



UNIVERSIDAD DE VALLADOLID

ESCUELA TÉCNICA SUPERIOR DE
INGENIEROS DE TELECOMUNICACIÓN

TRABAJO FIN DE Grado

GRADO EN INGENIERÍA DE
TELECOMUNICACIÓN

**ANÁLISIS DE ELECTROENCEFALOGRAMAS PARA LA
DETECCIÓN AUTOMÁTICA DE LAS FASES DEL SUEÑO**

Autor:

Dña. Eva Calvo Merino

Tutores:

Dr. D. Gonzalo César Gutiérrez Tobal

Dr. D. Roberto Hornero Sánchez

Valladolid, 12 de Junio de 2020

TÍTULO: **Análisis de electroencefalogramas para la detección automática de las fases del sueño**

AUTOR: **Dña. Eva Calvo Merino**

TUTORES: **Dr. D. Gonzalo César Gutiérrez Tobal**
Dr. D. Roberto Hornero Sánchez

DEPARTAMENTO: **Teoría de la Señal y Comunicaciones e Ingeniería Telemática**

TRIBUNAL

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

VOCAL: **Dr. D. Jesús Poza Crespo**

SECRETARIA: **Dra. Dña. María García Gadañón**

SUPLENTE 1: **Dr. D. Carlos Gómez Peña**

SUPLENTE 2: **Dr. D. Miguel López Coronado**

AGRADECIMIENTOS

En primer lugar, quiero agradecer a mis tutores Gonzalo César Gutiérrez Tobal y Roberto Hornero Sánchez por brindarme la oportunidad de realizar este Trabajo Fin de Grado, por su labor como guías en este primer acercamiento a lo que sería un trabajo de investigación, y por resolver las dudas que se me han planteado a lo largo del curso.

A los miembros del Grupo de Ingeniería Biomédica de la Universidad de Valladolid, por ayudarme en las diferentes etapas del desarrollo de este trabajo.

A todos mis compañeros y amigos, quienes me han acompañado a lo largo de estos 4 años de carrera, haciéndome que fuese una etapa inolvidable.

Y finalmente, a mi familia, por estar ahí siempre que lo he necesitado, apoyándome y animándome.

Muchas gracias a todos.

RESUMEN DEL TFG

Las enfermedades del sueño son cada vez más comunes debido al estresante estilo de vida de la sociedad actual. Un paso fundamental en su estudio y diagnóstico es detectar correctamente las diferentes fases del sueño. Avances en áreas como el *Deep learning* han permitido desarrollar métodos que automatizan esta detección, presentando una alternativa a la clasificación mediante inspección visual realizada hasta la fecha.

En este trabajo se ha indagado en el uso de redes neuronales convolucionales (CNN) como clasificadores de fases del sueño, usando para ello la señal de electroencefalograma (EEG). El comportamiento de esta señal difiere entre niños y adultos. Sin embargo, los estudios publicados hasta ahora se han centrado únicamente en pacientes adultos, lo que provoca que los modelos de clasificación no sean fácilmente generalizables.

Conseguir un método de clasificación basado en CNN que permita una detección precisa de las fases del sueño en niños, y comprobar si se puede entrenar un modelo que alcance resultados óptimos al evaluar sujetos de diferentes edades, son los objetivos principales de este trabajo. Para ello, se han usado dos amplias bases de datos públicas procedentes de los estudios *Sleep Heart Health Study* (SHHS) y *Childhood Adenotonsillectomy Trial* (CHAT), que contienen 5793 registros de adultos y 453 registros de niños, respectivamente. El proceso de entrenamiento y optimización de la red CNN se ha probado modificando el número de capas y su parámetro de regularización, este último buscando asegurar que no haya sobreentrenamiento. Tras conseguir un modelo con alto rendimiento al clasificar la población adulta, se ha evaluado dicho modelo en los registros pediátricos. El mismo procedimiento se ha realizado de manera inversa, probando en la población adulta un modelo entrenado únicamente con niños. Además, se ha obtenido un modelo conjunto usando registros de ambas bases de datos en los grupos de entrenamiento/validación/test. Para homogeneizar las señales de las dos bases, se ha implementado re-muestreo a la misma frecuencia, re-referenciado a la media de los canales utilizados en cada caso, y estandarización para igualar los límites de amplitud.

Los resultados muestran que los modelos entrenados con registros de una única base de datos clasifican con alta precisión siempre que se apliquen sobre sujetos en los mismos rangos de edad, consiguiéndose una precisión del 0.815 y un kappa de Cohen de 0.738 en el caso de sujetos adultos y precisión de 0.84 y kappa de 0.77 en el caso de niños, lo que es coherente con estudios previos. No obstante, al clasificar un grupo de edad diferente, estos valores disminuyen. Sin embargo, el modelo entrenado con registros de diferentes edades sí que consigue detectar de manera precisa registros de ambas bases, llegando a una precisión de 0.81 y a un kappa de 0.75 al evaluarlo en un grupo de test conjunto. Estos resultados sugieren la necesidad de incluir sujetos de diferentes edades en el entrenamiento para conseguir modelos más generalizables.

