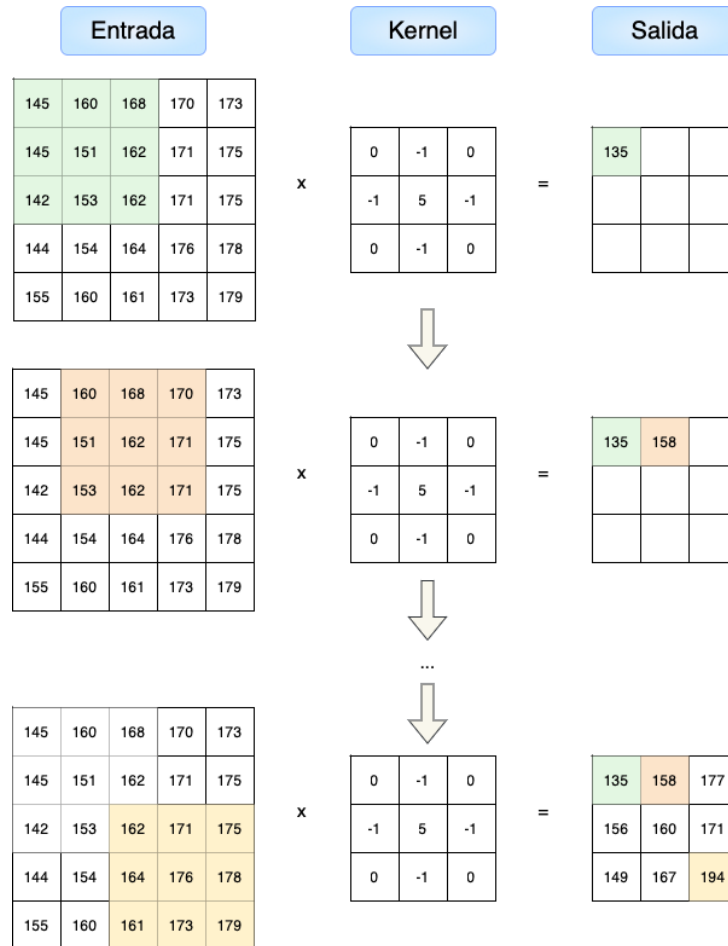
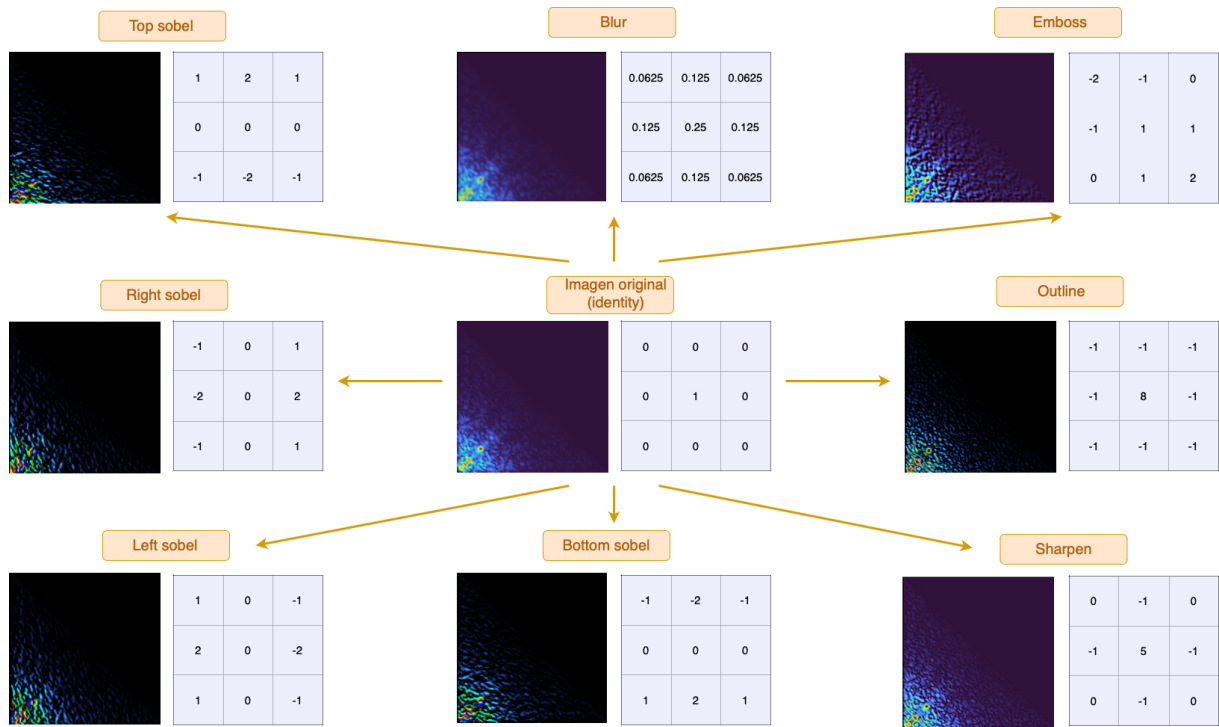


Figura 4.7: Operación de convolución

El *kernel* puede ser personalizado, según las necesidades específicas del problema, actuando como un filtro para resaltar diferentes aspectos de la imagen; o puede ser uno predefinido, como *blur*, que elimina el énfasis sobre las diferencias entre los valores de elementos adyacentes; *bottom sobel*, *left sobel*, *right sobel* o *top sobel*, que muestran únicamente las diferencias entre valores adyacentes en una dirección determinada; *emboss*, que aporta una sensación de profundidad enfatizando las diferencias de los valores en una dirección dada; *outline*, que resalta las diferencias grandes o *sharpen*, que enfatiza las diferencias entre píxeles adyacentes [65]. En la figura 4.8 se muestran estos ocho diferentes *kernels* de tamaño 3x3 aplicados sobre una imagen, en concreto, sobre un *bispectrum*.

Además del *kernel*, otro de los parámetros configurables que definen una operación de convolución es el *padding*. Este parámetro permite añadir bordes artificiales a la matriz de los datos de entrada, haciendo que el *kernel* cubra una mayor parte de la imagen y se



Figura 4.8: Aplicación de diferentes *kernels* predefinidos a una imagen

opere más veces sobre los píxeles de los bordes originales, preservando la información que éstos contienen. En la figura 4.9 se muestra un ejemplo de un *padding* de ceros alrededor de una matriz de tamaño 3x3, convirtiéndola en una matriz de tamaño 5x5. Este tipo de *padding* también permite controlar el tamaño de salida de la convolución de los datos.

El parámetro *stride* controla el movimiento del *kernel* sobre las matrices de datos de entrada. Cuando el *stride* vale 1 en una determinada dimensión, el desplazamiento será de una posición. Por ejemplo, en la figura 4.7, el *stride* es de [1, 1], haciendo que el *kernel* se desplace una posición a la derecha (de izquierda a derecha) y una posición hacia abajo (de arriba a abajo) en cada movimiento.

Dados los parámetros de tamaño de *kernel* ( $K$ ), *stride* ( $S$ ) y *padding* ( $P$ ), la dimensión de los datos de salida tras la operación de convolución ( $O$ ), conocida la dimensión de los datos de entrada ( $I$ ), se puede calcular a partir de la siguiente ecuación (ver [1]):

$$O = \frac{I - K + 2P}{S} + 1 \quad (4.2)$$

## Operación de *pooling*

La operación de *pooling*, también conocida como submuestreo, se suele aplicar después de una capa convolucional y una activación lineal en las CNN. Esta capa permite modificar la salida de una capa convolucional, disminuir su tamaño y resumir sus características más

|   |    |    |    |   |
|---|----|----|----|---|
| 0 | 0  | 0  | 0  | 0 |
| 0 | 12 | 35 | 32 | 0 |
| 0 | 67 | 78 | 98 | 0 |
| 0 | 25 | 94 | 73 | 0 |
| 0 | 0  | 0  | 0  | 0 |

Figura 4.9: Aplicación de un *padding* de ceros a una matriz de datos

importantes, contribuyendo a que las predicciones de este tipo de redes sean invariantes frente a pequeñas traslaciones en los datos de entrada [60].

El uso de *pooling* resulta fundamental cuando se trabaja con datos de entrada de dimensiones variables. Las redes neuronales necesitan recibir unos datos de entrada con dimensiones uniformes, por lo que, si estos tienen dimensiones diferentes, es preciso realizar un *pooling* para homogeneizarlas.

Además de reducir el tamaño de los datos, la operación de *pooling* disminuye los requisitos computacionales necesarios para el entrenamiento la red. También ayuda a reducir el sobreajuste en las CNN [60], ya que estas tienden a recordar la posición exacta de las características y pequeñas variaciones en esta posición pueden forzar a la red a efectuar cálculos adicionales.

Existen diferentes tipos de *pooling*, aunque los más comunes son: *max pooling*, que toma como salida el valor máximo de los datos de entrada; *average pooling*, el cual emplea como salida el valor medio de los datos de entrada; y *min pooling*, que toma el valor mínimo de los datos de entrada como salida [94]. En la figura 4.10 se puede ver un ejemplo de aplicación de cada uno de estos tres tipos de *pooling* 2x2 sobre una matriz de datos de tamaño 4x4, reduciendo su tamaño a una matriz 2x2.

Al igual que la operación de convolución, la operación de *pooling* se puede definir por los tres mismos parámetros: tamaño del *kernel*, que especifica las dimensiones de los datos de entrada sobre los que se aplica el *pooling*; *padding*, añade bordes a los datos de entrada; y *stride*, que indica el desplazamiento del *kernel* durante la operación de *pooling*. Las dimensiones de salida de los datos de una operación de *pooling* se pueden calcular de la misma forma que se mencionó previamente para la operación de convolución.

## Aplanado

Tras las capas convolucionales, se realiza un aplanado (también llamado *flattening*) de los datos. Este proceso consiste en transformar la salida multidimensional de las capas convolucionales en una matriz unidimensional. La razón principal es que las siguientes capas, las capas *fully-connected*, requieren una matriz unidimensional como entrada.

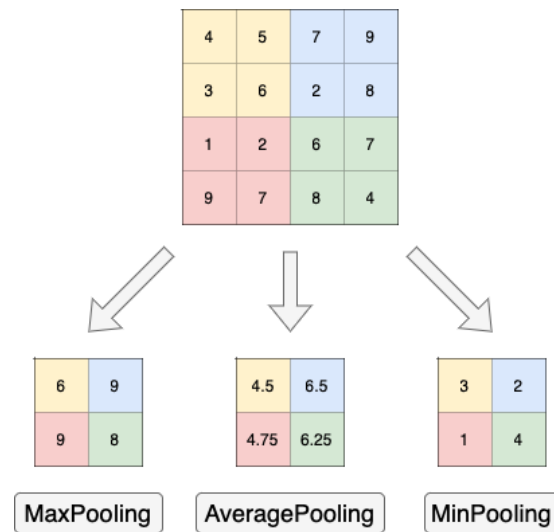


Figura 4.10: Aplicación de una operación de *pooling* sobre una matriz de datos

Además, el aplanado permite condensar los mapas de características extraídos en las capas convolucionales previas, facilitando su procesamiento posterior, de forma que el modelo sea capaz de aprender patrones complejos y sea capaz de interpretar de forma efectiva estos patrones y realizar predicciones precisas.

Cuando se aplica esta función sobre una matriz multidimensional, como se puede ver en la figura 4.11a, simplemente se comprimen todas las dimensiones en una única, dando lugar a una matriz unidimensional. Por ejemplo, en una matriz con dimensiones  $[10, 5, 4]$ , el vector tendrá una longitud igual al producto de cada una de las dimensiones de la matriz, es decir: 200.

## Capas *fully-connected*

El último bloque de las redes **CNN** es un conjunto de capas *fully-connected*. Estas capas se corresponden con un perceptrón multicapa clásico, donde cada una de sus neuronas está conectada con todas las neuronas de la siguiente capa, pero no están conectadas con neuronas de su misma capa ni con neuronas de capas anteriores [90].

La última capa *fully-connected* se corresponde con la salida de la **CNN** y genera las predicciones finales del modelo. En problemas de regresión, esta capa tiene una sola neurona. En cambio, en problemas de clasificación, el número de neuronas de esta capa depende del número de clases a predecir: para  $k$  clases serán necesarias  $k$  neuronas en la capa de salida. Por ejemplo, en un problema de clasificación binaria, como se muestra en la figura 4.11b, la capa de salida consta de dos neuronas.

Además de su función como generadoras de predicciones finales, las capas *fully-connected* también son responsables de la combinación lineal de características extraídas en las capas convolucionales previas. Esta combinación permite capturar relaciones complejas entre las características y llevar a cabo tareas de aprendizaje complejas.

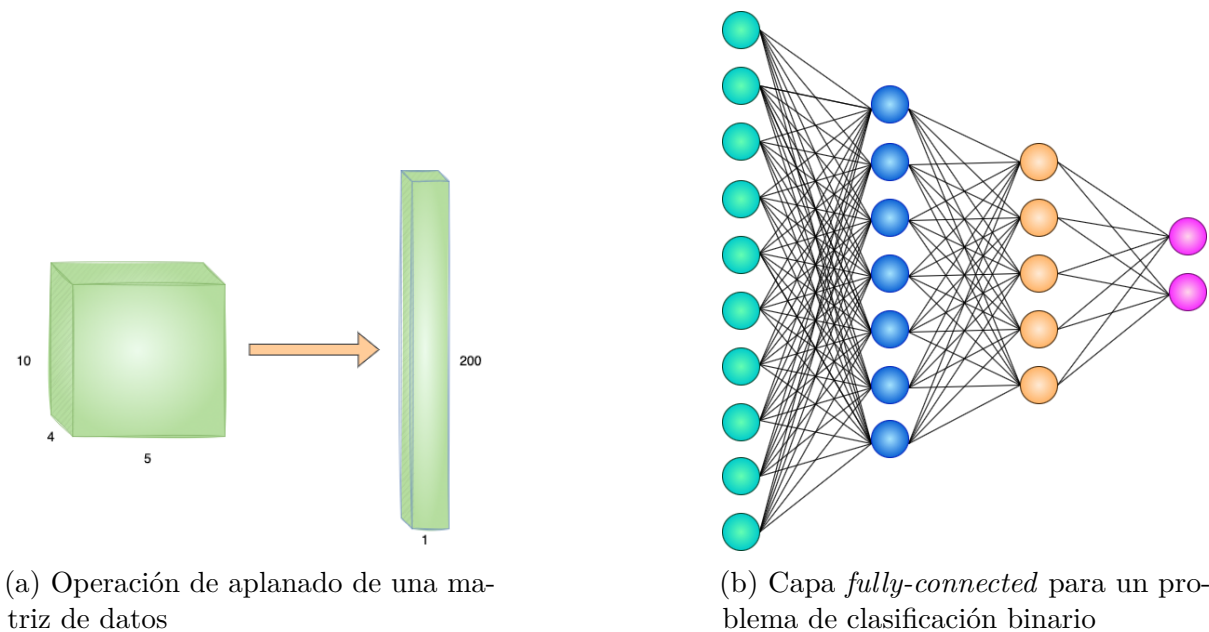


Figura 4.11: Aplanado y capa *fully-connected* en una CNN

## 4.2. Arquitecturas de redes neuronales convolucionales

En esta sección, se van a describir algunas de las arquitecturas más reconocidas y eficaces de redes neuronales convolucionales, y que han demostrado un alto rendimiento al ser aplicadas mediante *transfer learning* a diversos problemas. El *transfer learning* es una técnica que permite reutilizar un modelo previamente entrenado con una gran base de datos para abordar un nuevo problema con un conjunto de datos diferente [98]. Esta técnica no solo optimiza el tiempo y los recursos necesarios para el entrenamiento de los modelos, sino que también mejora la precisión y eficacia al aprovechar el conocimiento adquirido por el modelo original.

En concreto, en esta sección se van a analizar tres grupos de arquitecturas bien conocidos del estado del arte de las CNN: ResNet, EfficientNet y DenseNet. Estas arquitecturas se emplearán en etapas posteriores de este trabajo para la estimación de la severidad de la apnea del sueño en los pacientes que participaron en el estudio del SHHS.

### ResNet

Las redes residuales, o ResNet, son una arquitectura de redes convolucionales introducida en el artículo *Deep Residual Learning for Image Recognition* [27]. Esta arquitectura se desarrolló para abordar uno de los grandes problemas en el entrenamiento de redes neuronales profundas: el desvanecimiento o la explosión del gradiente. Para ello, ResNet presenta el concepto de “bloques residuales”. Estos bloques permiten la creación de unas conexiones rápidas que facilitan la propagación del gradiente y permiten entrenar modelos

mucho más profundos sin los problemas habituales de desvanecimiento del gradiente. La eficacia de ResNet fue demostrada tras ganar el *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) en 2015, con un modelo 8 veces mas profundo que las redes *Visual Geometry Group VGG* [77], y obteniendo un resultado con un error del 3,57% en el conjunto de test ImageNet. Actualmente, esta arquitectura es ampliamente utilizada para técnicas de *transfer learning* en múltiples aplicaciones.

Para explicar el funcionamiento de esta arquitectura y de los bloques residuales, supongamos que  $\mathcal{F}$  representa el conjunto de funciones que una arquitectura de red específica puede alcanzar. Es decir, para todo los  $f \in \mathcal{F}$ , existe algún conjunto de parámetros que se pueden obtener a través del entrenamiento en un conjunto de datos adecuado. Considerando  $f^*$  como la función que se pretende encontrar, si, por suerte, se encuentra en  $\mathcal{F}$ , entonces podremos identificarla directamente dentro de esta clase de funciones. En cambio, esto no suele ocurrir, más bien, se debe tratar de encontrar una función  $f_{\mathcal{F}}^*$ , que sea la mejor aproximación dentro de  $\mathcal{F}$ . El uso de regularización permite controlar la complejidad de  $\mathcal{F}$  y, generalmente, un mayor número de datos de entrenamiento y el uso de una arquitectura más potente  $\mathcal{F}'$  mejoran  $f_{\mathcal{F}}^*$  y proporcionan mejores resultados. En cambio, no se puede garantizar que  $\mathcal{F}$  esté contenido en  $\mathcal{F}'$ , por lo que no se puede asegurar que  $f_{\mathcal{F}'}^*$  sea mejor, incluso, puede ser peor [95].

De esta forma se puede ver que el uso de funciones no anidadas y el aumento de su complejidad no garantiza que se pueda reducir la distancia respecto a  $f^*$ , y puede que nunca se encuentre. En cambio, con las funciones anidadas  $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{F}_n$ , se evita este problema, puesto que las clases de las funciones más grandes siempre contienen a las más pequeñas y se garantiza que aumentarlas incrementa su capacidad para representar funciones más complejas. En la figura 4.12 se presenta un ejemplo de funciones anidadas (derecha) y no anidadas (izquierda), y en el que se muestra que las funciones no anidadas puede que nunca encuentre  $f^*$ , en cambio, en las funciones anidadas, cada función contiene a la anterior, por lo que a medida que se incrementa el número de funciones, se reduce la distancia a  $f^*$ .

En la figura 4.13 se muestra la comparación de un bloque de redes neuronales común (izquierda) y un bloque residual (derecha). Asumiendo que la entrada de estos bloques es  $x$ , y que  $f(x)$  es el mapeo que se pretende obtener mediante el aprendizaje del bloque, el bloque de la izquierda deberá aprender directamente del mapeo  $f(x)$ , mientras que el bloque residual aprende del mapeo residual  $g(x) = f(x) - x$ . Si el mapeo identidad  $f(x) = x$  es el mapeo deseado, entonces el mapeo residual equivale a  $g(x) = 0$ , por lo que el aprendizaje resulta más sencillo. En el bloque residual de la derecha se puede ver cómo hay una conexión, llamada conexión residual, que lleva  $x$  desde la entrada hasta un operador suma. De esta forma, con los bloques residuales, las entradas se propagan más rápidamente a través de las conexiones residuales entre capas.

Los modelos de redes residuales siguen un diseño similar a las redes VGG, con filtros de tamaño 3x3 en sus convoluciones. Cada bloque residual cuenta con dos capas convolucionales de tamaño 3x3, manteniendo el mismo número de canales de salida. Cada capa convolucional va seguida de una normalización del *batch* y una función de activación ReLU.

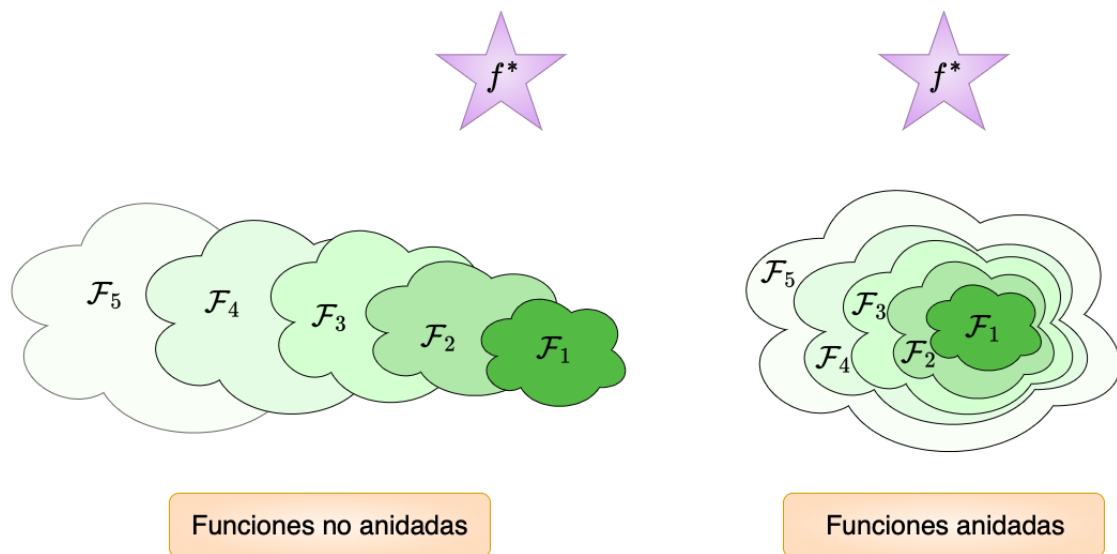


Figura 4.12: Comparación de funciones anidadas y no anidadas en busca de  $f^*$

Al final de cada bloque se añade la entrada original directamente, antes de la función de activación ReLU final, a través de una conexión residual. Este diseño requiere que la salida de las dos capas convolucionales tenga el mismo tamaño que la entrada, para que se puedan sumar correctamente.

Existen diferentes variantes de la arquitectura ResNet, como: ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152 y ResNet-200. Cada una de estas variantes recibe su nombre según el número total de capas que posee. La principal diferencia entre estos modelos es la profundidad de la red, lo que aumenta el número de bloques residuales y, por tanto, su capacidad para aprender características más complejas. Las arquitecturas ResNet-18 y ResNet-34 emplean únicamente bloques residuales simples con dos capas convolucionales de tamaño 3x3. Por otro lado, ResNet-50, ResNet-101, ResNet-152 y ResNet-200 emplean bloques denominados *bottleneck*, que combinan una capa convolucional de tamaño 1x1 para reducir la dimensionalidad de los datos, una capa de tamaño 3x3 para procesar los datos y otra capa de tamaño 1x1 para restaurar la dimensionalidad original de los datos, permitiendo aprender características más detalladas, aunque a costa de un mayor coste computacional [27].

## EfficientNet

Las redes EfficientNet fueron presentadas en 2019 en el artículo *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks* [80], con el objetivo de mejorar la precisión y la eficiencia de las redes neuronales convolucionales mediante un enfoque de escalado de modelos, de forma que, equilibrando la profundidad, el ancho y la resolución del modelo, se obtenga una mejora en el rendimiento. En las redes neuronales convolucionales es habitual escalar tan solo una de estas tres dimensiones, a pesar de que dos o las tres dimensiones

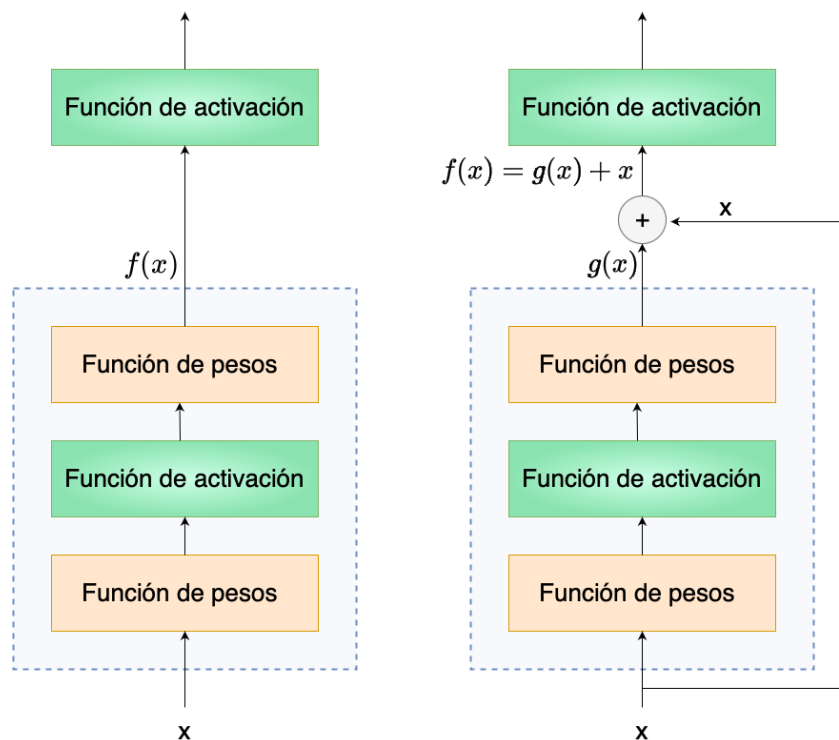


Figura 4.13: Bloque de redes neuronales común (izquierda) y bloque residual (derecha)

se pueden escalar; pero este ajuste se debe realizar de forma manual, por lo que resulta tedioso y frecuentemente produce una precisión y eficiencia subóptimas.

Mediante los modelos de redes EfficientNet es posible equilibrar las tres dimensiones de anchura, profundidad y resolución, escalando una de ellas con una proporción constante. El método empleado para ello consiste en escalar uniformemente estas tres dimensiones con un conjunto de coeficientes de escala fijos, que son determinados mediante un pequeño *grid search* en el modelo original pequeño. Por ejemplo, si se desean emplear  $2^N$  veces más recursos informáticos, se debe aumentar la profundidad de la red en  $\alpha^N$ , el ancho en  $\beta^N$  y la resolución en  $\gamma^N$ . En la figura 4.14 se puede ver una comparación de cómo, a partir de una determinada arquitectura convolucional, se puede realizar un escalado en anchura, profundidad y resolución (izquierda), y cómo los modelos EfficientNet realizan un escalado compuesto de estas tres dimensiones de forma simultánea (derecha).

Existen diferentes variantes de los modelos EfficientNet, entre los que se encuentran: EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, EfficientNet-B3, EfficientNet-B4, EfficientNet-B5, EfficientNet-B6 y EfficientNet-B7. La principal diferencia entre estos modelos radica en su tamaño y su capacidad de procesamiento, siendo B0 el modelo base y el más pequeño, y B7 el más grande y potente [80].

En la figura 4.15 se muestra una comparación del *accuracy* obtenido sobre el conjunto de datos de ImageNet de todas estas variantes de los modelos EfficientNet con otras arquitecturas reconocidas del estado del arte. Como se puede apreciar, las distintas



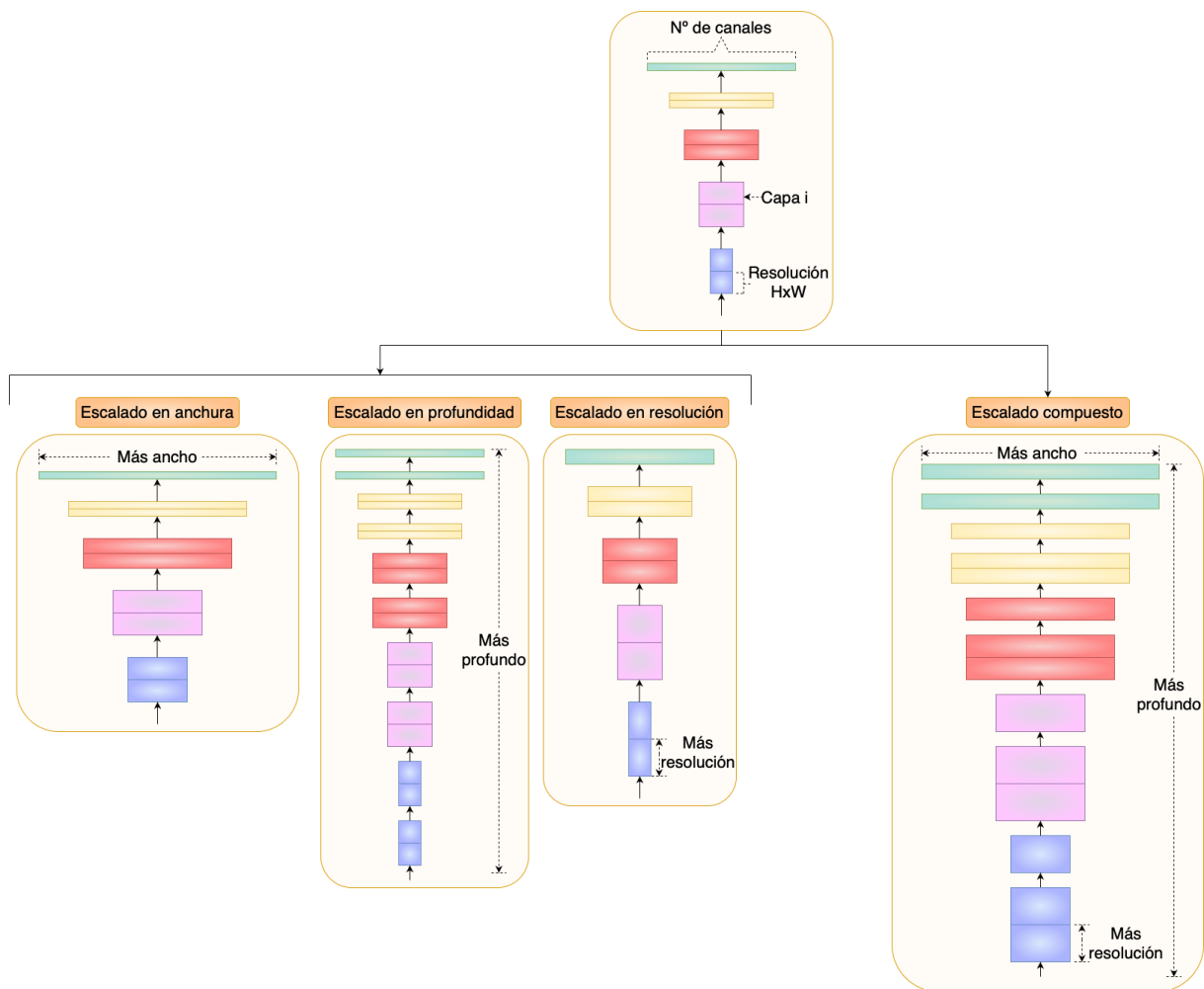


Figura 4.14: Comparación de escalado de dimensiones (ancho, profundidad y resolución) en CNN convencionales (izquierda), y escalado compuesto en arquitectura EfficientNet (derecha)

variantes de arquitectura EfficientNet logran obtener precisiones altas y superiores a la mayoría de las otras arquitecturas, al mismo tiempo que requieren de un número de parámetros muy inferior.

## DenseNet

Las redes convolucionales densamente conectadas, también conocidas como DenseNet, fueron introducidas en 2017 en el artículo *Densely Connected Convolutional Networks* [29]. Esta arquitectura se deriva de la familia de redes ResNet, abordando un problema identificado en estas últimas. Mientras que en ResNet cada capa está conectada solo con la siguiente capa, en DenseNet cada capa está conectada con todas las capas siguientes en una estructura densamente conectada, como se puede ver en la figura 4.16.



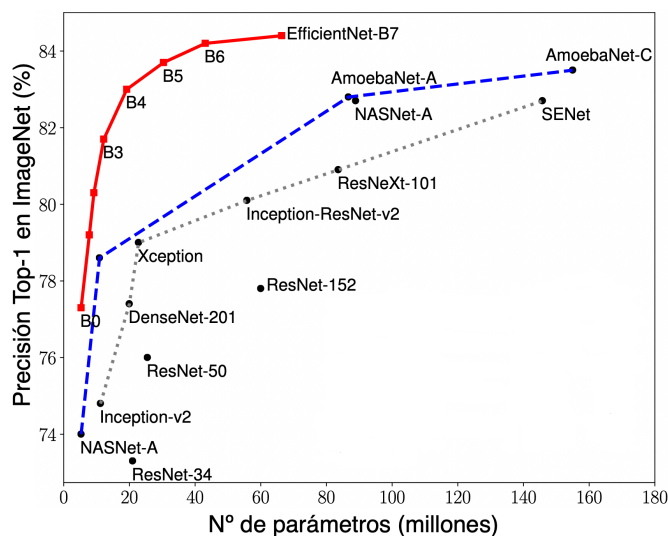


Figura 4.15: Comparación del rendimiento de las distintas variantes de modelo EfficientNet con otras arquitecturas reconocidas del estado del arte (extraído de [80])

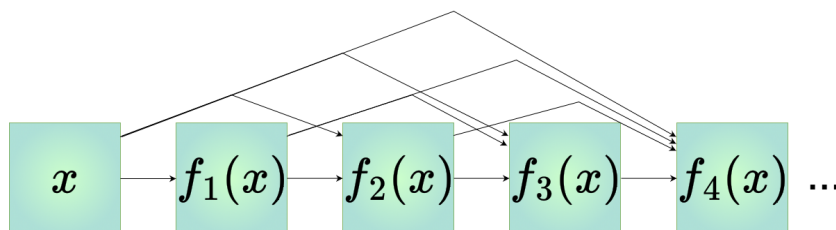


Figura 4.16: Conexiones en modelo DenseNet

La idea principal detrás de DenseNet es similar a la descomposición de Taylor para funciones, pero aplicado a redes neuronales convolucionales. La descomposición de Taylor descompone una función en términos de orden creciente. En el punto  $x = 0$ , esta descomposición se puede escribir de la siguiente forma [95]:

$$f(x) = f(0) + x \cdot \left( f'(0) + x \cdot \left( \frac{f''(0)}{2!} + x \cdot \left( \frac{f'''(0)}{3!} + \dots \right) \right) \right) \quad (4.3)$$

Por el contrario, ResNet descompone las funciones de la forma:  $f(x) = x + g(x)$ , es decir, descompone las funciones en un término lineal sencillo y en otro no lineal más complejo. DenseNet surge por la necesidad de capturar información más allá de dos términos. En la figura 4.17 se puede ver una comparación entre un bloque ResNet (izquierda) y un bloque Densenet (derecha). En un bloque ResNet la información de ambos términos se suma, en cambio, en un bloque DenseNet la información se concatena, de forma que obtiene un mapeo de sus datos de entrada ( $x$ ) a los valores obtenidos después de aplicar una secuencia cada vez más compleja de funciones [95]:

$$x \rightarrow [x, f_1(x), f_2([x, f_1(x)]), f_3([x, f_1(x), f_2([x, f_1(x)])]), \dots] \quad (4.4)$$

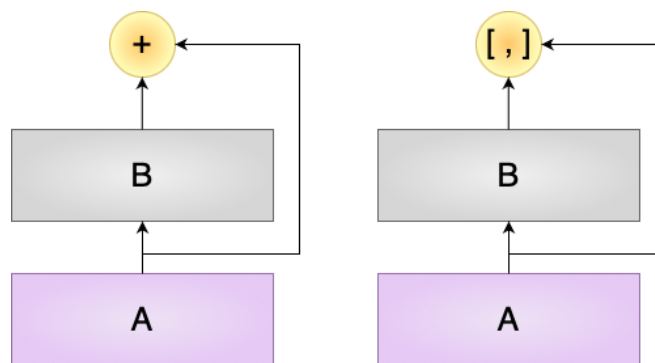


Figura 4.17: Comparación ResNet (izquierda) y DenseNet (derecha)

Cada bloque en una arquitectura DenseNet toma la salida de todas las capas anteriores como entrada y las concatena junto con la salida de la operación convolucional que se realiza en ese bloque. Esta concatenación crea una conexión densa entre las diferentes capas de la red, permitiendo que la información fluya de forma más directa y completa a lo largo de la red, y que cada capa tenga acceso directo a la información de todas las capas anteriores, facilitando el aprendizaje de representaciones más complejas.

En una arquitectura DenseNet, los bloques densos se intercalan con capas de transición que reducen la dimensionalidad de la salida. Estas capas combinan operaciones de *pooling* con *stride* 2 y convoluciones de dimensión 1x1 para controlar el número total de canales y evitar que la expansión de la red sea demasiado grande y tenga un coste computacional excesivo.

Existen diferentes variantes de los modelos DenseNet, aunque los más conocidos son: DenseNet-121, DenseNet-161, DenseNet-169 y DenseNet-201. Estas variantes se nombran de acuerdo al número total de capas que contienen, al igual que ocurría con las distintas variantes del modelo ResNet.

### 4.3. Inteligencia Artificial Explicable

A pesar de su alto rendimiento en una amplia variedad de tareas, los modelos de redes neuronales, especialmente los de *deep learning*, son a menudo considerados “cajas negras”. Esto se debe a su gran profundidad y complejidad, lo que dificulta enormemente la comprensión de las razones detrás de sus decisiones. Debido a esta opacidad, surge el campo de la Inteligencia Artificial Explicable (XAI, por sus siglas en inglés). XAI se centra en el desarrollo de métodos y técnicas que permitan explicar el funcionamiento interno de estos modelos, proporcionando interpretaciones comprensibles de las decisiones tomadas por los algoritmos de *deep learning* [72].

Esta falta de transparencia plantea grandes desafíos, especialmente en aplicaciones críticas donde se requiere confianza y explicabilidad, como en el campo de la medicina. En este contexto, es crucial que los médicos y otros profesionales de la salud comprendan las razones detrás de las recomendaciones y diagnósticos generados por los modelos de *deep learning*, para poder tomar decisiones informadas y seguras en el tratamiento de los pacientes.

En el contexto de la estimación de la severidad de la apnea del sueño de un paciente mediante CNN, también resulta necesario justificar las decisiones tomadas por los modelos. Es por esto que en este trabajo se han estudiado dos métodos de explicabilidad para tratar de comprender las predicciones proporcionadas por los modelos. En primer lugar, una técnica extendida en el estado del arte: Grad-Cam; y en segundo lugar, una técnica basada en ablaciones para validar la técnica de Grad-Cam empleada previamente.

## Grad-Cam

La técnica de *Gradient-weighted Class Activation Mapping* (Grad-Cam), propuesta en 2019 en [73], tiene como objetivo la generación de una representación visual que justifique las decisiones tomadas por los modelos de *deep learning*. Para ello, identifica las regiones de una imagen que son importantes para una predicción específica de la red a través de la generación de un mapa de activación ponderado mediante el cálculo de gradientes. Su principal utilidad radica en su gran capacidad para visualizar y comprender las características que más influyen en las decisiones tomadas por los modelos, facilitando así la interpretabilidad de los modelos de redes neuronales más complejos.

Grad-Cam es una generalización del algoritmo *Class Activation Mapping* (CAM) que emplea los gradientes de las capas convolucionales para identificar las regiones más sensibles de la entrada que influyen en la predicción final [73]. Estos mapas basados en gradientes son mapas de calor que se obtienen de cada capa individual del modelo.

Para obtener el mapa de calor de Grad-Cam por clase:  $L_{Grad-CAM}^C \in \mathbb{R}^{u \times v}$ , con un ancho de  $u$  y un alto de  $v$  para una clase  $c$ , en primer lugar se calcula el gradiente de la puntuación para la clase  $y^c$ , con respecto al mapa de activación ponderado  $A^k$  de una capa convolucional. Siendo  $Z$  el número de mapas de características (filtros), estos gradientes se retropropagan sobre las dimensiones de altura y anchura para obtener la importancia de los pesos  $\alpha_k^c$  [73]:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (4.5)$$

Durante el cómputo de  $\alpha_k^c$ , mientras se propagan hacia atrás los gradientes respecto a las activaciones, el cálculo exacto equivale a productos matriciales sucesivos de las matrices de pesos y el gradiente con respecto a las funciones de activación hasta la capa de convolución final a la que se propagan los gradientes. Por lo tanto, este peso  $\alpha_k^c$  captura la importancia del mapa de características  $k$  para una clase objetivo  $c$ . Al realizar una

combinación ponderada de los mapas de activación hacia delante, seguido de una ReLU, se obtiene [73]:

$$L_{Grad-CAM}^C = ReLU\left(\sum_k \alpha_k^c A^k\right) \quad (4.6)$$

Los mapas de calor generados por Grad-Cam tienen el mismo tamaño que los mapas de características de las capas convolucionales. Posteriormente, se aplica una ReLU a la combinación lineal de los mapas, puesto que solamente interesan las características que tienen una influencia positiva en la clase de interés. Esto significa que se destacan los píxeles cuyo incremento de intensidad también aumenta  $y^c$ , mientras que los píxeles con influencia negativa se suprimen, ya que probablemente pertenecen a otras categorías de la imagen [73]. En la figura 4.18 se muestra un ejemplo de aplicación de esta técnica de Grad-Cam sobre un bispectrum.

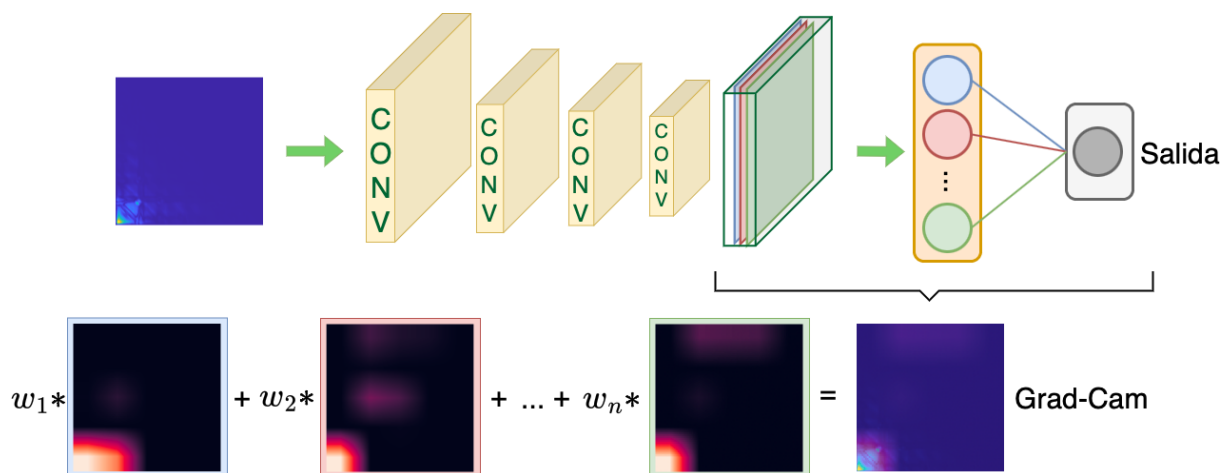


Figura 4.18: Aplicación de la técnica de Grad-Cam a un bispectrum

## Ablaciones sobre los datos de entrada

Las ablaciones, en el contexto de la inteligencia artificial, se refieren a la eliminación o modificación de un componente de un sistema para investigar cómo se comporta sin dicho componente [53] [55]. La hipótesis principal es que, al alterar las características más importantes, se debería observar un deterioro de las capacidades predictivas del modelo, en comparación con la alteración de características menos significativas. Esta metodología permite evaluar la relevancia y el impacto de cada característica en el rendimiento del modelo, facilitando su optimización y una mejor comprensión de su funcionamiento [26]. Además, como se presenta en [78], las redes neuronales profundas pueden ser muy sensibles a pequeñas alteraciones en las entradas, lo que significa que incluso un cambio mínimo y casi imperceptible en una imagen puede hacer que la red la clasifique incorrectamente. El uso de ablaciones sobre los datos de entrada permite ver si este problema se agrava o se

mitiga al eliminar o modificar componentes específicos de los datos, proporcionando una visión más clara sobre la robustez del modelo frente a perturbaciones.

El uso de ablaciones en el campo del *machine learning* y *deep learning* sigue siendo relativamente limitado. La mayoría de los estudios se enfocan en añadir o eliminar componentes del modelo [53] o en manipular características derivadas de los datos [52]. Investigaciones más avanzadas aplican ablaciones directamente en los datos para problemas de visión artificial, como en [55], donde se analiza el impacto de eliminar patrones repetidos en imágenes y se evalúa si las redes neuronales pueden mantener un alto nivel de precisión en sus predicciones [59].

Para el problema de estimación de la severidad de la apnea del sueño a partir de los bispectrums de las señales respiratorias de los pacientes, se han aplicado las ablaciones sobre los datos de entrada. Como se muestra en la figura 4.19, para analizar la capacidad predictiva del modelo y los aspectos en los que se fija para realizar sus predicciones, se toman dos imágenes: una representando un bispectrum de una señal respiratoria normal, sin eventos apneicos, y otra de una señal con presencia de eventos apneicos.

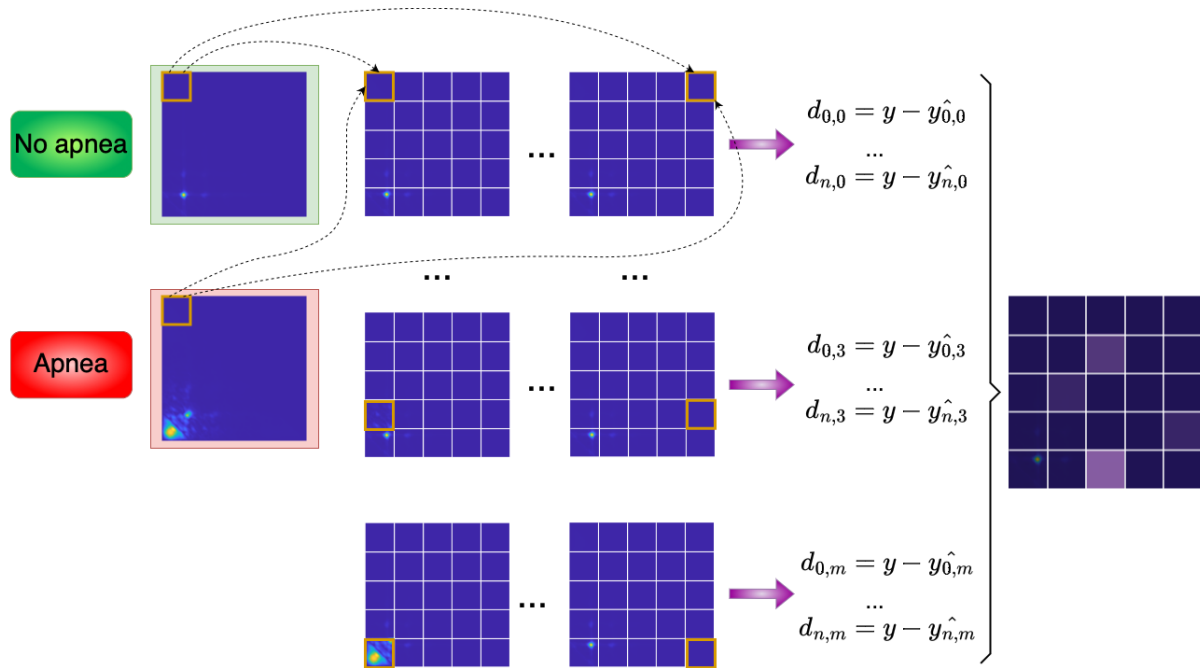


Figura 4.19: Aplicación de ablaciones sobre un bispectrum sin eventos apneicos

Para el proceso de aplicación de las ablaciones sobre los datos de entrada, se selecciona una de las imágenes, por ejemplo, la señal sin eventos apneicos, y se divide en una cuadrícula. A continuación, se itera sobre cada celda de esta cuadrícula, sustituyendo en cada paso una celda por su equivalente en la otra imagen de bispectrum con eventos apneicos. En cada iteración, se pasa la imagen modificada por la red neuronal entrenada para obtener una predicción del número de eventos apneicos presentes en ese fragmento de señal. A partir de esta estimación, se calcula la diferencia entre el valor predicho

del bispectrum modificado ( $\hat{y}$ ) y el valor real ( $y$ ) del número de eventos presentes en el fragmento de señal representado en el bispectrum:

$$d = y - \hat{y} \quad (4.7)$$

A partir de todas las diferencias calculadas en cada iteración, se genera un mapa de calor que muestra, con tonalidades más intensas, las celdas donde la diferencia es mayor. Una diferencia significativa entre la estimación del bispectrum con una ablación y el bispectrum sin modificar indica que la red neuronal está utilizando esa región específica de la ablación para extraer características relevantes. Por el contrario, una diferencia mínima sugiere que la ablación en esa porción del bispectrum no afecta significativamente a la predicción, lo que implica que dicha región contiene pocas características importantes para determinar el número de eventos apneicos en la señal.

---

## 5: Sujetos y señales

---

Para el desarrollo de este proyecto, se han analizado las señales obtenidas a través de polisomnografías nocturnas del Sleep Heart Health Study. En particular, se utilizarán las señales de esfuerzo respiratorio torácico y abdominal registradas en este estudio para abordar el problema de estimación de la severidad de la apnea del sueño mediante técnicas de *deep learning*. En este capítulo detalla el estudio del Sleep Heart Health Study, se explica la motivación detrás de la selección de las señales de esfuerzo respiratorio y se describe el proceso mediante el cual estas señales han sido procesadas y transformadas en bispectrums, adaptándolas para su aplicación en el problema planteado en este proyecto.

### 5.1. Sleep Heart Health Study

El Sleep Heart Health Study ([SHHS](#)) es un estudio diseñado para la investigación de diferentes trastornos respiratorios del sueño, entre los que se encuentra la [AOS](#), que se han identificado como factores de riesgo para el desarrollo de enfermedades cardiovasculares [67] [96]. La base de datos de este estudio ([SHHS](#), ClinicalTrials.gov Identifier: NCT00005275) es accesible bajo petición a través del National Sleep Research Resource ([NSRR](#)) (<https://sleepdata.org/datasets/shhs>).

El [SHHS](#) es un estudio multicentro desarrollado por el National Heart Lung y por el Blood Institute. Este estudio se compone de dos etapas en las que participantes adultos mayores de 40 años se someten a una [PSG](#) en su domicilio para evaluar la presencia de [AOS](#) u otros trastornos respiratorios del sueño. El principal objetivo de este estudio es probar la relación de determinados trastornos del sueño y un incremento en el riesgo de padecer enfermedades coronarias, ictus, accidentes cerebrovasculares, aumento longitudinal de la presión arterial, así como la mortalidad por todas las causas anteriores [67].

Este estudio se llevó a cabo en dos etapas: SHHS1 y SHHS2. La primera etapa (SHHS1) se desarrolló entre los años 1995 y 1998, y participaron un total de 6441 individuos. La segunda etapa (SHHS2) se llevó a cabo entre los años 2001 y 2003, y participaron un total de 2651 individuos. Todos estos individuos ya habían participado en la primera etapa del estudio [67].