Palabras clave: Fases del sueño, Electroencefalograma (EEG), Redes neuronales convolucionales (CNN), *Deep learning*, edad.

ABSTRACT

Sleep disorders are very common nowadays due to the stressful lifestyle of the current society. A fundamental step in the study and diagnosis of these disorders is to successfully detect the different sleep stages. Recent investigation in fields like Deep learning has led to the development of methods that automatize this detection, becoming an alternative to the visual classification mostly used up to the date.

This project explores the application of convolutional neural networks (CNN) as methods for sleep staging classification, using the brain signal of the electroencephalogram (EEG). The behavior of this signal changes between children and adults. However, studies published up to the date mostly focus on grown up patients, issue that causes a poor generalization of the classification models when applied to other age ranges.

Finding a classification method based on CNN that shows an accurate detection of children's sleep stages, and training a model that reaches high performance when evaluated with subjects of different ages, are the two main goals of this work. In order to achieve these goals, two large public data bases have been used, coming from the Sleep Heart Health Study (SHHS) and the Childhood Adenotonsillectomy Trial (CHAT), and containing 5793 adults' recordings and 453 children's recordings, respectively. The process of training and optimizing the neural network has been conducted by varying the number of convolutional layers and the dropout percentage, the latter being used to minimize the risk of model overfitting. Once an effective model for the classification of adults' recordings is found, it gets tested with the pediatric recordings. The same procedure is followed the other way around, testing with the recordings of adults a model trained only using kids' signals. Furthermore, a mixed model is obtained by including subjects from both data bases in the training/validation/test groups. With the aim of homogenize the signals of the two data bases, three different actions have been taken: re-sampling the recordings to the same frequency, applying an average reference, and standardizing the signals to keep them with in the same amplitude limits.

The results show that the models trained with just one of the data bases only classify accurately recordings from subjects of that data base, obtaining a Kappa coefficient of 0.74 and an accuracy of 0.82 when just using grown up subjects and a Kappa of 0.77 and accuracy of 0.84 with only children. However, when testing these models on subjects of different age from the ones in the training set the level of performance decreased significantly. On the contrary, the mixed model does succeed when classifying recordings from both age ranges, obtaining an accuracy of 0.81 and a Kappa of 0.75 in the classification of a test group formed by the same number of adults and children. These results support the need to consider subjects of different ages when developing methods for the automatic detection of sleep stages, so the models obtained can adapt to a wider range of patients.

Keywords: Sleep stages, Electroencephalogram (EEG), Convolutional neural networks (CNN), Deep learning, age.

INDICE GENERAL

CAPÍTULO 1 : INTRODUCCIÓN	1
1.1. PROCESADO DE SEÑALES BIOMÉDICAS.....	1
1.1.1. Obtención de señales bioeléctricas.....	3
1.1.2. Tipos de señales bioeléctricas	4
1.2. ELECTROENCEFALOGRAMA (EEG).....	6
1.2.1. Posicionamiento de los electrodos	9
1.3. FASES DEL SUEÑO.....	10
1.4. HIPÓTESIS Y OBJETIVOS.....	12
1.5. ESTRUCTURA DEL TFG	13
CAPÍTULO 2 DETECCIÓN AUTOMÁTICA DE LAS FASES DEL SUEÑO	15
2.1. ANTECEDENTES	15
2.2. OBTENCIÓN DE DATOS.....	16
2.3. PREPROCESADO	17
2.4. SELECCIÓN DE CARACTERÍSTICAS.....	17
2.5. CLASIFICACIÓN	19
2.5.1. <i>K-means</i>	19
2.5.2. LDA (Linear discriminant analysis).....	20
2.5.3. SVM (Support Vector Machines).....	20
2.5.4. ANN (Artificial Neural Networks) clásicas	20
2.5.5. Ensemble learning.....	21
2.5.6. LSTM (Long short-term memory).....	22
2.5.7. CNN (Convolutional neural network).....	22
2.6. Comparación y elección del método a implementar	22
CAPÍTULO 3 : SUJETOS Y SEÑALES	29
3.1. SHHS (<i>Sleep Heart Health Study</i>).....	29
3.1.1. Población bajo estudio	29
3.1.2. Características de las señales EEG.....	30
3.2. CHAT (<i>Childhood Adenotonsillectomy Trial</i>)	31
3.2.1. Población bajo estudio	31
3.2.2. Características de las señales EEG.....	32
3.3. ENTRENAMIENTO, VALIDACIÓN Y TEST	32