Todos los participantes en SHHS cumplieron una serie de criterios. En primer lugar, todos debían tener al menos 40 años, puesto que los individuos jóvenes tienen un menor riesgo de padecer enfermedades cardiovasculares. En segundo lugar, se excluyeron a las personas con apnea del sueño siguiendo un tratamiento de presión positiva continua en las vías respiratorias o con dispositivos orales, las personas que empleaban oxigenoterapia domiciliaria y a aquellos a los que se les hubiera practicado una traqueotomía. El motivo es que la realización de una polisomnografía en estos individuos resulta más compleja, además de que el tratamiento que siguen puede alterar el riesgo de desarrollar enfermedades cardiovasculares. Además, se reclutó a un número de hombres y mujeres aproximadamente igual, de forma que se pudiera aplicar el estudio a ambos sexos. También se dio prioridad a las minorías étnicas, para garantizar que la muestra sea representativa de la diversidad étnica de la población general de los Estados Unidos. En este estudio también se llevó a cabo un sobremuestreo de sujetos menores de 65 años y que roncan habitualmente, para enriquecer la muestra con sujetos más jóvenes, en los que roncar multiplica entre tres y diez veces el riesgo de padecer AOS. Por último, no se excluyeron del estudio sujetos con enfermedades cardiovasculares ni hipertensión para poder analizar si las personas con estas enfermedades tienen un riesgo diferente de desarrollar enfermedades cardiovasculares en comparación con las personas que no las tienen al inicio del estudio [67].

Todos los pacientes que participaron en el estudio fueron reclutados a través de correo, llamadas telefónicas y durante las visitas médicas programadas para revisiones de otros estudios en los que estaban participando. Entre estos estudios se incluye el Arteriosclerosis Risk in Communities Study (ARIC), cuyo objetivo principal es la investigación sobre el origen y naturaleza de la arteroesclerosis y la arteroesclerosis clínica y subclínica en cuatro comunidades americanas; el Cardiovascular Health Study (CHS), cuyo objetivo es la identificación de los factores de riesgo de enfermedades coronarias e ictus en adultos; el Framingham Heart Study (FHS), en el que se estudió tanto a hombres como a mujeres con problemas cardíacos y a sus parejas e hijos; el Strong Heart Study, que se centra en estimar la mortalidad y morbilidad causadas por enfermedades cardiovasculares, además de comparar los factores de riesgo entre nativos americanos; el New York Hypertension Cohorts, en el que se estudió la hipertensión, la relación entre el estrés laboral y la presión sanguínea, así como una evaluación cardiovascular completa en individuos de tres poblaciones diferentes a un proyecto denominado Psychosocial Factors and Cardiovascular Disease. Por último, entre estos estudios también se incluyen el Tucson Epidemiologic Study of Airways Obstructive Diseases y el Health and Environment Study, en los que se estudió la etiología y la historia de las enfermedades obstructivas de las vías respiratorias, respectivamente [67].

Para la creación de la base de datos de SHHS, en primer lugar, se llevó a cabo un análisis comparativo de los métodos de recolección de datos de los distintos estudios precedentes que previamente se han mencionado. Para ello, se han recogido una serie de variables que pueden agruparse en: críticas, de subconjuntos específicos y específicas de cada base de datos. En concreto, para el primer grupo se analizan variables como raza, sexo, edad, si el paciente padece alguna enfermedad cardiovascular, hipertensión, diabetes, consumo de tabaco, alcohol o cafeína, altura, peso, etc. [67].



Posteriormente, los pacientes rellenaron a un cuestionario sobre sus hábitos de sueño, con el objetivo de, en primer lugar, determinar el nivel de ronquidos de los participantes para sobremuestrear los individuos que roncan habitualmente, identificar los potenciales participantes que deben ser excluidos del estudio, obtener información sobre los datos de sueño y síntomas de los participantes que no están disponibles en otros estudios, y recopilar información de los hábitos de sueño de una muestra muy grande. Este cuestionario fue respondido no solo por los participantes en el [SHHS](#), sino en todos los estudios originales que se han mencionado previamente [67]. Estos cuestionarios que respondieron los participantes contenían preguntas sobre sus hábitos de sueño, su hábito de fumar, así como sobre enfermedades cardiovasculares, problemas respiratorios, altura, peso, sexo, etnia, educación, estado civil o consumo de medicamentos [76].

Por último, los pacientes se sometieron a una polisomnografía nocturna en su domicilio. Inicialmente se realizó un examen domiciliario a cada uno de los pacientes en el que se recogió su presión arterial en reposo, peso, circunferencia del cuello, preguntas relativas a enfermedades cardiovasculares, respiratorias y tabaquismo, medicación actual, calidad de vida, somnolencia de Epworth modificada y calidad de sueño matutino. La [PSG](#) se llevó a cabo a través de un polisomnógrafo de Compumedics P-Series Sleep Monitoring System. Este sistema consta de una Patient Interface Box (PIB “headbox”) que contiene amplificadores y filtros a los que se conectan electrodos y sensores. Este PIB se conecta a través de un cable a un dispositivo de almacenamiento, que incluye un ordenador (PCMCIA card), una batería recargable de 15 horas y un oxímetro. El PIB, los cables de los electrodos y los sensores se fijan al cuerpo del paciente mediante una malla (“bib”). Durante la visita del médico en la casa del paciente se colocan los sensores y se calibran los equipos. Los datos se recogen en tiempo real en la memoria PCMCIA y se envían a un centro de lectura de polisomnografía, donde se revisan todos los datos recogidos para comprobar su calidad técnica [67].

Durante una polisomnografía, se recogen múltiples mediciones para su posterior análisis para el diagnóstico de [AOS](#). Estas mediciones incluyen registros de [EEG](#) en C3/A2 y C4/A1, adquiridos a una frecuencia de muestreo de 125 Hz. Además, se registra la actividad ocular mediante un [EOG](#) en ambos ojos, muestreado a 50 Hz para detectar movimientos oculares rápidos durante el sueño REM. La actividad muscular se registra mediante un [EMG](#) muestreado a una frecuencia de 125 Hz. Se monitorizan también los movimientos torácicos ([THOR](#)) y abdominales ([ABDO](#)) a través de bandas de pletismografía inductiva, muestreados a 10 Hz para evaluar la respiración durante el sueño. El flujo de aire nasal y oral se registra mediante un termopar nasal-oral a 10 Hz para evaluar la respiración. La saturación de oxígeno en la sangre se mide mediante una oximetría de pulso en la yema del dedo, muestreada a 1 Hz. También se registra el ritmo cardíaco mediante un [ECG](#) con una frecuencia de muestreo de 125 Hz o 250 Hz, según el protocolo específico del estudio. La frecuencia cardíaca derivada del [ECG](#) se muestrea a una frecuencia de 1 Hz. Por último, se registra la posición del cuerpo utilizando un sensor de calibre de mercurio y la luz ambiental mediante un sensor de luz para contextualizar las condiciones del entorno durante el estudio. Es importante destacar que el ronquido no se registró

debido a la dificultad para definirlo objetivamente y medirlo con precisión en el entorno de la polisomnografía.

## 5.2. Sujetos y señales a estudiar

En este trabajo se ha analizado la base de datos [SHHS](#) que se acaba de describir. En concreto, de todas las señales recogidas a través de [PSG](#) en este estudio, se ha decidido analizar la información contenida únicamente en dos señales relacionadas con la respiración del sujeto: [THOR](#) y [ABDO](#), recogidas mediante bandas de pletismografía inductiva durante una noche de sueño en los domicilios de los pacientes.

Estas dos señales no se han analizado de forma exhaustiva en otros estudios del estado del arte, debido, principalmente, a la inestabilidad de los sensores empleados para la adquisición de estas señales [40]. En cambio, la [AASM](#) reconoce la importancia de estas dos señales para diferenciar entre eventos apnéicos centrales y obstructivos [9]. Cuando la respiración de un paciente es normal, se ha observado que los movimientos torácicos y abdominales del paciente se producen de forma simultánea. En cambio, cuando ocurre un evento obstructivo, se produce un fenómeno denominado “respiración paradójica”, que surge por un intento del cuerpo del paciente por superar la obstrucción de las vías respiratorias. En este tipo de respiración, los movimientos torácicos y los abdominales se producen en direcciones opuestas [40] [86]. Por otro lado, cuando se produce un evento central, el movimiento respiratorio, tanto torácico como abdominal, se reduce o incluso se anula completamente [40] [86]. En la figura 5.20 se puede ver a izquierda (figura 5.20a) un evento de tipo central, a la derecha (figura 5.20b) un evento de tipo obstructivo y en la parte inferior (figura 5.20c), un fragmento de señal sin ningún tipo de evento. En estas tres imágenes se están representando un fragmento de un minuto de una señal torácica (naranja) y una obstructiva (azul) de un paciente del conjunto de datos [SHHS](#) empleado en este trabajo.

Como ya se ha mencionado previamente, el número total de [PSG](#) realizadas en las dos etapas del estudio fue de 9092. En concreto, 6441 individuos participaron en la primera etapa de SHHS1, y 2651 individuos repitieron la polisomnografía a la que ya se habían sometido, 5 años después, en la segunda etapa de SHHS2. La frecuencia de muestreo de la mayoría de las señales recogidas es de 10 Hz; en cambio, un pequeño grupo de señales, concretamente 35, fueron recogidas con una frecuencia de muestreo inferior, de 8 Hz. La decisión tomada ha sido descartar estas señales debido a que representan un número casi insignificante frente al número total de señales disponibles en el estudio del [SHHS](#). Además, se ha descartado el remuestreo de las señales de 8 Hz a 10 Hz puesto que se podrían introducir artefactos y distorsiones que podrían comprometer la calidad de las señales. Por este motivo, el número total de señales empleadas en este trabajo es de 8257, en concreto, 5722 del SHHS1 y 2535 del SHHS2.

Todos los sujetos participantes en este estudio han sido diagnosticados, y sus señales se han etiquetado siguiendo las reglas dictadas por la [AASM](#) para puntuar los eventos de apnea e hipopnea. Para la evaluación de la severidad de la enfermedad de la apnea en los

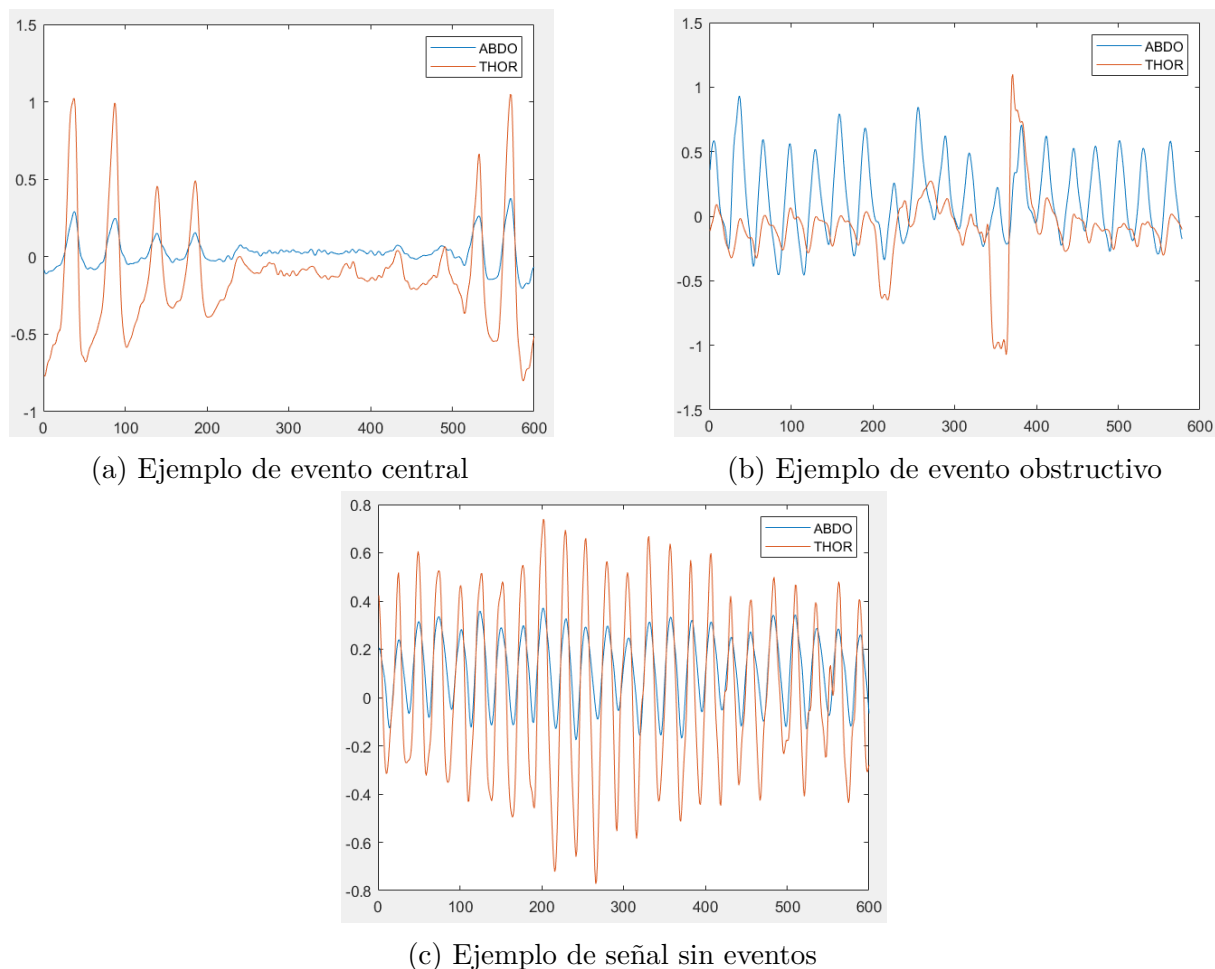


Figura 5.20: Fragmentos de señal con un evento central, un evento obstructivo y sin eventos

sujetos se ha empleado el **IAH** y el **IAC**, que miden el número de eventos apneicos totales y centrales, respectivamente, en una hora de sueño de un paciente. Se han empleado los umbrales habituales en el estado del arte para determinar la severidad de la apnea de los pacientes: no apnea:  $\text{IAH} < 5$  e/h, leve  $5 \text{ e/h} \leq \text{IAH} < 15$  e/h, moderada  $15 \text{ e/h} \leq \text{IAH} < 30$  e/h y severa  $\text{IAH} \geq 30$  e/h.

El total de las 8257 señales válidas empleadas en este estudio se han dividido en tres conjuntos de datos para su uso posterior en la experimentación con modelos de *deep learning*: entrenamiento, validación y test. Como ya se ha comentado previamente, los individuos que participaron en la segunda etapa del estudio habían participado previamente en la primera, por lo que la distribución de las señales entre estos tres conjuntos se debe realizar con cautela, para evitar que se produzcan sesgos en resultados posteriores. Es por esto que diferentes señales de un mismo individuo no pueden distribuirse en diferentes grupos, por lo que todas las señales de los individuos que participaron en ambas etapas del estudio se emplearon en el conjunto de datos de test. El resto de señales, tanto las de los individuos participantes en SHHS1, que no participaron en SHHS2, como los

individuos participantes en SHHS2, cuyas señales en SHHS1 se eliminaron previamente por ser recogidas con una frecuencia de muestreo inferior a 10 Hz, se emplearon tanto para el conjunto de datos de entrenamiento como para el conjunto de datos de validación. Todos estos individuos restantes se han repartido de forma aleatoria entre los conjuntos de entrenamiento (80 %) y el conjunto de validación (20 %). En la figura 5.21 se muestra un esquema de cómo se han distribuido estos individuos en los tres conjuntos de entrenamiento, validación y test, y se indica la cifra exacta de individuos distribuidos a cada conjunto de datos.

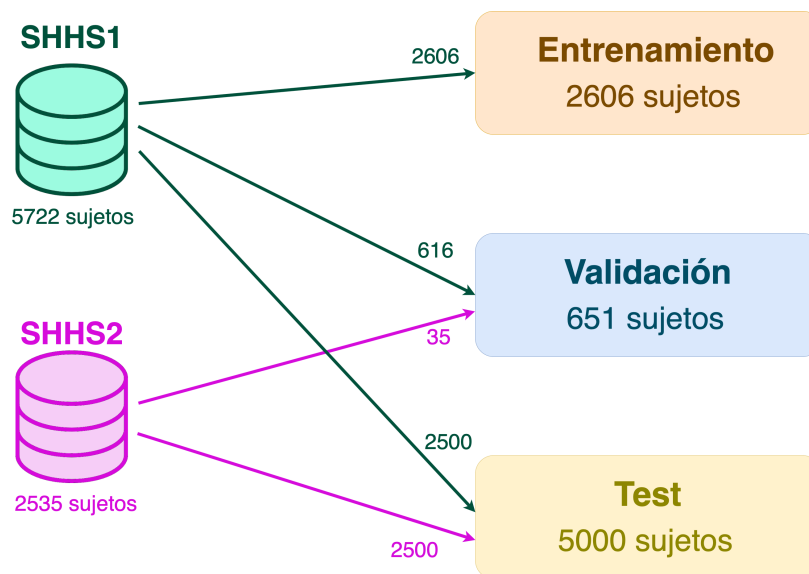


Figura 5.21: Distribución de los datos de entrenamiento en los conjuntos de entrenamiento, validación y test

En la tabla 5.11 se muestra una serie de datos sociodemográficos que caracterizan cada uno de los conjuntos de datos en los que se dividen las señales del SHHS y del total de las señales. En concreto, para cada conjunto de datos se muestra el número de sujetos de cada grupo, su edad (media y desviación típica), el número y la proporción de hombres en cada grupo, su IMC (media y desviación típica), el IAC e IAH (media y desviación típica) y el número de individuos diagnosticado en cada uno de los diferentes grupos de severidad mencionados previamente, así como la proporción que representa esta cifra respecto al total de individuos de cada grupo. Como se puede apreciar, el número de individuos con un menor número de eventos apnéicos es minoritario para los tres conjuntos de datos en los que se divide las señales recogidas en el SHHS.

### 5.3. Procesamiento de señales

Todas las señales empleadas en este trabajo han sido procesadas antes de ser utilizadas para el entrenamiento de modelos de *deep learning*. En primer lugar, el formato en el que se registran los datos en el estudio SHHS es .edf (European Data Format). Este formato es

| Datos                    | Total          | Entrenamiento  | Validación    | Test           |
|--------------------------|----------------|----------------|---------------|----------------|
| Nº de sujetos            | 8257           | 2606 (31.56 %) | 651 (7.88 %)  | 5000 (60.56 %) |
| Edad (años)              | 65 [16]        | 64 [21]        | 62 [12]       | 65 [16]        |
| Nº de hombres            | 3887 (47.08 %) | 1275 (48.93 %) | 322 (49.46 %) | 2290 (45.80 %) |
| IMC (kg/m <sup>2</sup> ) | 27.54 [6.10]   | 27.15 [6.18]   | 28.02 [6.32]  | 27.68 [6.02]   |
| IAC (e/h)                | 0 [0.26]       | 0 [0.33]       | 0.12 [0.26]   | 0 [0.26]       |
| IAH (e/h)                | 21.68 [21.52]  | 25.62 [23.08]  | 26.08 [26.48] | 19.38 [19.51]  |
| #No apnea                | 338 (4.09 %)   | 77 (2.95 %)    | 11 (1.69 %)   | 250 (5.00 %)   |
| #Leve                    | 2244 (27.18 %) | 529 (20.30 %)  | 143 (21.97 %) | 1572 (31.44 %) |
| #Moderado                | 3003 (36.37 %) | 958 (36.76 %)  | 214 (32.87 %) | 1831 (36.62 %) |
| #Severo                  | 2672 (32.36 %) | 1042 (39.99 %) | 283 (43.47 %) | 1347 (26.94 %) |

Tabla 5.11: Tabla resumen de los datos sociodemográficos de los participantes en el estudio

el habitual para la recogida de datos en polisomnografías y encefalogramas, puesto que es sencillo y flexible para el almacenamiento de señales biológicas. De estos archivos se extrae la información que se va a emplear en este trabajo: las señales en crudo de respiración torácica y abdominal, y el etiquetado de los eventos.

Las señales, al haber sido recogidas en un entorno real de una noche de sueño de un paciente en su domicilio, poseen artefactos que pueden empeorar la calidad de las predicciones en etapas posteriores del estudio. Estos artefactos pueden estar provocados por dos motivos principalmente: el movimiento o cambio de posición al dormir del paciente o ruido generado por el instrumental empleado para realizar las mediciones; por ejemplo, un mal ajuste del mismo alrededor del cuerpo del paciente. Por este motivo, se ha decidido llevar a cabo un filtrado de todas las señales.

El filtro empleado es un filtro de paso bajo, que permite el paso de las frecuencias más bajas, atenuando las frecuencias más altas y reduciendo así el efecto del ruido. Se ha utilizado un diseño FIR (Finite Impulse Response) en lugar del método IIR (Infinite Impulse Response) empleado habitualmente, dado que FIR afecta de forma lineal a la fase de las señales, evitando el desfase que puede provocar IIR. Además, el método FIR permite disminuir el ruido en un análisis prospectivo lineal a lo largo del tiempo. Se ha empleado una ventana de Hamming para minimizar las posibles distorsiones de fase y ondulaciones dentro de la banda de paso. La frecuencia de corte es de 1.5 Hz y la frecuencia de muestreo es de 10 Hz, como ya se ha mencionado previamente, lo que significa que se toman 10 muestras por segundo. En la figura 5.22 se muestra un segmento de una señal abdominal en el que se aplica el filtro descrito, y en el que se puede apreciar que la señal filtrada presenta una mayor suavidad, al mismo tiempo que conserva a la perfección su forma original.

Tras el filtrado, se realizó un recorte en los extremos de las señales para eliminar los artefactos provocados por los movimientos del paciente al inicio del sueño y al despertarse. Por este motivo, se decidió descartar los primeros y los últimos 20 minutos de cada registro de sueño de los pacientes, asegurando así que no se introducen variaciones no

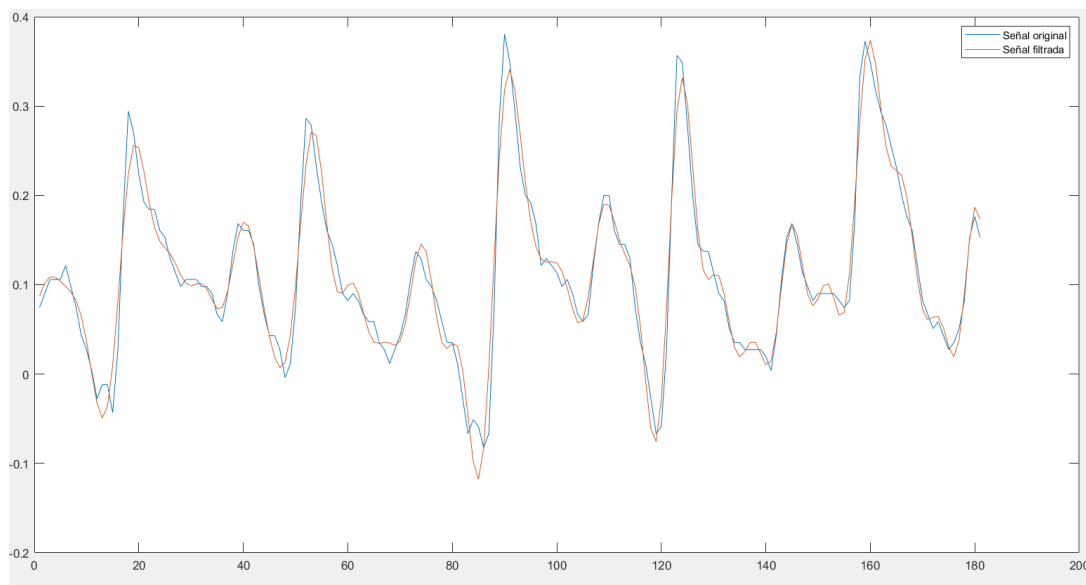


Figura 5.22: Comparación de una señal observada y esta misma señal tras aplicar un filtro de paso bajo, con FIR como método de diseño y ventana de Hamming

representativas en las señales de movimiento respiratorio durante el sueño del paciente [49].

Tras el filtrado y recorte de las señales, se ha procedido a seguir dos enfoques diferentes para el procesamiento de las señales. El primero de ellos ha sido la segmentación de las señales en fragmentos de 20 minutos no solapados, a partir de los que se calcularán los bispectrums que servirán como entrada para las redes neuronales entrenadas posteriormente. La elección de este intervalo para la segmentación de las señales no es arbitraria, sino que se debe a diferentes estudios del estado del arte que han demostrado que 20 minutos resulta ser la duración más adecuada para la detección de eventos de apnea e hipopnea [11] [85]. El segundo enfoque empleado es el uso de la señal completa de sueño del paciente. Éste, permite tener una visión global de los eventos que se producen a lo largo de toda la noche. Además, permite identificar tendencias y variaciones a más largo plazo, que pueden no ser evidentes en fragmentos más cortos [12] [13] [14].

Por último, todas las señales, tanto los fragmentos de 20 minutos como las señales de sueño completas, fueron estandarizadas mediante *z-score*. Al ajustar cada señal o fragmento de 20 minutos a una media de 0 y una desviación estándar de 1, se eliminan las diferencias de escala y amplitud entre las señales de diferentes pacientes y se minimizan los efectos causados en las señales por otras particularidades no relacionadas con la AOS. Esto hace que las señales sean comparables y facilita el entrenamiento posterior de modelos de *deep learning*.

## 5.4. Construcción de bispectrums

Un bispectrum es una herramienta empleada en el análisis de señales y series temporales para el estudio de las interacciones de las frecuencias de una misma señal. Es un *Higher Order Spectrum* (HOS) basado en el cumulante de tercer orden, es decir, la descomposición espectral de la asimetría del histograma de series temporales [6]. Este análisis HOS contiene información tanto de amplitud como de fase, y puede utilizarse para caracterizar desviaciones de gaussianidad, estacionariedad y linealidad [48]. Se representa como la matriz resultante del cómputo de la transformada de Fourier bidimensional del cumulante de tercer orden de la señal mediante la siguiente ecuación [6] [15]:

$$B(f_x, f_y) = X(f_x) \cdot X(f_y) \cdot X^*(f_x + f_y), \quad f_x, f_y = 0, \dots, f_N \quad (5.8)$$

Donde,  $f_x$  y  $f_y$  son las frecuencias asociadas a los ejes  $x$  e  $y$  de la matriz bispectral,  $f_N$  es la frecuencia de Nyquist (frecuencia de muestreo/2),  $X(t)$  es la transformada discreta de Fourier, y cada punto  $(f_x, f_y)$  de la matriz bispectral representa el grado de acoplamiento de fase entre los componente de frecuencia [6] [15].

El bispectrum mantiene la información de amplitud y fase, por lo que se puede emplear para analizar interacciones entre patrones de una señal. Además, también permite detectar cambios en la gaussianidad de las series temporales. De esta forma, valores iguales a cero en un bispectrum indican que los componentes de la señal se distribuyen de forma normal (componentes gaussianos), mientras que valores distintos de cero indican que el proceso tiene componentes no gaussianos [6] [15].

Además, el bispectrum es capaz de detectar desviaciones de la linealidad de una señal mediante el acoplamiento de fase entre sus componentes de frecuencia. El acoplamiento de fase entre tres componentes armónicas de frecuencias  $(f_1, f_2$  y  $f_3)$ , y los ángulos de fase  $(\phi_1, \phi_2$  y  $\phi_3)$ , se define como  $f_3 = f_1 + f_2$  y  $\phi_3 = \phi_1 + \phi_2$ . De esta forma, la existencia de acoplamiento de fase indica que existen relaciones de dependencia no lineales entre las componentes armónicas de la señal [6] [15].

Los bispectrums tienen propiedades de simetría, por lo que el cálculo de una región triangular resulta suficiente para describirlo completamente. Este área se conoce como la “región computacional no redundante del bispectrum”, y satisface la propiedad de  $f_x \geq 0, f_y \geq f_x$  y  $f_x + f_y \leq f_N$  [6].

El cálculo de la transformada discreta de Fourier implica el uso de una ventana temporal que se mueve a lo largo de la señal para recoger la variación de la frecuencia en cada instante de tiempo. La longitud de esta ventana afecta a la resolución temporal y de frecuencia, por lo que debe existir un compromiso entre ambas [6]. En este trabajo, para la construcción de los bispectrums se ha empleado una ventana de Hamming de 30 segundos, para asegurarnos de que el segmento contenga suficiente información de frecuencia sobre los eventos apnéicos, sin perder resolución temporal.

Los bispectrums se han construido tanto para los fragmentos de 20 minutos como para las señales completas, tanto de las señales torácicas como de las abdominales. Además, se



ha empleado un solapamiento del 50 % y se han limitado las frecuencias a un intervalo entre 0 y 1.5 Hz. Una representación de frecuencias más altas no resulta necesario dado que previamente las señales fueron filtradas en 1.5 Hz, por lo que a partir de esta frecuencia las señales no contienen ninguna información. Además, todos los bispectrums han sido normalizados.

En la figura 5.12 se pueden ver diferentes bispectrums de fragmentos de señales de 20 minutos. En la parte (a) se muestran dos bispectrums de la misma señal respiratoria de un paciente sin eventos apneicos. En la parte izquierda se muestra la señal de movimiento respiratorio torácico, y en la parte derecha la señal de movimiento respiratorio abdominal. En la parte (b) de la figura se muestran otros dos bispectrums de otra señal que sí posee eventos apneicos. De nuevo, en la parte izquierda se muestra la señal de movimiento respiratorio torácico, y en la parte derecha la señal de movimiento respiratorio abdominal.

En esta figura se puede ver claramente cómo la respiración normal del paciente, sin eventos apneicos, se encuentra en una frecuencia de alrededor de 0.25 Hz, es decir, una respiración cada 4 segundos. En cambio, la ocurrencia de eventos apneicos provoca una reducción de la potencia de la banda de respiración normal, haciendo que aparezcan nuevos focos de acoplamiento a bajas frecuencias, alrededor de 0.1 Hz.

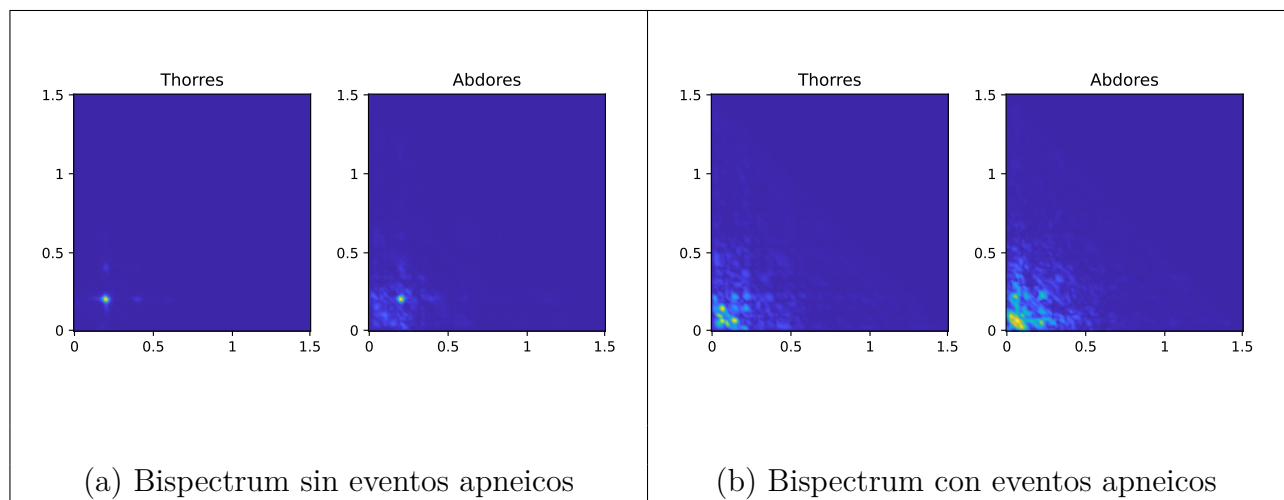


Tabla 5.12: Representación de bispectrum sin (a) y con (b) eventos apneicos para las señales de respiración torácica y abdominal



---

## 6: Metodología experimental

---

El objetivo de este estudio es la predicción del grado de severidad de la apnea del sueño de pacientes, a partir de señales fisiológicas recogidas durante una noche de sueño en su domicilio. Para ello, se han utilizado diversas arquitecturas de redes neuronales, incluyendo tanto algunas reconocidas en el estado del arte, como arquitecturas de diseño propio, ya implementadas en fases anteriores de la investigación.

Como ya se ha mencionado previamente, se han seleccionado las señales de esfuerzo respiratorio [THOR](#) y [ABDO](#) para la construcción de los bispectrums que sirven como datos de entrada para los modelos de redes neuronales que se van a entrenar. Tal y como recomienda la [AASM](#), el estudio de estas dos señales de forma conjunta resulta indispensable para identificar los eventos apneicos centrales y poder distinguirlos de los de tipo obstructivo [9]. La diferenciación de estos tipos de eventos resulta fundamental para proporcionar a los pacientes un tratamiento personalizado y eficaz.

En este capítulo se van a describir las diferentes aproximaciones seguidas para la resolución del problema propuesto, así como una descripción detallada de las redes neuronales empleadas, desde la entrada de los dos datos, las arquitecturas empleadas, la estrategia de aprendizaje seguida y la estrategia para la validación de los modelos empleados.

### 6.1. Variantes del problema

Dada la complejidad del problema abordado y la gran cantidad de ruido presente en las señales estudiadas, se ha decidido enfocar la estimación de la severidad de la apnea del sueño desde múltiples perspectivas, con el objetivo de identificar la solución más eficaz y más adecuada para resolver este problema.

1. **Problema clásico de regresión:** el objetivo de esta variante es la predicción del número de eventos apneicos, tanto totales como centrales, en cada fragmento de señal de 20 minutos. A partir de la predicción del número de eventos apneicos en cada fragmento, se calcula la estimación del [IAH](#) y del [IAC](#) de cada individuo.

2. **Problema clásico de clasificación:** esta variante trata de ser una simplificación del problema, en la que la red tan solo deba predecir si existe o no algún tipo de evento apneico en un fragmento de señal de 20 minutos. El umbral definido para determinar la clase con presencia o ausencia de eventos es 5 eventos por hora, ya que en este umbral se registra una alta prevalencia de AOS, debido a la sensibilidad de las técnicas de grabación y los criterios de puntuación actuales [74]. Este problema de clasificación se aplica tanto a la predicción de la presencia o ausencia de cualquier tipo de evento apneico, como de los eventos centrales en concreto.
3. **Problema de regresión acompañado de una regresión lineal:** esta variante toma como base el problema clásico de regresión, pero tras la estimación en cada época del número de eventos apneicos en cada fragmento de 20 minutos, se calcula el IAH o el IAC de cada individuo y se ajusta con estos una regresión lineal cuyo resultado se emplea para el cálculo de las métricas de validación del modelo. El propósito del uso de esta regresión lineal es la corrección de la tendencia de las CNN a subestimar o sobrestimar el IAH y el IAC.
4. **Problema de regresión acompañado de una regresión lineal con un nuevo mapa de color:** en esta variante se emplea la misma arquitectura y estrategias de aprendizaje que la anterior, pero se introduce una modificación en la representación de los bispectrums. En lugar de utilizar el mapa de colores previamente empleado, se opta por uno nuevo que abarca una gama más amplia de tonalidades. Este cambio tiene como objetivo resaltar con mayor intensidad los eventos presentes en los bispectrums, facilitando una diferenciación visual más clara entre ellos, para que las redes convolucionales también sean capaces de distinguir de forma más efectiva los eventos presentes en cada señal. En las figuras 6.23a y 6.23b se presenta la misma señal de esfuerzo respiratorio (ABDO a la izquierda y THOR a la derecha). En la figura 6.23a se puede ver el mapa de color empleado en las tres primeras variantes, mientras que en la figura 6.23b se puede ver el nuevo mapa de color empleado en esta variante para la misma señal. Como se puede apreciar, en este nuevo mapa de color, las alteraciones bispectrales causadas por los eventos apneicos aparecen más marcadas y resulta más sencillo visualizarlas.

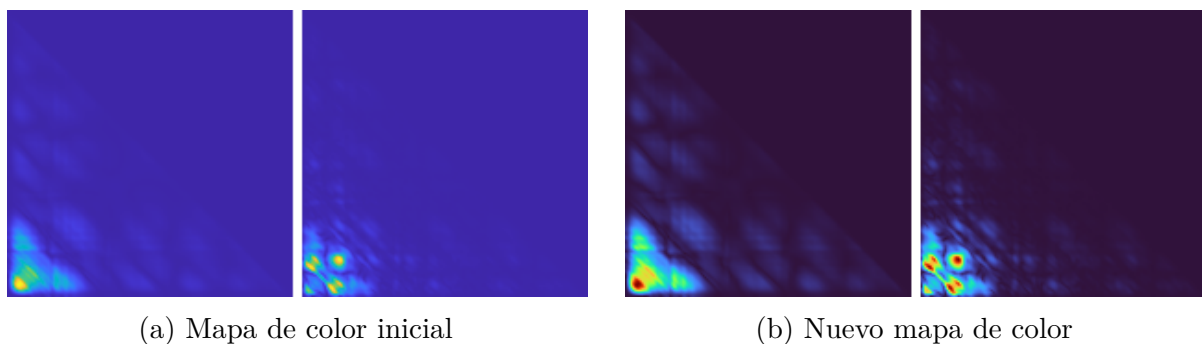


Figura 6.23: Comparación de los mapas de color empleados en las distintas variantes

5. **Problema de regresión acompañado de una regresión lineal para la señal completa:** esta última variante emplea la misma arquitectura y estrategias de aprendizaje que las dos anteriores, pero en vez de emplear bispectrums que representan fragmentos de 20 minutos de las señales, se emplean bispectrums que representan la señal completa de una noche de sueño de un paciente. Además, la salida de los modelos de redes convolucionales será la estimación del [IAH](#) o del [IAC](#), en lugar de la estimación del número de eventos apneicos totales o centrales presentes en el fragmento de señal. El motivo se encuentra en que el uso de la señal completa permite tener una visión global de los eventos a lo largo de toda la noche, así como las interacciones entre las frecuencias, que pueden no ser evidentes en fragmentos más cortos [\[12\]](#) [\[13\]](#) [\[14\]](#).

## 6.2. Entrada de datos

Los datos que se van a tomar como entrada para los modelos de redes neuronales convolucionales constan de dos imágenes de tamaño  $(155, 155, 3)$ . Estas imágenes son los dos bispectrums de las señales de esfuerzo respiratorio [ABDO](#) y [THOR](#) provenientes de fragmentos de 20 minutos o de señales completas, según la variante del problema. Estas dos señales son complementarias, dado que reflejan diferentes formas de esfuerzo respiratorio (torácico y abdominal). Por este motivo, resulta esencial analizarlas de manera conjunta para obtener una comprensión completa del patrón respiratorio del paciente.

Para integrar ambas señales, los dos bispectrums se combinan en una única imagen, pero distribuidos en canales diferentes, sumando un total de 6 canales al considerar los tres canales RGB de cada imagen. Este enfoque permite que las redes neuronales convolucionales procesen simultáneamente ambas fuentes de información. La figura [6.24a](#) muestra el esquema de esta distribución. En algunos experimentos excepcionales, se ha empleado una alternativa donde las imágenes se disponen una al lado de la otra, resultando en una entrada a la red con dimensiones  $(310, 155, 3)$ , como se muestra en la figura [6.24b](#).

## 6.3. Arquitecturas

Con el objetivo de la estimación de la severidad de la apnea del sueño para los pacientes participantes en el [SHHS](#), se han empleado diferentes arquitecturas de redes neuronales. En primer lugar, se han empleado algunas arquitecturas reconocidas del estado del arte, como ResNet, EfficientNet o DenseNet; todas ellas ya descritas previamente en el documento. Concretamente, de la familia ResNet se han empleado los modelos con 18 y 50 capas; de la familia EfficientNet se han empleado las variantes B0, B1 y B2; y de la familia DenseNet se ha empleado la variante 121. El motivo se encuentra fundamentalmente en su balance óptimo entre precisión y coste computacional.

Gran parte de la experimentación realizada con arquitecturas del estado del arte ha utilizado modelos basados en la arquitectura ResNet. Una de las variantes implementadas es un *encoder* doble, donde cada bispectrum de las señales torácicas y abdominales se

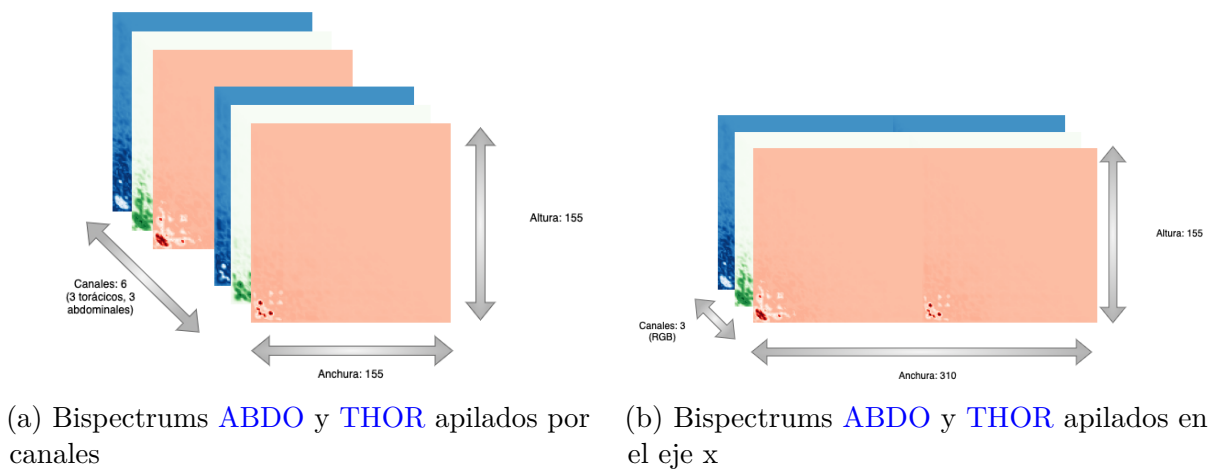


Figura 6.24: Distribución de los bispectrums ABDO y THOR como datos de entradas de los modelos

utiliza como entrada para un modelo ResNet18 independiente. Las características extraídas de cada uno de estos modelos se combinan antes de la capa final del modelo, que es una capa *fully-connected*. Este enfoque permite capturar y fusionar la información específica y complementaria de ambas señales respiratorias, mejorando así la capacidad del modelo para detectar y analizar eventos respiratorios complejos con mayor precisión. En la figura 6.25 se muestra un ejemplo de este modelo de doble ResNet18 desarrollado de forma específica para el problema de estimación de la severidad de la apnea del sueño a partir de dos señales.

Otra variante empleada de la arquitectura ResNet es ResNet-RS, introducida por Google en 2021 [7]. Esta arquitectura se presentó como una mejora en el rendimiento para diversas tareas de visión, manteniendo la eficiencia computacional. ResNet-RS aporta dos nuevas estrategias de escalado. La primera consiste en el escalado de la profundidad del modelo, para mejorar el rendimiento sin incrementar significativamente el riesgo de sobreajuste. La segunda estrategia trata de aumentar la resolución de la imagen de manera más gradual que en modelos anteriores, como los modelos de la familia EfficientNet. Gracias a estas mejoras, ResNet-RS logra un entrenamiento más rápido que los modelos ResNet convencionales, al mismo tiempo que mantienen precisiones similares en conjuntos de datos como ImageNet.

La arquitectura ConvNeXt, presentada por Facebook en 2022 [42], también se ha empleado en los experimentos de una de las variantes del problema abordado. Está inspirada en los modelos basados en *transformers* para obtener una versión mejorada de ResNet. Esta arquitectura se ha diseñado como alternativa a los *vision transformers* jerárquicos en tareas como clasificación de imágenes, detección de objetos y segmentación semántica. ConvNeXt tiene distintas variantes, como ConvNeXt-T, ConvNeXt-S, ConvNeXt-B, ConvNeXt-L y ConvNeXt-XL, cada una con un número variable de canales y bloques en cada etapa. A diferencia de otros modelos, ConvNeXt no incorpora *vision transformers* jerárquicos ni

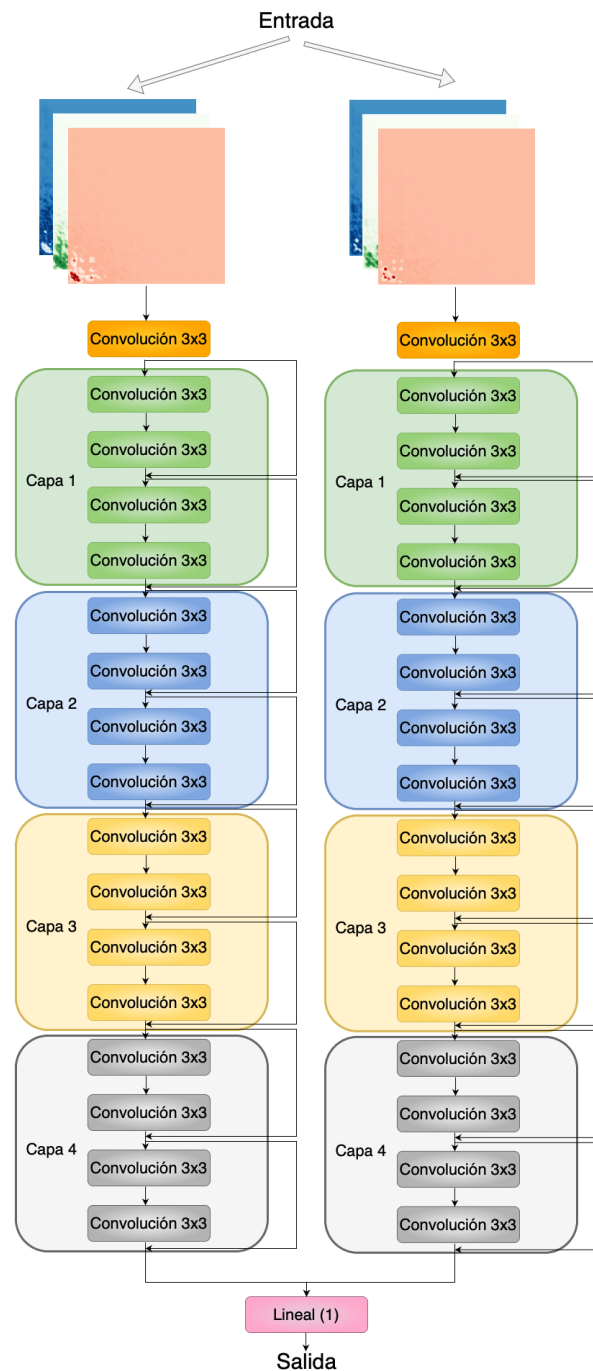


Figura 6.25: Arquitectura Double ResNet 18

módulos basados en atención, sino que se centra en modernizar una red convolucional tradicional (como ResNet) hacia un diseño similar al de un *vision transformer* jerárquico como Swin.

Además de modelos reconocidos en el estado del arte, se ha empleado otra arquitectura de CNN personalizada, diseñada especialmente para el problema de la predicción de la

severidad de la apnea del sueño a partir de las señales de esfuerzo respiratorio torácico y abdominal [4]. Esta arquitectura, presentada en la figura 6.26, ha demostrado ser particularmente eficaz para estimar el número de eventos centrales y está inspirado en otros modelos desarrollados para la estimación de la AOS pediátrica utilizando señales de saturación de oxígeno en sangre ( $SpO_2$ ) y/o flujo aéreo [33] [85].

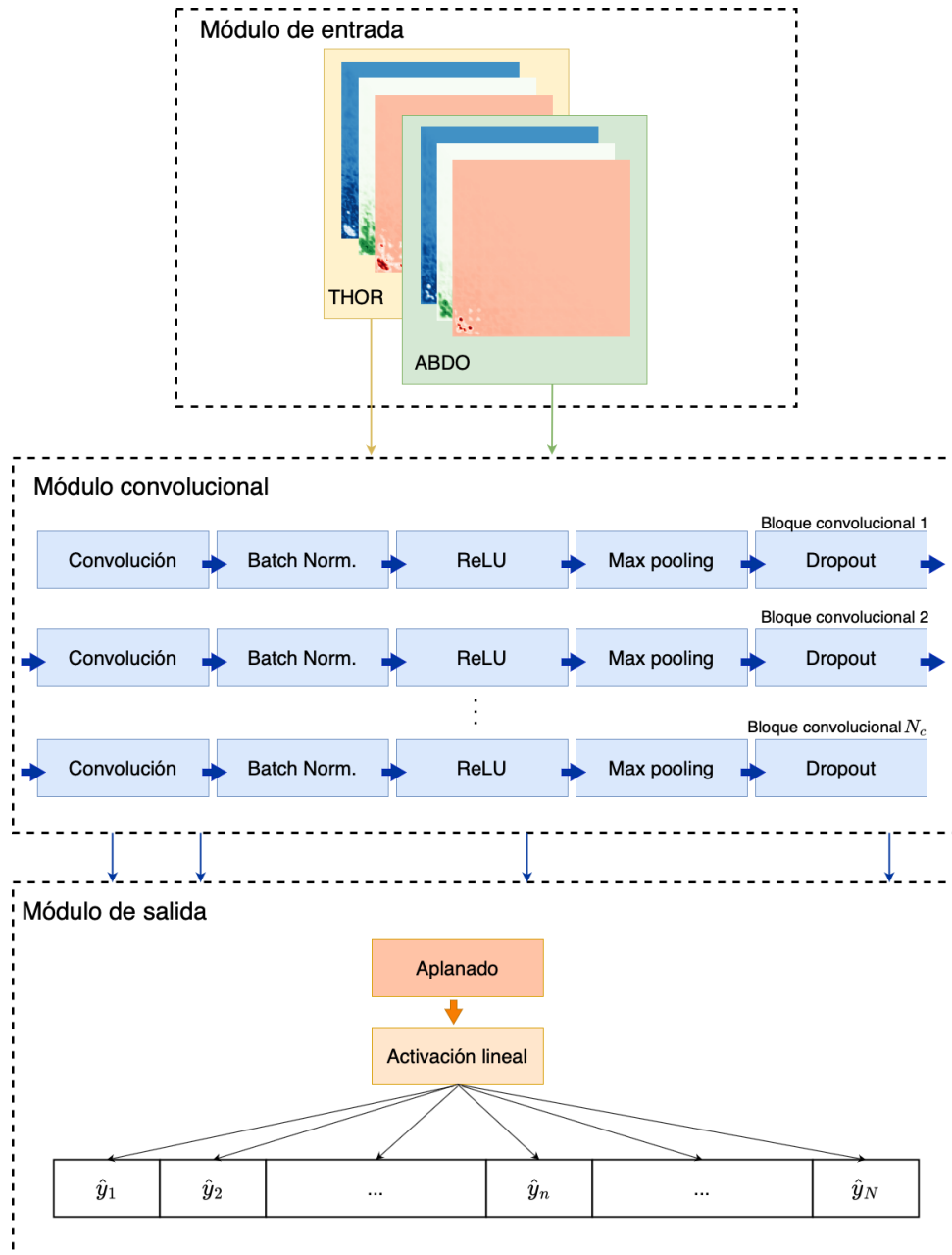


Figura 6.26: Arquitectura de red CNN de diseño propio

Esta arquitectura de diseño propio toma como entrada los bispectrums de las señales de esfuerzo respiratorio THOR y ABDO y está compuesta por  $N_c$  bloques convolucionales.

Cada bloque consta de una capa convolucional bidimensional, una normalización del *batch*, una capa de activación ReLU, una capa de *max pooling* y una capa de *dropout*. La capa convolucional aplica un *kernel* a los datos de entrada para generar un mapa de características. El tamaño del *padding* es 0, el *stride* utilizado es 1, y el número de bloques convolucionales empleado es 4 y 6. Además, se ha empleado un *dropout* de 0.1 al final de cada bloque convolucional para reducir el sobreajuste reduciendo la interdependencia entre las neuronas de la red. Tras los bloques convolucionales se realiza un aplanado de los mapas de características extraídos en una matriz unidimensional y finalmente se envía esta matriz a una función de activación lineal que devuelve la salida de la estimación del número de eventos contenidos en cada fragmento de señal, o el IAH o el IAC presente en una señal completa de sueño.

## 6.4. Estrategia de aprendizaje

Una vez seleccionadas las arquitecturas de redes convolucionales del estado del arte y desarrolladas las arquitecturas de diseño propio para el problema de estimación de la severidad de la apnea del sueño a partir de las señales de esfuerzo respiratorio torácico y abdominal, se procede a entrenar estos modelos. Para ello, se ha empleado un total de 150 épocas para todas las variantes del problema, excepto para la variante del problema clásico de regresión, en la que se han empleado 50 épocas, y para la variante que analiza la señal completa de una noche de sueño de un paciente, en lugar de fragmentos de 20 minutos, en la que se emplean 2000 épocas para cada entrenamiento. El motivo se encuentra en que, en las variantes que emplean fragmentos de 20 minutos como datos de entrada, el número total de instancias de entrenamiento es de 64364, mientras que para la variante que emplea la señal completa, el total de instancias es de 2606. Debido a esta menor cantidad de datos, comparada con las otras variantes, el modelo que usa la señal completa requiere un mayor número de épocas para aprender los patrones subyacentes en los datos de manera efectiva.

El optimizador empleado para todos los experimentos fue *Adaptive Moment Estimation Weight Decay* (AdamW) [45]. El algoritmo de Adam [37] destaca por su rápida convergencia, manejo eficiente de gradientes y robustez frente a la variabilidad de los datos y los hiperparámetros. El algoritmo de AdamW desacopla el decaimiento de los pesos de los gradientes y aplica en su lugar este decaimiento de los pesos directamente a los parámetros del modelo, permitiendo que el algoritmo tenga una mejor convergencia y capacidad de generalización que Adam.

Junto con el algoritmo AdamW se ha empleado una tasa de aprendizaje de 0.001, siguiendo el trabajo ya realizado en [4]. Por este mismo motivo, se ha empleado un tamaño de *batch* de 32. Además, este tamaño de *batch* permite mantener un equilibrio entre la eficiencia computacional y el rendimiento del modelo.

Como criterio de optimización en el caso de los problemas de regresión se ha empleado la función de pérdida de Huber, definida en la siguiente ecuación [30]:

$$L_{\delta} = \begin{cases} \frac{1}{2} \cdot (y - f(x))^2 & \text{si } |y - f(x)| \leq \delta \\ \delta \cdot \left( |y - f(x)| - \frac{1}{2} \cdot \delta \right) & \text{resto} \end{cases} \quad (6.9)$$

Donde  $L_{\delta}$  representa la pérdida de Huber,  $d$ , el punto en el que la pérdida pasa de un comportamiento cuadrático a uno no lineal,  $y$  es el número real de eventos apneicos y  $f(x)$  es el valor predicho por el modelo. Esta función combina las características del error cuadrático medio y el error absoluto medio, proporcionando una respuesta menos sensible a los valores atípicos en los datos [30] [4]. El parámetro  $d$  se ha fijado en 1, ya que ha demostrado ser una elección apropiada para distribuciones aleatorias [30].

En el caso de los problemas de clasificación, se ha empleado como función de pérdida la entropía cruzada, definida en la siguiente ecuación para una clasificación binaria:

$$L_{\delta} = -(y \cdot \log(f(x)) + (1 - y) \cdot \log(1 - f(x))) \quad (6.10)$$

## Optimización de hiperparámetros

En la búsqueda del mejor modelo de predicción de la severidad de la apnea del sueño, se ha trabajado desde los cinco enfoques mencionados previamente. A lo largo de la experimentación desde estos cinco enfoques y en la búsqueda del mejor modelo, se ha tratado de optimizar una serie de hiperparámetros, con el objetivo de encontrar los que mejores resultados proporcionen. La tabla 6.13 muestra los hiperparámetros considerados y sus respectivos espacios de búsqueda.

| Hiperparámetros                  | Espacio de búsqueda                                                     |
|----------------------------------|-------------------------------------------------------------------------|
| <b>Modelos</b>                   | ResNet, EfficientNet, DenseNet, ResNetRS, ConvNext, modelos propios     |
| <b>Tamaño <i>batch</i></b>       | 4, 32                                                                   |
| <b>Función de <i>loss</i></b>    | entropía cruzada (clasificación), Huber, L1, MSE, smooth L1 (regresión) |
| <b>Tasa de aprendizaje</b>       | 0.001, <i>fit-one-cycle</i>                                             |
| <b>Inicialización de pesos</b>   | aleatoria, no aleatoria                                                 |
| <b>Umbral clasificación</b>      | 0, 1, 5, 10, 15 (todos los eventos), 0, 1, 5 (central)                  |
| <b>Frecuencias representadas</b> | 1 Hz, 1.5 Hz                                                            |
| <b>Distribución de los datos</b> | THOR, ABDO, independientes, conjunto, apilados                          |

Tabla 6.13: Hiperparámetros explorados

En primer lugar, como ya se ha comentado en ocasiones anteriores, se han evaluado diversas familias de arquitecturas del estado del arte, como ResNet, EfficientNet, DenseNet,



modelos como ResNetRS o ConvNext y modelos de diseño propio. Se ha empleado por defecto un tamaño de *batch* de 32, en cambio, en algunos experimentos se ha explorado un tamaño de 4, tratando de obtener mejores resultados, penalizando el tiempo de ejecución del modelo.

Respecto a la función de *loss*, en el caso de los problemas de clasificación, se ha empleado una entropía cruzada y en el caso de los problemas de regresión, se ha empleado fundamentalmente la función de pérdida de Huber, aunque también se han explorado otras funciones de *loss* como L1, *Mean Squared Error* (MSE) o *smooth L1*. La tasa de aprendizaje utilizada en la mayoría de los experimentos ha sido 0.001, aunque también se ha explorado la posibilidad de ajustarla de forma dinámica empleando el método de *fit-one-cycle*. Además, en el caso de los experimentos entrenados a partir de modelos del estado del arte, se ha explorado la posibilidad de emplear una inicialización aleatoria de los pesos, como emplear los pesos de estos modelos pre-entrenados.

En los experimentos entrenados para la variante de clasificación, se ha llevado a cabo una exploración de los posibles umbrales que definen la presencia o ausencia de apnea en las señales respiratorias. Por defecto, como ya se ha mencionado previamente, este umbral se ha definido en 5 e/h, tanto para todos los tipos de eventos, como para los centrales. También se han explorado los umbrales 0, 1, 10 y 15 para todos los tipos de eventos, y 0 y 1 para los eventos de tipo central. Debido a las frecuencias de respiración normal y las frecuencias a las que se producen los eventos apneicos, también se ha explorado la posibilidad de representar en los bispectrums una frecuencia menor a la del límite de 1.5 Hz establecido en el filtrado de las señales, en concreto 1 Hz, con el objetivo de facilitar el aprendizaje a los modelos convolucionales, limitando la región de la imagen en la que identificar los eventos apneicos.

Por último, principalmente en los experimentos de la variante del problema de clasificación, se ha tratado de averiguar cuál es la mejor forma de introducir los datos en el modelo. Por defecto, como se recomienda en el estado del arte, se introducen las dos señales de esfuerzo respiratorio de forma conjunta en el modelo, distribuidas en dos canales. También se ha explorado la posibilidad de apilar ambas señales en el eje x, tratarlas como señales independientes, o emplear únicamente una u otra señal (THOR o ABDO), con el objetivo de ver si alguna de las dos señales contenía información más relevante para la estimación de la severidad de la apnea del sueño.

## Regresión sobre el IAH

En este estudio se trata de predecir la severidad de la apnea del sueño de los pacientes de dos formas diferentes, en primer lugar, estimando el IAH a partir del número de eventos apneicos, tanto obstructivos, como centrales e hipopneas. En segundo lugar, estimando el IAC a partir del número de eventos apneicos centrales. En el caso de las variantes que estiman el número de eventos apneicos a partir de los fragmentos de señal de 20 minutos, el IAH y el IAC por sujeto se calculan a partir de la suma de eventos por segmentos contenidos en la señal, dividida por la suma de la duración en horas de los segmentos contenidos

en la señal de una noche completa de sueño. En el caso de la variante que emplea como datos de entrada los bispectrums de las señales, el IAH y el IAC por sujeto se obtienen directamente de la estimación proporcionada por los modelos de redes convolucionales.

Tras el cálculo del IAH y del IAC por sujeto, para tres de las variantes del problema se ajusta un modelo de regresión lineal siguiendo la siguiente ecuación:

$$AHI_i = (\beta \cdot AHI_{CNN_i}) + \epsilon, \quad i = \text{global, central} \quad (6.11)$$

Donde  $AHI_{CNN_i}$  es el IAH o IAC por individuo,  $\beta$  es el coeficiente de la pendiente y  $\epsilon$  es el *intercept* del modelo de regresión lineal obtenido a través del método de mínimos cuadrados ordinarios (OLS). Tanto  $\beta$  como  $\epsilon$  se han ajustado empleado el conjunto de datos de entrenamiento. El objetivo de esta regresión lineal es corregir la tendencia de las CNN a subestimar o sobrestimar el IAH y el IAC. Uno de los principales motivos es que las CNN tienden a sobreestimar las hipopneas al detectar reducciones del esfuerzo respiratorio que no van acompañadas de desaturaciones de oxígeno del 3 % o despertares [4].

## 6.5. Estrategia de validación

Para la evaluación del rendimiento de cada experimento se ha empleado la técnica de *hold-out*, en la que el conjunto de datos se divide en tres grupos: entrenamiento, validación y test. Esta técnica de *hold-out* se ha empleado debido a la gran cantidad de datos disponibles y la necesidad de una evaluación eficiente del rendimiento del modelo. Aunque otras técnicas como la validación cruzada ofrecen una evaluación más robusta, *hold-out* proporciona un equilibrio adecuado entre complejidad computacional y precisión de la evaluación.

El conjunto de entrenamiento se utiliza para ajustar los pesos del modelo, el conjunto de validación sirve para ajustar los hiperparámetros y prevenir el sobreajuste, y el conjunto de test permite evaluar el rendimiento final del modelo. El conjunto de entrenamiento consta de 2606 sujetos, el conjunto de validación de 651 y el de test de 5000. La justificación del reparto del total de los 8257 individuos entre los tres conjuntos ya se mencionó previamente y se muestra en la figura 5.21.

Para los problemas de regresión, una vez estimado el IAH o el IAC para cada individuo, se discretiza siguiendo los cuatro grupos de severidad definidos por la AASM: no apnea ( $IAH < 5$  e/h), leve ( $5 \text{ e/h} \leq IAH < 15$  e/h), moderada ( $15 \text{ e/h} \leq IAH < 30$  e/h) y grave ( $IAH \geq 30$  e/h).

Tomando el IAH o el IAC discretizados, o las clases de la variante del problema de clasificación, se calcula el índice kappa de Cohen ( $k$ ) para evaluar el rendimiento del modelo [16]. Este índice mide la concordancia entre las observaciones y los valores reales, teniendo en cuenta que el acuerdo se podría dar por azar, por lo que es una medida más

robusta que el simple cálculo de una tasa de acierto. A continuación se muestra la formula empleada para su cálculo:

$$k = \frac{P_o - P_e}{1 - P_e} \quad (6.12)$$

Donde  $P_o$  es la proporción de acuerdo observado entre los evaluadores y  $P_e$  es la proporción de acuerdo esperado por azar. Esta métrica puede tomar valores entre 0 y 1, representando 0 un acuerdo peor que el azar (no hay acuerdo) y 1, un acuerdo perfecto.

El índice kappa no solo se utiliza para evaluar el rendimiento de cada uno de los modelos sobre el conjunto de datos de test para determinar cuál es el que proporciona mejores resultados. También se utiliza para evaluar la concordancia entre las predicciones y los valores reales en el conjunto de datos de validación. Esto permite seleccionar el modelo que proporciona los mejores resultados y determinar en qué época del modelo se obtienen, lo que ayuda a reducir el sobreajuste.

---

## 7: Resultados

---

En este capítulo se van a revisar los resultados de los experimentos entrenados tras el diseño experimental comentado en el capítulo anterior. Se van a analizar los resultados para cada una de las cinco variantes del problema abordadas en este trabajo, tratando de ver si alguna de ellas muestra una mejor comprensión de los datos y por lo tanto, también mejores resultados. Para cada variante, se muestran los resultados del índice kappa y la tasa de acierto (*accuracy*) para cada modelo, tanto en entrenamiento como en validación. A continuación se van a seleccionar los problemas de cada variante que mejores resultados proporcionen, se analizarán sus resultados sobre el conjunto de test y a través de las técnicas de [IA](#) explicable, se van a tratar de justificar los resultados obtenidos.

### 7.1. Problema clásico de regresión

Para la variante del problema clásico de regresión, se han explorado diferentes arquitecturas del estado del arte, así como arquitecturas de diseño propio. En la tabla [7.14](#) se muestran los resultados obtenidos para cada experimento para la estimación del [IAH](#) a partir de todos los eventos apneicos, y en la tabla [7.15](#), los resultados para la estimación del [IAC](#) a partir de los eventos apneicos centrales.

| Modelo          | Kappa train | Kappa val.   | Accuracy train | Accuracy val. |
|-----------------|-------------|--------------|----------------|---------------|
| Resnet 18       | 0.813       | 0.296        | 0.877          | 0.533         |
| Efficientnet b0 | 0.545       | 0.301        | 0.704          | 0.531         |
| Efficientnet b1 | 0.557       | <b>0.321</b> | 0.712          | 0.559         |
| Efficientnet b2 | 0.568       | 0.298        | 0.719          | 0.525         |
| Densenet 121    | 0.763       | 0.261        | 0.844          | 0.504         |
| Red propia 4c   | 0.242       | 0.248        | 0.515          | 0.516         |
| Red propia 6c   | 0.242       | 0.249        | 0.515          | 0.498         |

Tabla 7.14: Resultados problema clásico de regresión (todos los eventos apneicos)

Para la estimación del [IAH](#), se puede ver que la familia EfficientNet proporciona resultados ligeramente mejores sobre el conjunto de validación, especialmente en el caso de la variante b1. Sin embargo, todos los experimentos proporcionan resultados con una baja capacidad de generalización, y en los modelos tomados del estado del arte, se evidencia un claro sobreajuste, dado que el índice kappa es significativamente más alto en entrenamiento que en validación.

| Modelo                 | Kappa train | Kappa val.   | Accuracy train | Accuracy val. |
|------------------------|-------------|--------------|----------------|---------------|
| <b>Resnet 18</b>       | 0.641       | <b>0.452</b> | 0.987          | 0.991         |
| <b>Efficientnet b0</b> | 0.357       | 0.220        | 0.980          | 0.988         |
| <b>Efficientnet b1</b> | 0.368       | 0.368        | 0.982          | 0.985         |
| <b>Efficientnet b2</b> | 0.377       | 0.327        | 0.981          | 0.985         |
| <b>Densenet 121</b>    | 0.349       | 0.380        | 0.980          | 0.986         |
| <b>Red propia 4c</b>   | 0.098       | 0.220        | 0.976          | 0.989         |
| <b>Red propia 6c</b>   | 0           | 0            | 0.976          | 0.988         |

Tabla 7.15: Resultados problema clásico de regresión (eventos centrales)

Para la estimación del [IAC](#), por lo general, se obtienen resultados ligeramente superiores a los obtenidos para la estimación del [IAH](#). El modelo ResNet 18 destaca por su alto índice kappa en validación, en comparación con el resto de experimentos. También resulta importante destacar la elevada tasa de acierto obtenida en validación para la estimación del [IAC](#). Esto sugiere que los resultados podrían estar influenciados por el gran desbalanceo entre las clases de los distintos grupos de severidad, por lo que es posible que la red esté clasificando todos o casi todos los sujetos en una misma clase, en vez de tratar de capturar adecuadamente la diversidad y distribución de las clases.

## 7.2. Problema de clasificación

Dados los discretos resultados obtenidos en la primera variante del problema, se ha tratado de simplificar hacia una clasificación de los sujetos, en vez de estimar su [IAH](#) o [IAC](#). En la tabla 7.16 y 7.17 se muestran los resultados del problema de clasificación de los sujetos según la presencia o ausencia de eventos apneicos de cualquier tipo o eventos apneicos centrales, respectivamente.

Para la clasificación de los sujetos en base al número total de eventos apneicos, se aprecia como el mejor modelo sigue siendo el modelo EfficientNet b1, con un resultado ligeramente mejor al obtenido en el problema clásico de regresión. De nuevo se evidencia un alto sobreajuste en los experimentos que emplean modelos del estado del arte, y una baja capacidad de generalización. Las redes de diseño propio muestran un desempeño mucho más bajo en comparación con los modelos preentrenados.

Para la clasificación de los sujetos en base al número de eventos centrales, se puede ver que, aunque el modelo Double ResNet 18 es el que proporciona un mejor resultado, todos

| Modelo           | Kappa train | Kappa val. | Accuracy train | Accuracy val. |
|------------------|-------------|------------|----------------|---------------|
| Resnet 18        | 0.997       | 0.357      | 0.998          | 0.696         |
| Double ResNet 18 | 0.97        | 0.347      | 0.985          | 0.68          |
| EfficientNet b1  | 0.950       | 0.371      | 0.976          | 0.695         |
| DenseNet 121     | 0.989       | 0.366      | 0.995          | 0.69          |
| Red propia 4c    | 0.381       | 0.338      | 0.707          | 0.685         |
| Red propia 6c    | 0           | 0          | 0.588          | 0.596         |

Tabla 7.16: Resultados clasificación con umbral 5 e/h (todos los eventos apneicos)

los resultados son peores que los obtenidos mediante la variante de regresión. La tasa de acierto es muy alta, con valores cercanos al 100 %, lo que nos está indicando que la red no está aprendiendo, sino clasificando a la gran mayoría de individuos en una de las dos clases.

| Modelo           | Kappa train | Kappa val. | Accuracy train | Accuracy val. |
|------------------|-------------|------------|----------------|---------------|
| Resnet 18        | 0.970       | 0.204      | 0.999          | 0.989         |
| Double ResNet 18 | 0.95        | 0.244      | 0.999          | 0.99          |
| EfficientNet b1  | 0.932       | 0.224      | 0.999          | 0.992         |
| DenseNet 121     | 0.986       | 0.196      | 1              | 0.985         |
| Red propia 4c    | 0           | 0          | 0.989          | 0.994         |
| Red propia 6c    | 0           | 0          | 0.989          | 0.994         |

Tabla 7.17: Resultados clasificación con umbral 5 e/h (eventos centrales)

En base a los resultados obtenidos para esta variante del problema de clasificación, se han dedicado esfuerzos a tratar de encontrar algún hiperparámetro que mejore el rendimiento diagnóstico. Para estos experimentos, se ha seleccionado el modelo ResNet 18, por su gran popularidad y amplio uso y buenos resultados en el estado del arte de *transfer learning*. En la tabla 7.18 y en la tabla 7.19 se muestran los resultados obtenidos en la exploración de los hiperparámetros para detectar la presencia o ausencia de eventos apneicos de cualquier tipo o centrales, respectivamente.

Para la clasificación de los sujetos en base al número total de eventos apneicos, se puede apreciar que casi ninguno de los experimentos mejora el resultado obtenido en validación para el modelo ResNet 18 de la tabla 7.16. Solo se observan mejoras en los modelos cuando se ajusta el umbral para determinar la presencia de eventos apneicos: cuanto mayor es este umbral, mejores resultados se obtienen. Esto podría deberse a que es común en el conjunto de datos que los sujetos tengan un [IAH](#) alrededor de 5 eventos por hora. Como resultado, existen fragmentos de señal con un número de eventos casi igual, pero que se clasifican

en grupos diferentes debido al umbral aplicado. Estas señales, aunque muy similares, son tratadas como distintas, lo que confunde a la red y dificulta su capacidad para generalizar correctamente.

El resultado obtenido con la representación de frecuencias más bajas, 1 Hz en lugar de 1.5 Hz es ligeramente inferior. El resto de experimentos en los que se exploran diferentes formas de introducir los datos en el modelo, inicialización aleatoria de los pesos del modelo y ajuste dinámico del *learning rate*, tampoco mejoran el resultado obtenido con el modelo ResNet 18 básico.

| Variantes modelo ResNet 18 | Kappa train | Kappa val. | Accuracy train | Accuracy val. |
|----------------------------|-------------|------------|----------------|---------------|
| Umbral 0                   | 0.950       | 0.331      | 0.983          | 0.763         |
| Umbral 1                   | 0.961       | 0.341      | 0.984          | 0.741         |
| Umbral 10                  | 0.967       | 0.396      | 0.983          | 0.706         |
| Umbral 15                  | 0.965       | 0.401      | 0.986          | 0.742         |
| 1Hz                        | 0.994       | 0.370      | 0.997          | 0.662         |
| THOR                       | 0.967       | 0.311      | 0.984          | 0.654         |
| ABDO                       | 0.965       | 0.339      | 0.983          | 0.670         |
| Señales apiladas           | 0.970       | 0.340      | 0.985          | 0.655         |
| Señales independientes     | 0.944       | 0.331      | 0.973          | 0.680         |
| Rand init                  | 0.9731      | 0.286      | 0.9870         | 0.6852        |
| Ajuste learning rate       | 0.927       | 0.343      | 0.965          | 0.685         |

Tabla 7.18: Resultados clasificación modelo ResNet 18 con exploración de grano grueso de diferentes hiperparámetros (todos los eventos apneicos)

Para las diferentes variantes del modelo ResNet 18 para la clasificación de los sujetos en base al número de eventos centrales, ninguno de los experimentos ha obtenido un resultado que mejore el obtenido en el experimento base del modelo ResNet 18. Como se ha visto para la clasificación teniendo en cuenta todos los tipos de eventos apneicos, el uso de un umbral más alto provoca que los resultados de los experimentos sean mejores. En este caso, un umbral de 5 e/h sí es adecuado ya que el número de eventos apneicos centrales en las señales es inferior al número total de eventos. Por otro lado, se puede ver que el uso de las señales apiladas en el eje x, después del método habitual, proporciona mejores resultados que el resto de experimentos de entrada de datos. También se aprecia que es posible que las señales de esfuerzo respiratorio abdominal proporcionen algo más de información que las torácicas, dado que tanto para la estimación de los eventos apneicos centrales como para el total de eventos de cualquier tipo, el índice kappa es superior.

| Variantes modelo<br>ResNet 18 | Kappa train | Kappa val. | Accuracy<br>train | Accuracy<br>val. |
|-------------------------------|-------------|------------|-------------------|------------------|
| Umbral 0                      | 0.990       | 0.151      | 0.999             | 0.919            |
| Umbral 1                      | 0.979       | 0.157      | 0.999             | 0.924            |
| 1Hz                           | 0.982       | 0.214      | 0.999             | 0.992            |
| THOR                          | 0.96        | 0.15       | 0.999             | 0.991            |
| ABDO                          | 0.966       | 0.227      | 0.999             | 0.992            |
| Señales apiladas              | 0.973       | 0.237      | 0.999             | 0.993            |
| Señales independientes        | 0.932       | 0.176      | 0.999             | 0.988            |
| Rand init                     | 0.95        | 0.187      | 0.999             | 0.989            |
| Ajuste learning rate          | 0.995       | 0.188      | 1                 | 0.99             |

Tabla 7.19: Resultados clasificación modelo ResNet 18 con exploración de grano grueso de diferentes hiperparámetros (eventos centrales)

### 7.3. Problema de regresión junto con una regresión lineal final

Dado el moderado rendimiento obtenido en clasificación, se ha decidido continuar con el enfoque inicial del problema como un problema de regresión, pero añadiendo al final una regresión lineal ajustada durante el entrenamiento del modelo para tratar de reducir posibles sobreestimaciones o subestimaciones de las redes neuronales. En las tablas 7.20 y 7.21 se muestran los resultados obtenidos para la exploración de diferentes arquitecturas, para la estimación del IAH y del IAC, respectivamente. Se toman como referencia los modelos entrenados en la primera variante del problema y se estudia de cada familia de arquitecturas el modelo con el que mejor resultado se obtiene. No se explora la red de diseño propio en esta variante debido al bajo rendimiento obtenido anteriormente.

Para la estimación del IAH, el mejor resultado se obtiene con el modelo EfficientNet b1, que obtiene un índice kappa muy similar al obtenido sin aplicar la regresión lineal final tras el paso de los datos por el modelo. El resto de modelos comunes con los experimentos de la primera variante también obtienen resultados similares. De las nuevas arquitecturas introducidas, el modelo ConvNext no logra converger hacia una solución, mientras que el modelo ResNetRS obtiene el segundo mejor resultado, muy cerca del obtenido por el modelo EfficientNet b1.

Para la estimación del IAC, el mejor resultado se obtiene con el modelo ResNetRS. Además, los modelos comunes con la variante sin regresión lineal final obtienen por lo general mejores resultados, por lo que podemos decir que para el caso de la estimación del IAC, el uso de esta regresión lineal sí que afecta de forma positiva al rendimiento de los modelos. El modelo ConvNext, al igual que ocurre con la estimación del IAH, no logra capturar la información presente en los bispectrums ni converger hacia una solución.



| Modelo           | Kappa train | Kappa val. | Accuracy train | Accuracy val. |
|------------------|-------------|------------|----------------|---------------|
| Resnet 18        | 0.945       | 0.294      | 0.964          | 0.539         |
| Double Resnet 18 | 0.93        | 0.302      | 0.954          | 0.539         |
| ResNetRS         | 0.786       | 0.312      | 0.86           | 0.55          |
| ConvNext         | 0           | 0          | 0.368          | 0.435         |
| EfficientNet b1  | 0.736       | 0.318      | 0.827          | 0.547         |
| DenseNet 121     | 0.756       | 0.307      | 0.84           | 0.539         |

Tabla 7.20: Resultados regresión junto con una regresión lineal (todos los eventos apneicos)

| Modelo           | Kappa train | Kappa val. | Accuracy train | Accuracy val. |
|------------------|-------------|------------|----------------|---------------|
| Resnet 18        | 0.923       | 0.543      | 0.997          | 0.992         |
| Double Resnet 18 | 0.914       | 0.305      | 0.996          | 0.983         |
| ResNetRS         | 0.483       | 0.557      | 0.983          | 0.989         |
| ConvNext         | 0           | 0          | 0.976          | 0.988         |
| EfficientNet b1  | 0.459       | 0.397      | 0.982          | 0.991         |
| DenseNet 121     | 0.447       | 0.495      | 0.982          | 0.989         |

Tabla 7.21: Resultados regresión junto con una regresión lineal (eventos centrales)

Para el problema de regresión acompañado de una regresión lineal, se ha tratado de encontrar algún hiperparámetro que mejore el rendimiento diagnóstico. Para estos experimentos, al igual que para la variante del problema de clasificación, se ha empleado el modelo ResNet 18 por su gran popularidad, amplio uso y buenos resultados en el estado del arte de *transfer learning*. En las tablas 7.22 y 7.23 se muestran los resultados obtenidos para la estimación del IAH y del IAC, respectivamente. En concreto, se han explorado diferentes funciones de *loss* y un tamaño del *batch* más reducido.

| Hiperparámetros ResNet18 | Kappa train | Kappa val. | Accuracy train | Accuracy val. |
|--------------------------|-------------|------------|----------------|---------------|
| Batch 4                  | 0.966       | 0.275      | 0.978          | 0.527         |
| L1 Loss                  | 0.946       | 0.308      | 0.964          | 0.544         |
| MSE Loss                 | 0.91        | 0.303      | 0.941          | 0.541         |
| Smooth L1 Loss           | 0.936       | 0.299      | 0.957          | 0.533         |

Tabla 7.22: Resultados regresión junto con una regresión lineal (todos los eventos apneicos)

El uso de un tamaño de *batch* más pequeño, tanto para la estimación del IAH como del IAC, no mejora los resultados obtenidos con un tamaño de *batch* más grandes. El uso de una función de *loss* diferente, para el caso de la estimación del IAH sí mejora, aunque ligeramente, los resultados obtenidos con la función de pérdida de Huber empleada para

la mayoría de los experimentos. En cambio, no se observaron mejoras al utilizar otras funciones de pérdida para estimar el IAC.

| Hiperparámetros<br>ResNet18 | Kappa train | Kappa val. | Accuracy<br>train | Accuracy<br>val. |
|-----------------------------|-------------|------------|-------------------|------------------|
| Batch 4                     | 0.983       | 0.269      | 0.999             | 0.98             |
| L1 Loss                     | 0           | 0          | 0.976             | 0.988            |
| MSE Loss                    | 0.896       | 0.464      | 0.995             | 0.986            |
| Smooth L1 Loss              | 0.883       | 0.38       | 0.995             | 0.988            |

Tabla 7.23: Resultados regresión junto con una regresión lineal (eventos centrales)

## 7.4. Problema de regresión con nuevo mapa de color

Dados los resultados que se han obtenido previamente, se han creado nuevos experimentos que traten de captar mejor los eventos presentes en los bispectrums. Para ello, se han representado con un nuevo mapa de color que abarca una gama más amplia de colores, para resaltar con mayor intensidad los eventos en los bispectrums. Se han revisado las mismas arquitecturas que las empleadas en la versión anterior del problema, a excepción de la arquitectura ConvNext, que como ya se vio, no logró capturar la información presente en los bispectrums. En las tablas 7.24 y 7.25 se muestran los resultados de estos experimentos para la estimación del IAH y del IAC, respectivamente.

Para la estimación del IAH, el mejor resultado se ha obtenido con el modelo Double Resnet 18, que obtiene un resultado ligeramente mejor que el modelo con el mapa de colores inicial. En cambio, para la mayoría de los experimentos con el nuevo mapa de color, se obtienen resultados ligeramente inferiores que con el mapa de color antiguo, por lo que podríamos decir que la estrategia de remarcar los eventos no está mejorando el aprendizaje de las redes convolucionales estudiadas.

| Modelo           | Kappa train | Kappa val. | Accuracy<br>train | Accuracy<br>val. |
|------------------|-------------|------------|-------------------|------------------|
| Resnet 18        | 0.95        | 0.287      | 0.967             | 0.551            |
| Double Resnet 18 | 0.759       | 0.305      | 0.843             | 0.547            |
| ResNetRS         | 0.959       | 0.267      | 0.973             | 0.524            |
| EfficientNet b1  | 0.881       | 0.282      | 0.922             | 0.53             |
| DenseNet 121     | 0.935       | 0.293      | 0.957             | 0.547            |

Tabla 7.24: Resultados regresión con nuevo mapa de color (todos los eventos apneicos)

Para la estimación del IAC, el mejor resultado se obtiene para el modelo DenseNet 121, aunque este resultado es ligeramente inferior al obtenido con el mapa de color original. El resto de los experimentos realizados con el nuevo mapa de color, tampoco mejora los resultados obtenidos previamente. De esta forma, al igual que se vio para la estimación

del IAH, el uso de este nuevo mapa de colores no facilita a los modelos el aprendizaje de patrones en los bispectrums para la estimación de la severidad de la apnea del sueño.

| Modelo           | Kappa train | Kappa val. | Accuracy train | Accuracy val. |
|------------------|-------------|------------|----------------|---------------|
| Resnet 18        | 0.976       | 0.345      | 0.999          | 0.983         |
| Double Resnet 18 | 0.820       | 0.309      | 0.992          | 0.974         |
| ResNetRS         | 0.984       | 0.360      | 0.999          | 0.989         |
| EfficientNet b1  | 0.933       | 0.328      | 0.997          | 0.988         |
| DenseNet 121     | 0.968       | 0.397      | 0.998          | 0.991         |

Tabla 7.25: Resultados regresión con nuevo mapa de color (eventos centrales)

## 7.5. Problema de regresión con señal completa

En esta última variante se emplea toda la noche de sueño de un paciente para la estimación del IAH y del IAC, dado que la señal completa proporciona una visión global de los eventos y de sus interacciones, que pueden no ser evidentes en los fragmentos de 20 minutos. En las tablas 7.26 y 7.28 se muestran los resultados de los experimentos con diferentes arquitecturas de redes convolucionales para la estimación del IAH y del IAC. Además, en las tablas 7.27 y 7.29 se muestran los resultados de varios experimentos en los que se modifica el tamaño del *batch*, la función de *loss* y la forma de introducir los datos en el modelo, tanto para la estimación del IAH como del IAC.

| Modelo           | Kappa train | Kappa val. | Accuracy train | Accuracy val. |
|------------------|-------------|------------|----------------|---------------|
| Resnet 18        | 0.992       | 0.211      | 0.995          | 0.496         |
| Double ResNet 18 | 0.986       | 0.227      | 0.991          | 0.499         |
| ResNet 50        | 0.983       | 0.207      | 0.989          | 0.507         |
| EfficientNet b1  | 0.988       | 0.171      | 0.992          | 0.476         |
| DenseNet 121     | 0.988       | 0.173      | 0.992          | 0.469         |
| ResNetRS         | 0.989       | 0.230      | 0.993          | 0.513         |
| Red propia 4c    | 0.79        | 0.129      | 0.864          | 0.455         |
| Red propia 6c    | 0.86        | 0.141      | 0.909          | 0.447         |

Tabla 7.26: Resultados regresión con señal completa (todos los eventos apneicos)

Para la estimación del IAH, el experimento que mejor índice kappa obtiene es el modelo ResNetRS, aunque se debe destacar que los resultados obtenidos en esta variante del problema no incrementan el rendimiento alcanzado en las variantes anteriores, en las que se estima el IAH a partir de fragmentos de 20 minutos en lugar de la señal completa. Se puede apreciar también que los modelos pequeños con una arquitectura de diseño propio, obtienen

unos resultados que no distan mucho de modelos mucho más grande como EfficientNet b1, lo que sugiere que con un diseño adecuado, es posible lograr un rendimiento competitivo sin necesidad de recurrir a arquitecturas de gran tamaño y complejidad. Por otro lado, el uso de un tamaño de *batch* 4 y los bispectrums apilados como datos de entrada para el modelo ResNet18 mejoran ligeramente el índice kappa obtenido en validación.

| Variantes modelo<br>ResNet 18 | Kappa train | Kappa val. | Accuracy<br>train | Accuracy<br>val. |
|-------------------------------|-------------|------------|-------------------|------------------|
| Batch 4                       | 0.985       | 0.221      | 0.99              | 0.513            |
| MSE loss                      | 0.98        | 0.187      | 0.987             | 0.482            |
| Señales apiladas              | 0.992       | 0.215      | 0.995             | 0.521            |

Tabla 7.27: Resultados regresión con señal completa con exploración de grano grueso de diferentes hiperparámetros (todos los eventos apneicos)

Para la estimación del IAC, el experimento que mejores resultados obtiene es de nuevo el que emplea el modelo ResNetRS, aunque estos resultados tampoco superan los obtenidos previamente en versiones anteriores del problema, tanto para el IAH como para el IAC. Como ocurre con la estimación del IAH, el uso de modelos de grandes dimensiones, en comparación con los modelos más pequeños de diseño propio, no parece que proporcione resultados muy superiores, además de evidenciar un claro sobreajuste. La exploración de los hiperparámetros para el modelo ResNet 18 nos proporciona ligeras mejoras respecto al modelo básico.

| Modelo           | Kappa train | Kappa val. | Accuracy<br>train | Accuracy<br>val. |
|------------------|-------------|------------|-------------------|------------------|
| Resnet 18        | 0.989       | 0.263      | 0.999             | 0.983            |
| Double ResNet 18 | 0.983       | 0.255      | 0.999             | 0.978            |
| ResNet 50        | 0.994       | 0.263      | 1                 | 0.983            |
| EfficientNet b1  | 0.983       | 0.141      | 0.999             | 0.982            |
| DenseNet 121     | 0.989       | 0.243      | 0.999             | 0.977            |
| ResNetRS         | 0.983       | 0.347      | 0.999             | 0.983            |
| Red propia 4c    | 0.879       | 0.216      | 0.992             | 0.974            |
| Red propia 6c    | 0.937       | 0.308      | 0.996             | 0.983            |

Tabla 7.28: Resultados regresión con señal completa (eventos centrales)

## 7.6. Análisis de los mejores resultados

Una vez analizados todos los experimentos realizados para cada una de las cinco versiones estudiadas del problema, se van a analizar las funciones de *loss* de los modelos de cada versión que han obtenido un índice kappa más alto en validación, para tratar

| Variantes modelo<br>ResNet 18 | Kappa train | Kappa val. | Accuracy<br>train | Accuracy<br>val. |
|-------------------------------|-------------|------------|-------------------|------------------|
| Batch 4                       | 0.977       | 0.326      | 0.998             | 0.982            |
| MSE loss                      | 0.966       | 0.288      | 0.998             | 0.982            |
| Señales apiladas              | 0.989       | 0.295      | 0.999             | 0.972            |

Tabla 7.29: Resultados regresión con señal completa con exploración de grano grueso de diferentes hiperparámetros (eventos centrales)

de encontrar la causa raíz del bajo rendimiento diagnóstico del problema. El resumen de los resultados de estos modelos se muestra en las tablas 7.30 y 7.31. En la figura 7.27 se muestran los gráficos con la función de *loss* para la estimación del IAH en la primera columna y del IAC en la segunda columna. En cada fila se representa la evolución de las funciones de *loss* para cada variante. Además, en color naranja se muestra la función de *loss* para el entrenamiento y en color verde, para la validación.

| Modelos                 | Kappa train | Kappa val. | Accuracy<br>train | Accuracy<br>val. |
|-------------------------|-------------|------------|-------------------|------------------|
| V1: Efficientnet b1     | 0.557       | 0.321      | 0.712             | 0.559            |
| V2: ResNet 18 umbral 15 | 0.965       | 0.401      | 0.986             | 0.742            |
| V3: EfficientNet b1     | 0.736       | 0.318      | 0.827             | 0.547            |
| V4: DoubleResNet        | 0.759       | 0.305      | 0.843             | 0.547            |
| V5: ResNet RS           | 0.989       | 0.230      | 0.993             | 0.513            |

Tabla 7.30: Resumen de resultados (todos los eventos apneicos)

| Modelos                        | Kappa train | Kappa val. | Accuracy<br>train | Accuracy<br>val. |
|--------------------------------|-------------|------------|-------------------|------------------|
| V1: ResNet 18                  | 0.641       | 0.452      | 0.987             | 0.991            |
| V2: ResNet 18 señales apiladas | 0.932       | 0.176      | 0.999             | 0.988            |
| V3: ResNet RS                  | 0.483       | 0.557      | 0.983             | 0.989            |
| V4: DenseNet 121               | 0.968       | 0.397      | 0.998             | 0.991            |
| V5: ResNet RS                  | 0.983       | 0.347      | 0.999             | 0.983            |

Tabla 7.31: Resumen de resultados (eventos centrales)

La versión del problema que mejores resultados proporciona empleando todos los eventos apneicos es la versión de clasificación de los sujetos en apnea/no apnea, mediante el modelo ResNet 18 y empleando un umbral de 15 eventos para considerar si un sujeto padece apnea o no. Por otro lado, la versión del problema que mejores resultados proporciona empleando únicamente los eventos apneicos centrales es la versión de regresión que emplea una regresión lineal tras el entrenamiento del modelo ResNet RS. Analizando las tablas

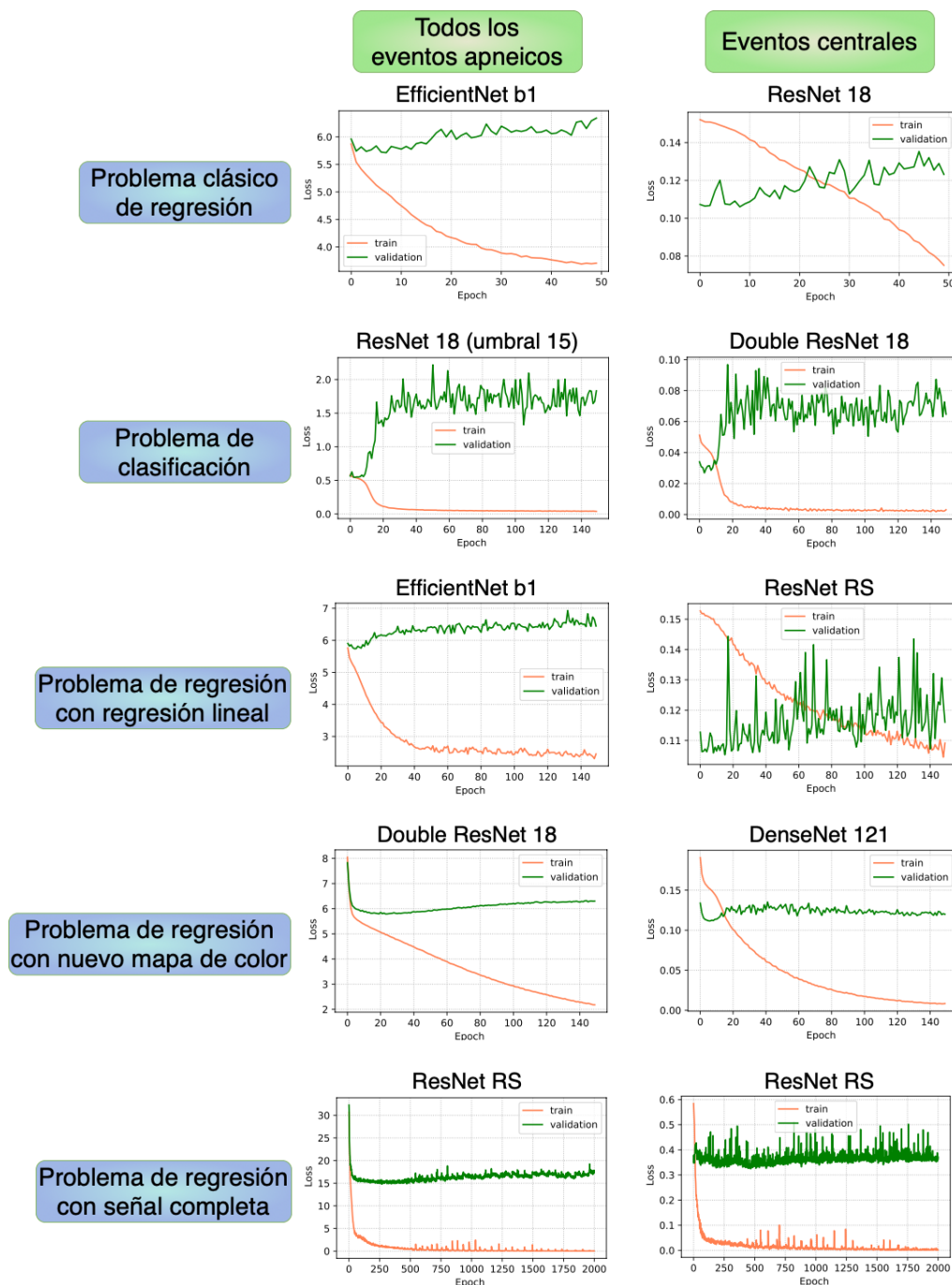


Figura 7.27: Evolución del *loss* para los mejores modelos de cada variante

7.30 y 7.31 se puede ver que existe un claro sobreajuste en el entrenamiento del primer modelo, aunque esta limitación resulta generalizada para el entrenamiento de la mayoría de los experimentos realizados en este estudio.

Analizando las funciones de *loss* de los modelos con un mayor índice kappa en validación se puede apreciar un claro sobreajuste. Así como aparentemente, la función de *loss* para el

entrenamiento evidencia que sí existe un progreso y un aprendizaje, no ocurre lo mismo para las funciones de *loss* en validación que, en la mayoría de los casos, no solo no mejora, sino que incluso empeora ligeramente a medida que pasan las épocas. Estas funciones de *loss* nos muestran un claro sobreajuste en los modelos, y que los modelos de redes convolucionales empleados no llegan a aprender los patrones presentes en los bispectrums.

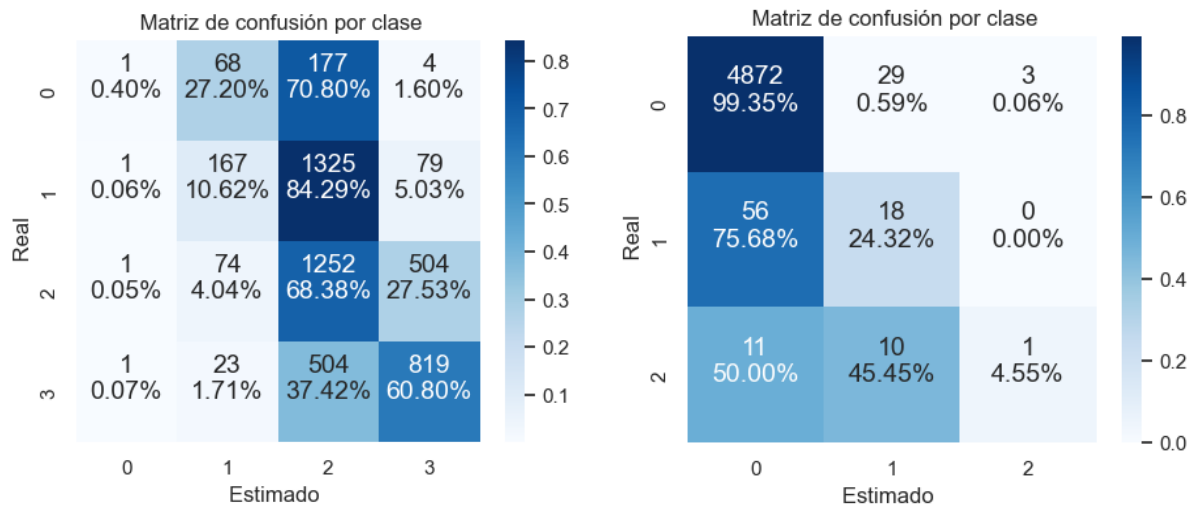
A continuación se van a analizar los resultados sobre test de los dos modelos que han obtenido un índice kappa mayor en validación. Para la estimación de IAH se va a seleccionar el modelo EfficientNet b1 de la versión del problema de regresión clásica, en lugar del modelo ResNet 18 de la versión del problema de clasificación, puesto que esta versión se empleó como simplificación del problema en busca de la causa raíz del bajo rendimiento diagnóstico de los modelos empleados. Además, los resultados de los experimentos de esta versión no son comparables, puesto que estiman la presencia o ausencia de eventos apneicos en fragmentos de 20 minutos, en lugar del IAH o del IAC. Para la estimación de IAC se ha empleado el modelo ResNet RS seguido de una regresión lineal simple. Los resultados sobre test: sensibilidad, especificidad, valor predictivo positivo (VPP), valor predictivo negativo (VPN), para los umbrales de 5 e/h, 15 e/h y 30 e/h, además del índice kappa para entrenamiento, validación y test y el *accuracy* sobre test, se muestran en la tabla 7.32. En la figura 7.28 se muestran las matrices de confusión calculadas sobre el conjunto de test para estos dos modelos, a la izquierda (figura 7.28a) para la estimación del IAH y a la derecha (figura 7.28b) para la estimación del IAC.

|                      | EfficientNet b1 |        |        | ResNetRS (central) |        |          |
|----------------------|-----------------|--------|--------|--------------------|--------|----------|
|                      | 5 e/h           | 15 e/h | 30 e/h | 5 e/h              | 15 e/h | 30 e/h * |
| Sensibilidad         | 0.004           | 0.130  | 0.839  | 0.993              | 0.999  | -        |
| Especificidad        | 0.999           | 0.969  | 0.608  | 0.303              | 0.045  | -        |
| VPP                  | 0.250           | 0.705  | 0.853  | 0.986              | 0.996  | -        |
| VPN                  | 0.950           | 0.660  | 0.583  | 0.475              | 0.25   | -        |
| Kappa E              | 0.557           |        |        | 0.483              |        |          |
| Kappa V              | 0.321           |        |        | 0.557              |        |          |
| Kappa T              | 0.169           |        |        | 0.297              |        |          |
| <i>Accuracy</i> test | 0.448           |        |        | 0.978              |        |          |

Tabla 7.32: Resultados sobre test para el mejor resultado de estimación del IAH y del IAC (\* No existen sujetos con IAC  $\geq$  30 e/h)

Como se puede apreciar, los resultados sobre el conjunto de test son moderados, especialmente para la estimación del IAH, en el que el índice kappa alcanza un valor de 0.169. Los resultados para la estimación del IAC tampoco son buenos, con un índice kappa de 0.297. Estos valores están por debajo de los obtenidos en [4], donde la estimación del IAH y del IAC empleando la señal en crudo y modelos más sencillos logró índices kappa de 0.396 y 0.590, respectivamente. Esto sugiere que, a diferencia de lo que ocurre con otras señales cardiorrespiratorias, los eventos de apnea o hipopnea no provocan cambios, o no lo suficientemente significativos en las interacciones no lineales, acoplamiento o gaussianidad de las señales THOR y ABDO. Es por esto que dichas alteraciones no están





(a) Matriz de confusión sobre test para la estimación del IAH (EfficientNet b1)

(b) Matriz de confusión sobre test para la estimación del IAC (ResNet RS)

Figura 7.28: Matrices de confusión calculadas sobre el conjunto de test

siendo capturadas en los correspondientes bispectrums de THOR y ABDO, y se puede estar omitiendo información relevante contenida en la señal en crudo, crucial para la correcta captura de los patrones subyacentes necesarios para la estimación precisa de los eventos respiratorios. Para la estimación del IAC, a pesar de obtener un índice kappa bajo, la tasa de acierto es muy elevada, de 0.978. Este resultado podría estar condicionado por el desbalanceo de los datos, de forma que el modelo tiende a clasificar a casi todos los sujetos hacia la clase mayoritaria, logrando una alta precisión sin realmente ser capaz de distinguir los diferentes umbrales de severidad del IAC.

Analizando la sensibilidad sobre el conjunto de datos de test, se puede ver cómo en el caso del modelo EfficientNet b1, tiene un valor moderado en el umbral de 5 e/h, que va mejorando a medida que el umbral aumenta. Por otro lado, la sensibilidad para la estimación del IAC con el modelo ResNet RS es alto para todos los umbrales. Si nos fijamos en la especificidad, para la estimación del IAH podemos ver que es muy alta, aunque se reduce para el umbral de severidad más alto (30 e/h). En el caso del modelo ResNet RS, la especificidad es baja para todos los umbrales. Para el VPP y el VPN se aprecia el mismo patrón de acierto que para la sensibilidad y especificidad, respectivamente, para ambos modelos. El modelo EfficientNet b1 muestra un aumento significativo en el VPP y una disminución del VPN a medida que crece el umbral de eventos, mientras que el ResNet RS mantiene un VPP alto, a costa de un VPN bajo. Es importante destacar que no existen resultados para el umbral de más de 30 e/h para la estimación del IAC, puesto que no existe ningún individuo que presente tantos eventos apneicos centrales.

En la matriz de confusión para la estimación del IAH utilizando el modelo EfficientNet b1, se aprecia una clara tendencia a sobrestimar la severidad de los casos con menor gravedad. Esto se puede apreciar en que muchos de los casos de no apnea (clase 0) y apnea



leve (clase 1) son clasificados incorrectamente en clases de mayor severidad, particularmente en la clase de apnea moderada (clase 2). Además, la matriz de confusión muestra que la mayoría de los sujetos tienden a ser clasificados en el grupo de severidad moderada (clase 2), que se corresponde con la clase mayoritaria. Este sesgo hace que el modelo tenga una baja especificidad y sensibilidad para las otras clases, como se ha visto en la tabla 7.32.

En la matriz de confusión para la estimación del IAC con el modelo ResNet RS se observa un patrón similar. Existe una gran tendencia a clasificar los sujetos en la clase sin presencia de apnea (clase 0), que a su vez es la clase mayoritaria de los sujetos con apnea central. Este sesgo hacia la clase mayoritaria resulta en una alta tasa de acierto, pero en un bajo índice kappa y una baja especificidad.

Para analizar cómo los modelos empleados obtienen sus predicciones, se han utilizado dos técnicas de XAI: GradCam y ablaciones sobre los datos de entrada. En el caso del modelo EfficientNet b1, se evaluó a un sujeto con IAH severo y cuya clasificación fue correcta. Las figuras 7.29 y 7.30 muestran la aplicación de GradCam y ablaciones sobre cuatro fragmentos de 20 minutos cada uno, correspondientes a este sujeto. Para el modelo ResNet RS se ha analizado un sujeto con IAC moderado (dado que ningún individuo tiene clasificación severa con eventos centrales) correctamente clasificado. Las figuras 7.31 y 7.32 muestran otros cuatro fragmentos de 20 minutos de este sujeto, aplicando GradCam y ablaciones respectivamente.

En el caso del modelo EfficientNet b1, se puede apreciar a través de GradCam cómo el modelo en algunos fragmentos parece centrarse en las zonas de bajas frecuencias, donde se encuentran generalmente los eventos apnéicos, pero, por lo general, su atención está dispersa y abarca áreas más amplias de las esperadas. Esto sugiere que, como ya se ha mencionado previamente, los eventos de apnea e hipopnea no provocan cambios suficientemente significativos en las interacciones no lineales, gaussianidad o acoplamiento, por lo que no se reflejan de forma adecuada en los bispectrums, haciendo que el modelo no sea capaz de enfocarse en las regiones de bajas frecuencias en las que se producen los eventos. Por otro lado, a través de la técnica de las ablaciones se puede ver que tan solo para uno de los fragmentos, el modelo se centra en las zonas de más bajas frecuencias. Por lo general, el modelo se enfoca en zonas en las que no se representa ningún evento para la estimación del IAH, lo que de nuevo sugiere que los bispectrums no reflejan de forma adecuada la ocurrencia de eventos apnéicos e hipopneas.

Para el caso del modelo ResNet RS, GradCam muestra cómo el modelo se centra en zonas más restringidas y de bajas frecuencias para las predicciones, lo que podría justificar el hecho de tener resultados ligeramente mejores para la estimación del IAC. A través de las ablaciones vemos que el modelo no se centra en este caso en ninguna de las celdas de más bajas frecuencias, pero sí en celdas adyacentes en todos los casos, lo que podría justificar también que los resultados obtenidos para la estimación del IAC sean ligeramente mejores que los obtenidos para la estimación del IAH.

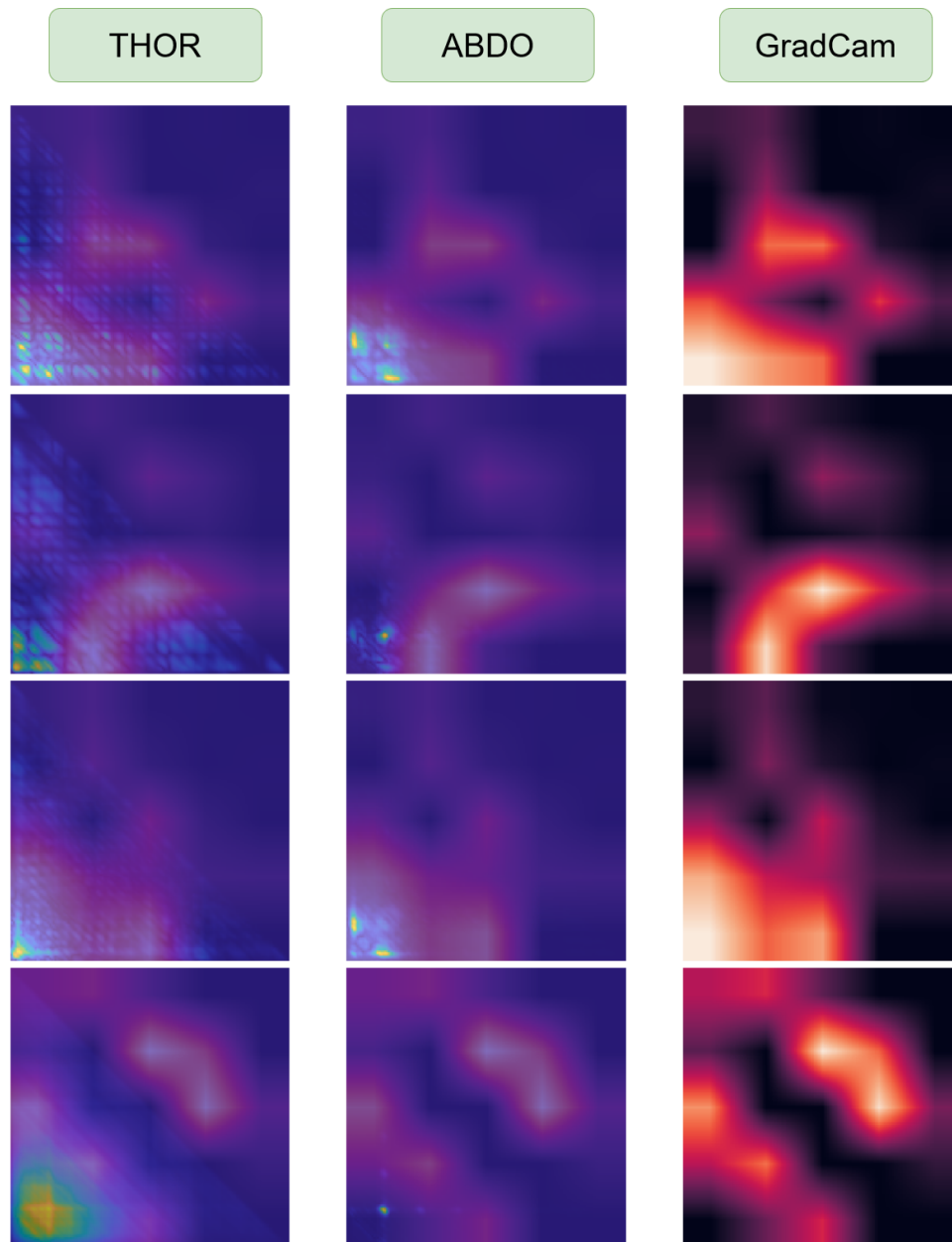


Figura 7.29: Gradcam sobre un sujeto con apnea severa y una predicción correcta de severidad del [IAH](#) con modelo EfficientNet b1

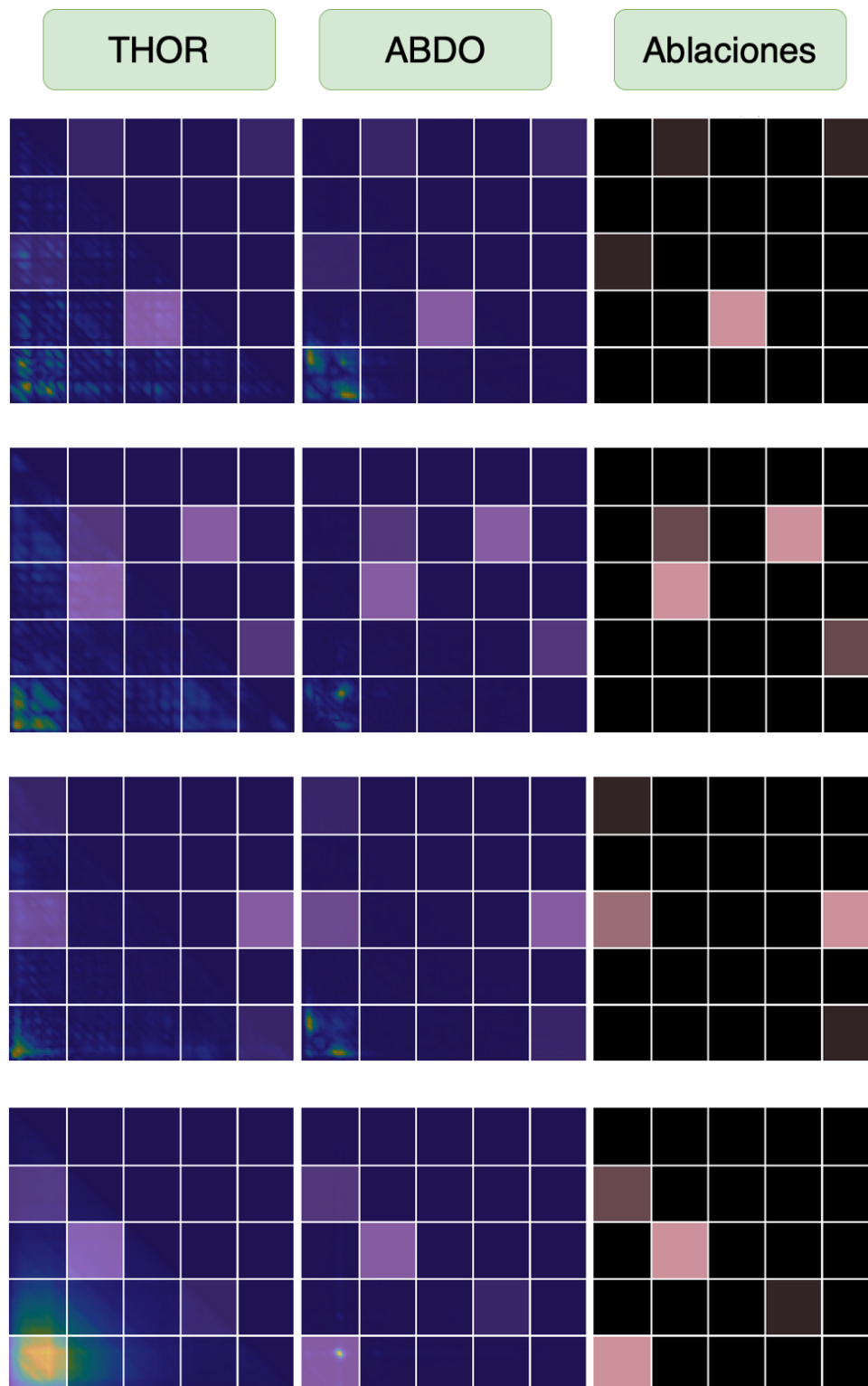


Figura 7.30: Ablaciones sobre un sujeto con apnea severa y una predicción correcta de severidad del IAHI con modelo EfficientNet b1

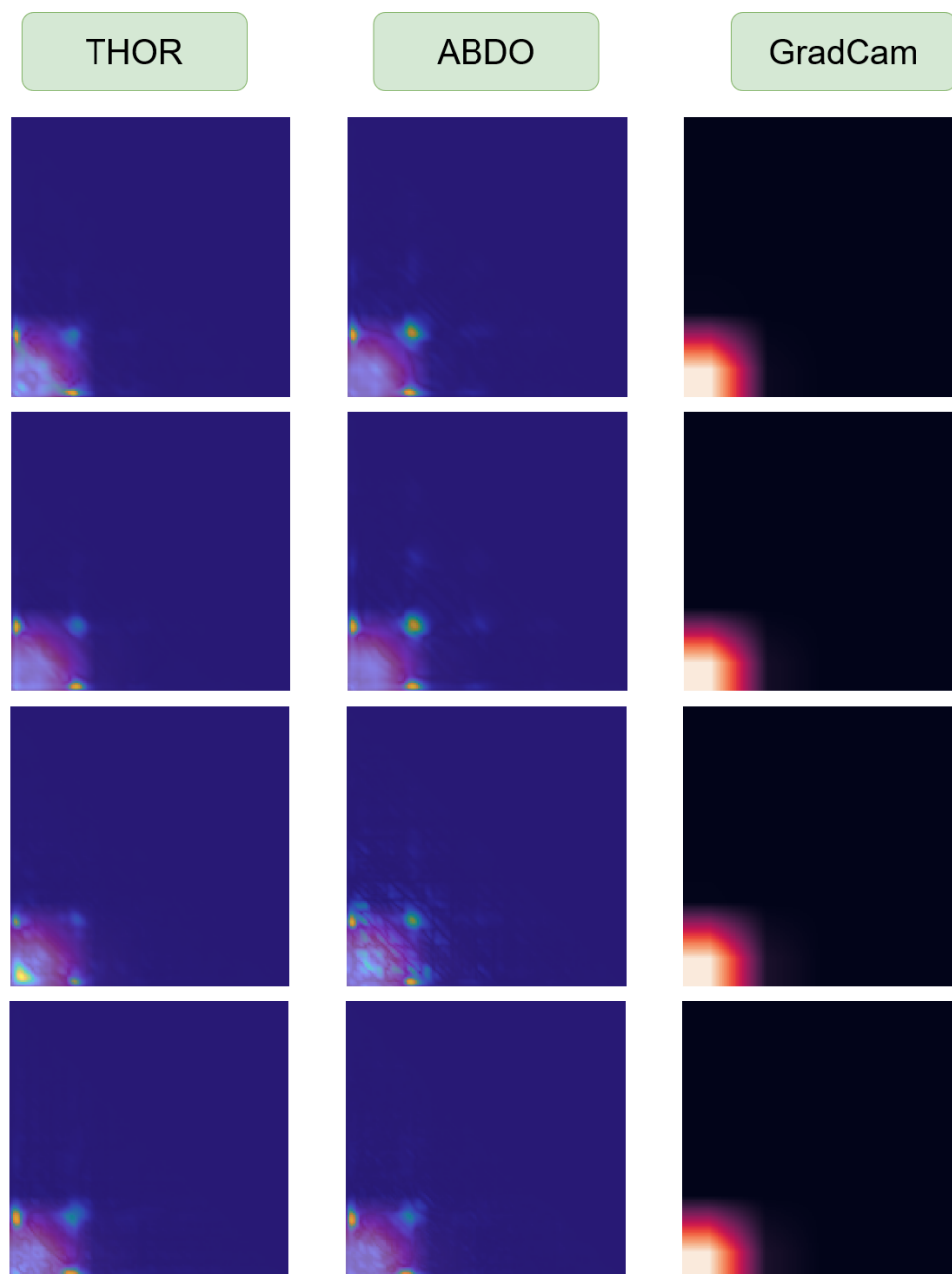


Figura 7.31: Gradcam sobre un sujeto con apnea severa y una predicción correcta de severidad del IAC con modelo ResNet RS

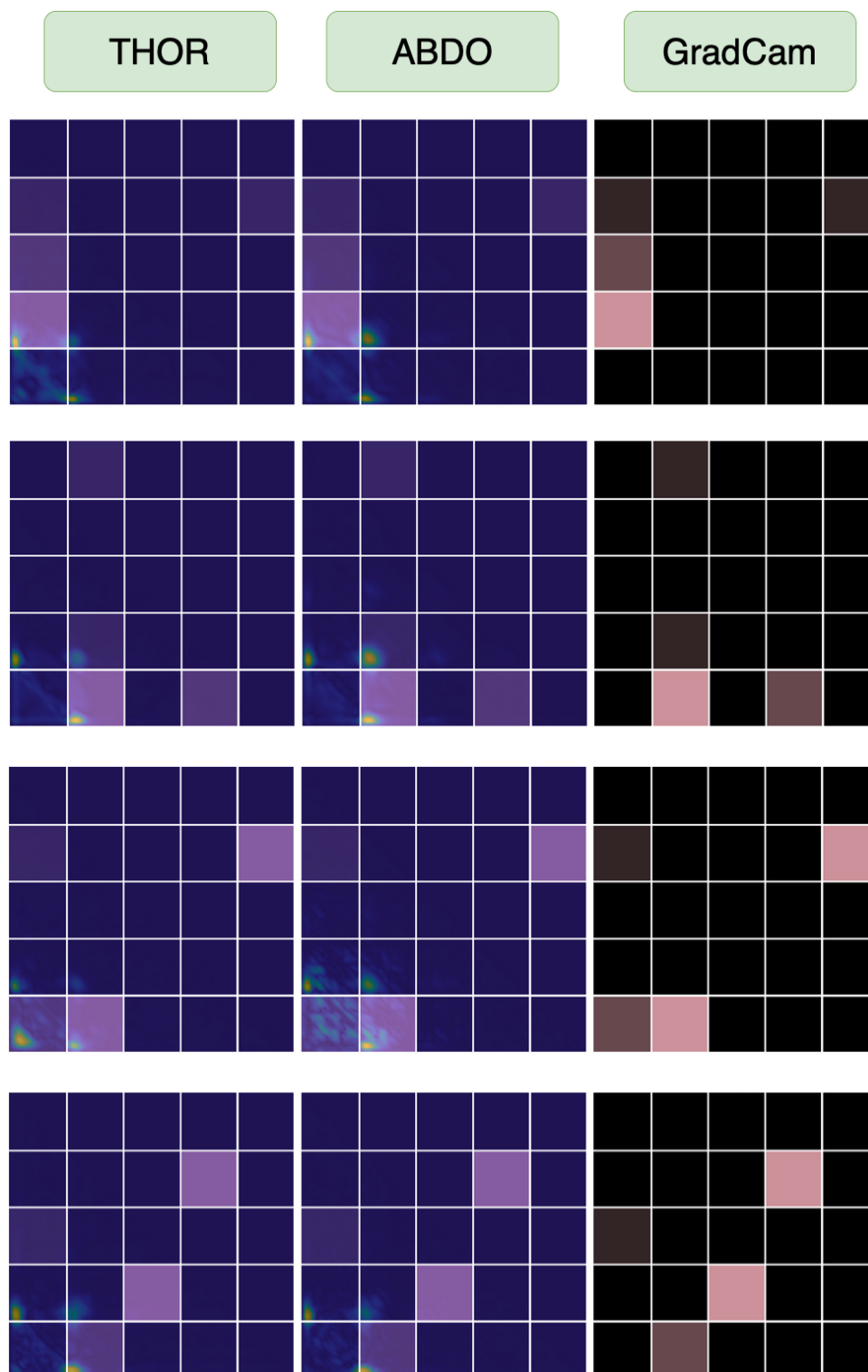


Figura 7.32: Ablaciones sobre un sujeto con apnea severa y una predicción correcta de severidad IAC con modelo ResNet RS

---

## 8: Conclusiones y líneas de trabajo futuras

---

En este capítulo se van a presentar las conclusiones extraídas tras el desarrollo del proyecto descrito en el presente documento, así como las limitaciones presentes en el estudio y las posibles futuras líneas de investigación.

### 8.1. Conclusiones

En este trabajo se han analizado diferentes modelos de redes neuronales convolucionales aplicados a la estimación de la severidad de la apnea del sueño de pacientes a partir de señales de esfuerzo respiratorio torácico y abdominal representadas en forma de bispectrums. Las señales empleadas en este proyecto han sido extraídas de la base de datos [SHHS](#), y fueron recogidas mediante bandas de pletismografía inductiva situadas en el tórax y en el abdomen de los pacientes durante una noche de sueño en sus domicilios.

Este problema de estimación de la severidad de la apnea del sueño, tanto para todos los tipos de eventos apneicos, como para los eventos centrales en particular, ha sido analizado desde cinco perspectivas diferentes, en las que se ha modificado el problema básico de regresión, como problema de clasificación, añadiendo una regresión lineal final para evitar sesgos en los resultados, variando el mapa de color empleado para la representación de los bispectrums y representando las señales completas en un único bispectrum en lugar de fragmentos de 20 minutos. Los resultados obtenidos en estas variantes no superan los resultados obtenidos en etapas previas del estudio [4] en las que se analizaron las señales en crudo en lugar de su representación en forma de bispectrum. Mientras que para la estimación del [IAH](#) se obtuvo un índice kappa sobre el conjunto de test de 0.396 y una tasa de acierto del 58.8 % empleando la señal en crudo, el modelo que ha proporcionado un mejor resultado empleando bispectrums, logra tan solo un índice kappa sobre test de 0.169 y una tasa de acierto de 44.80 %. Lo mismo ocurre para la estimación del [IAC](#), mientras que el uso de las señales en crudo obtiene un índice kappa de 0.590 y una tasa de acierto del 98.50 %, el uso de bispectrums logra un índice kappa de tan solo 0.297 y una tasa de acierto del 97.80 %.

El uso de [XAI](#), tanto la técnica de Grad Cam, como la aplicación de ablaciones sobre los datos de entrada han permitido ver cómo los modelos, por lo general, no se fijan en las regiones de bajas frecuencias, que son características de la respiración normal y la ocurrencia de eventos apneicos en los bispectrums. En cambio, en ocasiones se enfocan en regiones de frecuencias más altas, donde no se produce ningún evento relevante. Esto sugiere que los modelos no están siendo capaces de aprender los patrones existentes en las señales, y que pueden estar aprendiendo los patrones característicos de otras enfermedades del sueño o se estén viendo influenciados por la presencia de ruido en las señales.

Como conclusión, el estudio realizado muestra que la representación de las señales en forma de bispectrums no mejora el rendimiento diagnóstico de los modelos de redes convolucionales para la estimación de la severidad de la apnea del sueño. En primer lugar, se ha observado que la cantidad de ruido presente en las señales es alta. Este ruido incrementa su influencia negativa en las predicciones de los modelos al representar las señales en forma de bispectrums puesto que confunden a los modelos de redes convolucionales. A diferencia de lo que ocurre con otras señales cardiorrespiratorias, los eventos de apnea e hipopnea no provocan cambios significativos en las interacciones no lineales, el acoplamiento y la gaussianidad de las señales [THOR](#) y [ABDO](#). Por tanto, dichas alteraciones no pueden ser capturadas de manera efectiva en los bispectrums correspondientes de estas señales de esfuerzo respiratorio. El efecto negativo del uso de bispectrums es particularmente significativo en la estimación del [IAC](#). La reducción del rendimiento diagnóstico de los modelos es más notoria en el caso de la estimación del [IAC](#), lo que sugiere que el uso de bispectrums reduce el potencial de las señales de esfuerzo respiratorio torácico y abdominal para la distinción de los eventos apneicos centrales de los eventos apneicos obstructivos.

## 8.2. Limitaciones

El estudio realizado no está exento de limitaciones. En primer lugar, existe un gran desequilibrio entre los diferentes grupos de severidad en la base de datos del [SHHS](#). Mientras que el grupo de severidad más frecuente es el grupo de apnea moderada, con un 36.37 % de los individuos, el grupo de sujetos sin apnea representa tan solo un 4.09 % del total. En el caso de la estimación del [IAC](#), este desbalanceo es incluso más notorio, siendo el grupo de sujetos sin eventos apneicos, alrededor del 98 % del total de sujetos. Este desbalanceo de los datos hace que los entrenamientos de los modelos estén sesgados y las predicciones tiendan a la clase mayoritaria, reduciendo el rendimiento en la clasificación de las clases minoritarias.

Otra de las limitaciones del estudio es que todos los individuos analizados eran mayores de 40 años. Esto significa que los resultados obtenidos en este proyecto podrían no ser aplicables a otros grupos de población, como los niños, en los que la apnea del sueño es una enfermedad con una alta prevalencia y en los que se ha visto que los eventos apneicos causan alteraciones bispectrales en [ECG](#), saturación de oxígeno en sangre y [FA](#). Los resultados obtenidos deberían estudiarse en sujetos pediátricos, particularmente las

señales de [THOR](#) y [ABDO](#), para evaluar si las conclusiones extraídas son igualmente válidas para este grupo poblacional.

Por último, otra de las limitaciones del estudio son las inconsistencias encontradas en la base de datos del [SHHS](#). Por un lado, se han encontrado inconsistencias entre las anotaciones de los eventos respiratorios y los índices clínicos derivados de los estudios del sueño. Existe una falta de concordancia entre las anotaciones de los eventos respiratorios y los índices clínicos derivados de los estudios del sueño, que podrían deberse a la corrupción de los datos a lo largo del tiempo, así como a cambios en las recomendaciones clínicas y a los criterios de puntuación en los que se basó el estudio inicial del [SHHS](#) [4] [35]. Por este motivo, para la estimación de la severidad de la apnea del sueño se han tomado como referencia las anotaciones de los eventos directamente proporcionadas para cada sujeto en las señales de esfuerzo respiratorio. Por otro lado, las señales presentan gran cantidad de ruido, que afecta a la precisión de las estimaciones de la severidad de la apnea el sueño y empeora el rendimiento de los modelos. En este estudio, se ha decidido no aplicar técnicas específicas de reducción de ruido para las señales, dado que se han analizado con modelos de *deep learning*, diseñados para aprender patrones intrínsecos en los datos. Con este enfoque se ha podido ver el gran impacto directo del ruido y otras patologías en las predicciones al representar las señales en forma de bispectrums.

### 8.3. Líneas futuras

Como se ha comentado en las limitaciones del proyecto, los datos están desbalanceados, especialmente en el caso de la estimación del [IAC](#). Como líneas futuras de investigación se podría mitigar este desbalanceo aplicando técnicas de *data augmentation* e incrementado el número de señales de sujetos de los grupos de severidad minoritarios. Además, se podrían utilizar los modelos desarrollados para otras bases de datos independientes, incluyendo conjuntos de datos con sujetos pertenecientes a diferentes grupos poblacionales, como sujetos pediátricos, de forma que se pueda justificar que los resultados obtenidos son generalizables a otros estudios.

Respecto a los resultados moderados que se han obtenido, se pueden seguir dos enfoques de investigación futura. En primer lugar, se puede tratar de realizar un filtrado exhaustivo de las señales, eliminando aquellas con artefactos provocados por una colocación incorrecta de los sensores sobre el cuerpo del paciente o los provocados por los movimientos del sujeto durante el sueño. Además, se podría diseñar e implementar un algoritmo específico para la eliminación de artefactos en estas señales. Por otro lado, se puede enfocar el problema de la estimación de la severidad de la apnea del sueño empleando la señal en crudo y otros modelos de *deep learning* como *transformers*, [RNN](#) o una combinación con redes [CNN](#). Con el uso de diferentes arquitecturas se pretende mejorar la precisión de los modelos en la estimación del [IAH](#) y del [IAC](#), identificando nuevos patrones en las señales, no detectados por las redes [CNN](#) desarrolladas en este estudio. Por último, también se pueden aplicar otras técnicas de [XAI](#) para visualizar con más exactitud y precisión las regiones en las que se fija el modelo, como LIME, SHAP o variantes como WindowSHAP.



---

## Bibliografía

---

- [1] Pytorch. <https://pytorch.org>. Accedido: 28-05-2024.
- [2] AL-ANGARI, H. M., AND SAHAKIAN, A. V. Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier. *IEEE Transactions on Information Technology in Biomedicine* 16, 3 (2012), 463–468.
- [3] BADR, M. S., DINGELL, J. D., AND JAVAHERI, S. Central sleep apnea: a brief review. *Curr Pulmonol Rep* 8, 1 (2019), 14–21.
- [4] BARROSO-GARCÍA, V., FERNÁNDEZ-POYATOS, M., SAHELICES, B., ÁLVAREZ, D., GOZAL, D., HORNERO, R., AND GUTIÉRREZ-TOBAL, G. C. Prediction of the sleep apnea severity using 2d-convolutional neural networks and respiratory effort signals. *Diagnostics* 13, 20 (2023).
- [5] BARROSO-GARCÍA, V., GUTIÉRREZ-TOBAL, G. C., KHEIRANDISH-GOZAL, L., VAQUERIZO-VILLAR, F., ÁLVAREZ, D., DEL CAMPO, F., GOZAL, D., AND HORNERO, R. Bispectral analysis of overnight airflow to improve the pediatric sleep apnea diagnosis. *Computers in Biology and Medicine* 129 (2021), 104167.
- [6] BARROSO-GARCÍA, V., GUTIÉRREZ-TOBAL, G. C., KHEIRANDISH-GOZAL, L., VAQUERIZO-VILLAR, F., ÁLVAREZ, D., DEL CAMPO, F., GOZAL, D., AND HORNERO, R. Bispectral analysis of overnight airflow to improve the pediatric sleep apnea diagnosis. *Computers in Biology and Medicine* 129 (2021), 104167.
- [7] BELLO, I., FEDUS, W., DU, X., CUBUK, E. D., SRINIVAS, A., LIN, T.-Y., SHLENS, J., AND ZOPH, B. Revisiting resnets: Improved training and scaling strategies, 2021.
- [8] BENNETT, J. A., AND KINNEAR, W. Sleep on the cheap: The role of overnight oximetry in the diagnosis of sleep apnoea hypopnoea syndrome. *Thorax* 54, 11 (1999), 958–959.
- [9] BERRY, R. B., QUAN, S. F., ABREU, A. R., BIBBS, M. L., DELROSSO, L., HARDING, S. M., MAO, M.-M., PLANTE, D. T., PRESSMAN, M. R., TROESTER,

- M. M., AND ET AL. *The AASM Manual for the scoring of sleep and associated events: Rules, terminology and technical specifications*. American Academy of Sleep Medicine, 2020.
- [10] BIXLER, E. O., VGONTZAS, A. N., HAVE, T. T., TYSON, K., AND KALES, A. Effects of age on sleep apnea in men: I. prevalence and severity. *American journal of respiratory and critical care medicine* 157, 1 (1998), 144–148.
  - [11] BROUILLETTE, R. T., MORIELLI, A., LEIMANIS, A., WATERS, K. A., LUCIANO, R., AND DUCHARME, F. M. Nocturnal Pulse Oximetry as an Abbreviated Testing Modality for Pediatric Obstructive Sleep Apnea. *Pediatrics* 105, 2 (02 2000), 405–412.
  - [12] CHEN, J., WANG, C., LIN, C., HSU, M., YEH, C., AND HWANG, S. Predicting apnea-hypopnea index in patients with obstructive sleep apnea using unsegmented ecg-signal-based algorithms. *IEEEJ Transactions on Electrical and Electronic Engineering* 18, 9 (Jul 2023), 1550–1552.
  - [13] CHEN, J.-W., LIN, S.-T., WANG, C.-Y., LIN, C.-C., HSU, K.-C., YEH, C.-Y., AND HWANG, S.-H. A signal segmentation-free model for electrocardiogram-based obstructive sleep apnea severity classification. *Advanced Intelligent Systems* 5, 3 (Jan 2023).
  - [14] CHEN, J.-W., LIU, C.-M., WANG, C.-Y., LIN, C.-C., QIU, K.-Y., YEH, C.-Y., AND HWANG, S.-H. A deep neural network-based model for osa severity classification using unsegmented peripheral oxygen saturation signals. *Engineering Applications of Artificial Intelligence* 122 (2023), 106161.
  - [15] CHUA, K. C., CHANDRAN, V., ACHARYA, U. R., AND LIM, C. M. Application of higher order statistics/spectra in biomedical signals—a review. *Medical Engineering & Physics* 32, 7 (2010), 679–689.
  - [16] COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 1 (1960), 37–46.
  - [17] DE ALMEIDA, F. R., AYAS, N. T., OTSUKA, R., UEDA, H., HAMILTON, P., RYAN, F. C., AND LOWE, A. A. Nasal pressure recordings to detect obstructive sleep apnea. *Sleep and Breathing* 10 (2006), 62–69.
  - [18] DEHLINK, E., AND TAN, H.-L. Update on paediatric obstructive sleep apnoea. *Journal of Thoracic Disease* 8, 2 (2016).
  - [19] DEVIAENE, M., TESTELMANS, D., BUYSE, B., BORZÉE, P., VAN HUFFEL, S., AND VARON, C. Automatic screening of sleep apnea patients based on the spo2 signal. *IEEE Journal of Biomedical and Health Informatics* 23, 2 (2019), 607–617.
  - [20] DONOVAN, L. M., AND KAPUR, V. K. Prevalence and characteristics of central compared to obstructive sleep apnea: Analyses from the sleep heart health study cohort. *Sleep* 39, 7 (2016), 1353–1359.

- [21] ECKERT, D. J., JORDAN, A. S., MERCHIA, P., AND MALHOTRA, A. Central sleep apnea: Pathophysiology and treatment. *Chest* 131, 2 (2007), 595–607.
- [22] FLEMONS, W. W., LITTNER, M. R., ROWLEY, J. A., GAY, P., ANDERSON, W. M., HUDGEL, D. W., MCEVOY, R. D., AND LOUBE, D. I. Home diagnosis of sleep apnea: A systematic review of the literature. *Chest* 124, 4 (2003), 1543–1579.
- [23] GARCÍA-VICENTE, C., GUTIÉRREZ-TOBAL, G. C., JIMÉNEZ-GARCÍA, J., MARTÍN-MONTERO, A., GOZAL, D., AND HORNERO, R. Ecg-based convolutional neural network in pediatric obstructive sleep apnea diagnosis. *Computers in Biology and Medicine* 167 (2023), 107628.
- [24] GOODFELLOW, I., BENGIO, Y., AND COURVILLE, A. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [25] HAIDAR, R., MCCLOSKEY, S., KOPRINSKA, I., AND JEFFRIES, B. Convolutional neural networks on multiple respiratory channels to detect hypopnea and obstructive apnea events. In *2018 International Joint Conference on Neural Networks (IJCNN)* (2018), pp. 1–7.
- [26] HAMEED, I., SHARPE, S., BARCKLOW, D., AU-YEUNG, J., VERMA, S., HUANG, J., BARR, B., AND BRUSS, C. B. Based-xai: Breaking ablation studies down for explainable artificial intelligence, 2022.
- [27] HE, K., ZHANG, X., REN, S., AND SUN, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778.
- [28] HICKS, M. W., AND FOSTER, J. S. Adapting scrum to managing a research group.
- [29] HUANG, G., LIU, Z., VAN DER MAATEN, L., AND WEINBERGER, K. Q. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 2261–2269.
- [30] HUBER, P. J. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics* 35, 1 (1964), 73 – 101.
- [31] INSTITUTE, P. M. *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*, 6th ed. Project Management Institute, Newtown Square, PA, USA, 2017.
- [32] JAVAHERI, S. Sleep disorders in systolic heart failure: a prospective study of 100 male patients. *International journal of cardiology* 106, 1 (2006), 21–28.
- [33] JIMÉNEZ-GARCÍA, J., GARCÍA, M., GUTIÉRREZ-TOBAL, G. C., KHEIRANDISH-GOZAL, L., VAQUERIZO-VILLAR, F., ÁLVAREZ, D., DEL CAMPO, F., GOZAL, D., AND HORNERO, R. A 2d convolutional neural network to detect sleep apnea in children using airflow and oximetry. *Computers in Biology and Medicine* 147 (2022), 105784.

- [34] JIMÉNEZ-GARCÍA, J., GARCÍA, M., GUTIÉRREZ-TOBAL, G. C., KHEIRANDISH-GOZAL, L., VAQUERIZO-VILLAR, F., ÁLVAREZ, D., DEL CAMPO, F., GOZAL, D., AND HORNERO, R. An explainable deep-learning architecture for pediatric sleep apnea identification from overnight airflow and oximetry signals. *Biomedical Signal Processing and Control* 87 (2024), 105490.
- [35] KARHU, T., MYLLYMAA, S., NIKKONEN, S., MAZZOTTI, D. R., TÖYRÄS, J., AND LEPPÄNEN, T. Longer and deeper desaturations are associated with the worsening of mild sleep apnea: The sleep heart health study. *Frontiers in Neuroscience* 15 (2021).
- [36] KHAN, M. T., AND FRANCO, R. A. Complex sleep apnea syndrome. *Sleep Disorders* 2014 (2014), 1–6.
- [37] KINGMA, D. P., AND BA, J. Adam: A method for stochastic optimization, 2014.
- [38] LE CUN, Y., JACKEL, L., BOSER, B., DENKER, J., GRAF, H., GUYON, I., HENDERSON, D., HOWARD, R., AND HUBBARD, W. Handwritten digit recognition: Applications of neural network chips and automatic learning. *IEEE Communications Magazine* 27, 11 (Nov 1989), 41–46.
- [39] LECUN, Y., BOTTOU, L., BENGIO, Y., AND HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 11 (1998), 2278–2324.
- [40] LIN, Y.-Y., WU, H.-T., HSU, C.-A., HUANG, P.-C., HUANG, Y.-H., AND LO, Y.-L. Sleep apnea detection based on thoracic and abdominal movement signals of wearable piezoelectric bands. *IEEE Journal of Biomedical and Health Informatics* 21, 6 (2017), 1533–1545.
- [41] LIU, D., ARMITSTEAD, J., BENJAFIELD, A., SHAO, S., MALHOTRA, A., CISTULLI, P. A., PEPIN, J.-L., AND WOEHRLE, H. Trajectories of emergent central sleep apnea during cpap therapy. *Chest* 152, 4 (2017), 751–760.
- [42] LIU, Z., MAO, H., WU, C.-Y., FEICHTENHOFER, C., DARRELL, T., AND XIE, S. A convnet for the 2020s, 2022.
- [43] LOMBARDI, C., CARAVITA, S., AND PARATI, G. Central sleep apnea during continuous positive airway pressure therapy in obstructive sleep apnea patients. *Journal of Thoracic Disease* 9, 11 (2017), 4152–4156.
- [44] LOPEZ-JIMENEZ, F., KUNIYOSHI, F. H. S., GAMI, A., AND SOMERS, V. K. Obstructive sleep apnea: implications for cardiac and vascular disease. *Chest* 133, 3 (2008), 793–804.
- [45] LOSHCHILOV, I., AND HUTTER, F. Decoupled weight decay regularization, 2017.
- [46] MARKUN, L. C., AND SAMPAT, A. Clinician-focused overview and developments in polysomnography. *Current Sleep Medicine Reports* 6, 4 (2020), 309–321.

- [47] MARTÍN-MONTERO, A., GUTIÉRREZ-TOBAL, G. C., GOZAL, D., BARROSO-GARCÍA, V., ÁLVAREZ, D., DEL CAMPO, F., KHEIRANDISH-GOZAL, L., AND HORNERO, R. Bispectral analysis of heart rate variability to characterize and help diagnose pediatric sleep apnea. *Entropy* 23, 8 (2021).
- [48] MARTÍN-MONTERO, A., GUTIÉRREZ-TOBAL, G. C., GOZAL, D., BARROSO-GARCÍA, V., ÁLVAREZ, D., DEL CAMPO, F., KHEIRANDISH-GOZAL, L., AND HORNERO, R. Bispectral analysis of heart rate variability to characterize and help diagnose pediatric sleep apnea. *Entropy* 23, 8 (2021).
- [49] MARTÍN-MONTERO, A., GUTIÉRREZ-TOBAL, G. C., KHEIRANDISH-GOZAL, L., JIMÉNEZ-GARCÍA, J., ÁLVAREZ, D., DEL CAMPO, F., GOZAL, D., AND HORNERO, R. Heart rate variability spectrum characteristics in children with sleep apnea. *Pediatric Research* 89, 7 (2021), 1771–1779.
- [50] MCCARTHY, J. What is artificial intelligence? <https://www-formal.stanford.edu/jmc/whatisai.pdf>, 2007. Stanford University.
- [51] MCCULLOCH, W. S., AND PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics* 5, 4 (1943), 115–133.
- [52] MERRICK, L. Randomized ablation feature importance, 2019.
- [53] MEYES, R., LU, M., DE PUISEAU, C. W., AND MEISEN, T. Ablation studies in artificial neural networks. *ArXiv abs/1901.08644* (2019).
- [54] MOSTAFA, S. S., MENDONÇA, F., G. RAVELO-GARCÍA, A., AND MORGADO-DIAS, F. A systematic review of detecting sleep apnea using deep learning. *Sensors* 19, 22 (2019).
- [55] MOUSAVI, M., KHANAL, A., AND ESTRADA, R. Ai playground: Unreal engine-based data ablation tool for deep learning. In *International Symposium on Visual Computing* (2020).
- [56] MUÑOZ-LOMBO, J. P., GARRIDO-VALENCIA, G. A., AND PACHECO, R. Frequency and Factors Associated with Obstructive Sleep Apnea in Adults. *RAMR* 20, 4 (2020), 381–391.
- [57] NAUGHTON, M. T. Pathophysiology and treatment of cheyne-stokes respiration. *Thorax* 53, 6 (1998), 514–518.
- [58] NG, A. S., CHUNG, J. W., GOHEL, M. D., YU, W. W., FAN, K. L., AND WONG, T. K. Evaluation of the performance of using mean absolute amplitude analysis of thoracic and abdominal signals for immediate indication of sleep apnoea events. *Journal of Clinical Nursing* 17, 17 (2008), 2360–2366.
- [59] NGUYEN, A., YOSINSKI, J., AND CLUNE, J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 427–436.

- [60] NIRTHIKA, R., MANIVANNAN, S., RAMANAN, A., AND WANG, R. Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. *Neural Computing and Applications* 34, 7 (2022), 5321–5347.
- [61] OLSEN, M., MIGNOT, E., JENNUM, P. J., AND SORENSEN, H. B. D. Robust, ECG-based detection of Sleep-disordered breathing in large population-based cohorts. *Sleep* 43, 5 (11 2019), zsz276.
- [62] OLSON, A. L., AND ZWILLICH, C. The obesity hypoventilation syndrome. *The American journal of medicine* 118, 9 (2005), 948–956.
- [63] PARK, J. G., RAMAR, K., AND OLSON, E. J. Updates on definition, consequences, and management of obstructive sleep apnea. *Mayo Clinic proceedings* 86, 6 (2011), 549–554.
- [64] PAVWOSKI, P., AND SHELGIKAR, A. V. Treatment options for obstructive sleep apnea. *Neurology Clinical Practice* 7, 1 (2017), 77–85.
- [65] POWELL, V. Image kernels explained visually.
- [66] PUNJABI, N. M. The epidemiology of adult obstructive sleep apnea. *Proceedings of the American Thoracic Society* 5, 2 (2008), 136–143.
- [67] QUAN, S. F., HOWARD, B. V., IBER, C., KILEY, J. P., NIETO, F. J., O’CONNOR, G. T., RAPOPORT, D. M., REDLINE, S., ROBBINS, J., SAMET, J. M., AND WAHL, P. W. The Sleep Heart Health Study: Design, Rationale, and Methods. *Sleep* 20, 12 (12 1997), 1077–1085.
- [68] RETAMALES, G., GAVIDIA, M. E., BAUSCH, B., MONTANARI, A. N., HUSCH, A., AND GONCALVES, J. Towards automatic home-based sleep apnea estimation using deep learning. *npj Digital Medicine* 7, 1 (June 2024), 144.
- [69] ROLÓN, R., LARRATEGUY, L., DI PERSIA, L., SPIES, R., AND RUFINER, H. Discriminative methods based on sparse representations of pulse oximetry signals for sleep apnea–hypopnea detection. *Biomedical Signal Processing and Control* 33 (2017), 358–367.
- [70] RUNDO, J. V., AND DOWNEY, R. Polysomnography. In *Clinical Neurophysiology: Basis and Technical Aspects*, K. H. Levin and P. Chauvel, Eds., vol. 160 of *Handbook of Clinical Neurology*. Elsevier, 2019, pp. 381–392.
- [71] SAN, T., POLAT, S., CINGI, C., ESKIIZMIR, G., OGHAN, F., AND CAKIR, B. Effects of high altitude on sleep and respiratory system and theirs adaptations. *The Scientific World Journal* 2013 (2013), 1–7.
- [72] SARANYA, A., AND SUBHASHINI, R. A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal* 7 (2023), 100230.

- [73] SELVARAJU, R. R., COGSWELL, M., DAS, A., VEDANTAM, R., PARIKH, D., AND BATRA, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), pp. 618–626.
- [74] SENARATNA, C. V., PERRET, J. L., LODGE, C. J., LOWE, A. J., CAMPBELL, B. E., MATHESON, M. C., HAMILTON, G. S., AND DHARMAGE, S. C. Prevalence of obstructive sleep apnea in the general population: A systematic review. *Sleep Medicine Reviews* 34 (2017), 70–81.
- [75] SHABBIR, J., AND ANWER, T. Artificial intelligence and its role in near future, Apr 2018.
- [76] SILVA, G. E., AN, M.-W., GOODWIN, J. L., SHAHAR, E., REDLINE, S., RESNICK, H., BALDWIN, C. M., AND QUAN, S. F. Longitudinal evaluation of sleep-disordered breathing and sleep symptoms with change in quality of life: the sleep heart health study (shhs). *Sleep* 32, 8 (2009), 1049–1057.
- [77] SIMONYAN, K., AND ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *CoRR abs/1409.1556* (2014).
- [78] SZEGEDY, C., ZAREMBA, W., SUTSKEVER, I., BRUNA, J., ERHAN, D., GOODFELLOW, I., AND FERGUS, R. Intriguing properties of neural networks, 2013.
- [79] SÁNCHEZ-DE-LA TORRE, M., CUBILLOS, C., VEATCH, O. J., GARCIA-RIO, F., GOZAL, D., AND MARTINEZ-GARCIA, M. A. Potential pathophysiological pathways in the complex relationships between osa and cancer. *Cancers* 15, 4 (2023).
- [80] TAN, M., AND LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. *CoRR abs/1905.11946* (2019).
- [81] TRIBUTARIA, A. Tabla de amortización simplificada. [https://sede.agenciatributaria.gob.es/Sede/ayuda/manuales-videos-folletos/manuales-practicos/folleto-actividades-economicas/3-impuesto-sobre-renta-personas-fisicas/3\\_5-estimacion-directa-simplificada/3\\_5\\_4-tabla-amortizacion-simplificada.html](https://sede.agenciatributaria.gob.es/Sede/ayuda/manuales-videos-folletos/manuales-practicos/folleto-actividades-economicas/3-impuesto-sobre-renta-personas-fisicas/3_5-estimacion-directa-simplificada/3_5_4-tabla-amortizacion-simplificada.html). Accedido: 21-05-2024.
- [82] TURING, A. M. On computable numbers, with an application to the entscheidungsproblem. *Journal of Mathematical Society* 2, 42 (1936), 230–265.
- [83] UNIDAS, N. Transformar nuestro mundo: la agenda 2030 para el desarrollo sostenible. [https://unctad.org/system/files/official-document/ares70d1\\_es.pdf](https://unctad.org/system/files/official-document/ares70d1_es.pdf), Septiembre 2015. Asamblea General de las Naciones Unidas.
- [84] VAISH, R., DWIVEDI, U., TEWARI, S., AND TRIPATHI, S. Machine learning applications in power system fault diagnosis: Research advancements and perspectives. *Engineering Applications of Artificial Intelligence* 106 (2021), 104504.

- [85] VAQUERIZO-VILLAR, F., ÁLVAREZ, D., KHEIRANDISH-GOZAL, L., GUTIÉRREZ-TOBAL, G. C., BARROSO-GARCÍA, V., SANTAMARÍA-VÁZQUEZ, E., CAMPO, F. D., GOZAL, D., AND HORNERO, R. A convolutional neural network architecture to enhance oximetry ability to diagnose pediatric obstructive sleep apnea. *IEEE Journal of Biomedical and Health Informatics* 25, 8 (2021), 2906–2916.
- [86] VARADY, P., BONGAR, S., AND BENYO, Z. Detection of airway obstructions and sleep apnea by analyzing the phase relation of respiration movement signals. *IEEE Transactions on Instrumentation and Measurement* 52, 1 (2003), 2–6.
- [87] VGONTZAS, A. N., PAPANICOLAOU, D. A., BIXLER, E. O., HOPPER, K., LOTSIKAS, A., LIN, H.-M., KALES, A., AND CHROUSOS, G. P. Sleep Apnea and Daytime Sleepiness and Fatigue: Relation to Visceral Obesity, Insulin Resistance, and Hypercytokinemia. *The Journal of Clinical Endocrinology & Metabolism* 85, 3 (03 2000), 1151–1158.
- [88] WANG, E., KOPRINSKA, I., AND JEFFRIES, B. Sleep apnea prediction using deep learning. *IEEE Journal of Biomedical and Health Informatics* 27, 11 (2023), 5644–5654.
- [89] XIE, A., RANKIN, F., RUTHERFORD, R., AND BRADLEY, T. D. Effects of inhaled co2 and added dead space on idiopathic central sleep apnea. *Journal of applied physiology* 82, 3 (1985), 918–926.
- [90] YAMASHITA, R., NISHIO, M., DO, R. K. G., AND TOGASHI, K. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging* 9, 4 (2018), 611–629.
- [91] YOUNG, T., EVANS, L., FINN, L., AND PALTA, M. Estimation of the Clinically Diagnosed Proportion of Sleep Apnea Syndrome in Middle-aged Men and Women. *Sleep* 20, 9 (09 1997), 705–706.
- [92] YOUNG, T., PEPPARD, P. E., AND GOTTLIEB, D. J. Epidemiology of obstructive sleep apnea: a population health perspective. *American Journal of Respiratory and Critical Care Medicine* 165, 9 (2002), 1217–1239.
- [93] YUMINO, D., AND BRADLEY, T. D. Central sleep apnea and cheyne-stokes respiration. *Proceedings of the American Thoracic Society* 5, 2 (2008), 226–236.
- [94] ZAFAR, A., AAMIR, M., MOHD NAWI, N., ARSHAD, A., RIAZ, S., ALRUBAN, A., DUTTA, A. K., AND ALMOTAIRI, S. A comparison of pooling methods for convolutional neural networks. *Applied Sciences* 12, 17 (2022).
- [95] ZHANG, A., LIPTON, Z. C., LI, M., AND SMOLA, A. J. Dive into deep learning, Aug 2023.



- [96] ZHANG, G.-Q., CUI, L., MUELLER, R., TAO, S., KIM, M., RUESCHMAN, M., MARIANI, S., MOBLEY, D., AND REDLINE, S. The National Sleep Research Resource: towards a sleep data commons. *Journal of the American Medical Informatics Association* 25, 10 (05 2018), 1351–1358.
- [97] ZHANG, J.-W., ZHENG, C.-X., AND XIE, A. Bispectrum analysis of focal ischemic cerebral eeg signal using third-order recursion method. *IEEE Transactions on Biomedical Engineering* 47, 3 (2000), 352–359.
- [98] ZHUANG, F., QI, Z., DUAN, K., XI, D., ZHU, Y., ZHU, H., XIONG, H., AND HE, Q. A comprehensive survey on transfer learning. *Proceedings of the IEEE* 109 (2019), 43–76.