CAPÍTULO 4 : METODOLOGÍA	35
4.1. REDES NEURONALES CONVOLUCIONALES.....	35
4.1.1. Estructura general de una CNN.....	37
4.1.2. Ventajas frente a ANN convencionales.....	40
4.2. MODELO ORIGINAL	42
4.2.1. Arquitectura	42
4.2.2. Optimización	44
4.3. SELECCIÓN DE LAS ENTRADAS A LA RED	44
4.4. MEDIDAS DEL RENDIMIENTO DE CLASIFICACIÓN	45
4.5. APLICACIÓN DEL MODELO A NIÑOS.....	47
4.5.1. Re-referenciado.....	47
4.5.2. Estandarizado.....	48
4.6. MODELO ESPECÍFICO PARA NIÑOS	48
4.6.1. Dropout	48
4.6.2. Número de capas	50
CAPÍTULO 5 : RESULTADOS	51
5.1. Resultados con el modelo original	51
5.1.1. Aplicando Re-referenciado.....	53
5.1.2. Aplicando Re-referenciado y estandarizado	54
5.2. Resultados con el modelo entrenado con chat.....	55
5.3. Resultados con el modelo entrenado con chat y shhs.....	57
CAPÍTULO 6 : DISCUSIÓN	61
6.1. Discusión del modelo original	61
6.2. Discusión del modelo específico de niños	62
6.3. Discusión del modelo combinado	63
6.4. Comparación con otros estudios	64
6.5. Limitaciones	69
CAPÍTULO 7 : CONCLUSIONES Y LÍNEAS FUTURAS	71
7.1. CONTRIBUCIONES	71
7.2. CONCLUSIONES	71
7.3. LÍNEAS FUTURAS	73

ÍNDICE DE FIGURAS

CAPÍTULO 1: INTRODUCCIÓN

Figura 1.1: Diagrama de bloques describiendo los pasos principales en el análisis de señales biomédicas.	4
Figura 1.2: Empezando desde arriba se muestra la representaciones de un electroencefalograma (EEG), un electrocardiograma (ECG) y un electromiograma (EMG). Las tres sacadas de la base de datos MIT-BIH.....	5
Figura 1.3: Representación de todas las medidas realizadas durante una polisomnografía.....	7
Figura 1.4: Representaciones de la señal de electroencefalograma asociada a diferentes estados: (a) animado, (b) relajado, (c) adormecido, (d) dormido y (e) en sueño profundo.....	8
Figura 1.5: Distribución de los electrodos según el sistema internacional 10/20.	9

CAPÍTULO 4: METODOLOGÍA

Figura 4.1: Esquema de las etapas principales en las que se basa el entrenamiento de una red neuronal y como se suceden entre sí.....	36
Figura 4.2: Ejemplo de convolución 2-D.....	37
Figura 4.3: Composición de una capa de convolución 1D, incluyendo tanto la fase de convolución cómo el submuestreo. La función de activación no se representa.	38
Figura 4.4: Ejemplos de funciones de activación no lineales.....	39
Figura 4.5: Efecto equivalente de submuestrear o aplicar un stride.	39
Figura 4.6: Submuestreo con max pooling y average pooling.	40
Figura 4.7: Se muestra en sombreado una unidad de entrada x_3 , y las unidades de salida s_i que dependen de dicha unidad de entrada. Arriba las salidas se forman convolucionando con un filtro de tamaño 3 (caso de una CNN), por lo que solo 3 salidas se ven afectadas por una misma entrada. Abajo las salidas se forman por multiplicación matricial (caso de red neuronal), por lo que todas las salidas se ven afectadas.	41
Figura 4.8: Arquitectura de la red convolucional 1D usada en Sors.	43
Figura 4.9: Representación de una misma red neuronal con, derecha, y sin dropout, izquierda.	49

CAPÍTULO 5: RESULTADOS

- Figura 5.1: Evolución de la precisión y el coste a medida que se entrenaba el modelo, para los grupos de entrenamiento y validación. 52
- Figura 5.2: Matriz de confusión conseguida al evaluar, con el grupo de test de la base de datos SHHS, el modelo entrenado con el grupo de entrenamiento de SHHS (14 capas y no dropout). 52
- Figura 5.3: Matriz de confusión conseguida al evaluar, con todos los registros de la base de datos CHAT, el modelo entrenado con el grupo de entrenamiento de SHHS (14 capas y no dropout). 53
- Figura 5.4: Matrices de confusión conseguidas al evaluar, con el grupo de test de SHHS (izquierda) y con todos los registros de la base de datos CHAT (derecha) , el modelo entrenado con el grupo de entrenamiento de SHHS (14 capas y no dropout). En este caso todos los registros han sido previamente re-referenciados..... 54
- Figura 5.5: Matrices de confusión conseguidas al evaluar, con el grupo de test de SHHS (izquierda) y con todos los registros de la base de datos CHAT (derecha) , el modelo entrenado con el grupo de entrenamiento de SHHS (14 capas y no dropout). En este caso todos los registros han sido previamente re-referenciados y estandarizados..... 55
- Figura 5.6: Matrices de confusión conseguidas al evaluar, con el grupo de test de CHAT (derecha) y con todos los registros de la base de datos SHHS (izquierda) , el modelo entrenado con el grupo de entrenamiento de CHAT (14 capas y un índice de dropout de 0.4). Todos los registros han sido previamente re-referenciados y estandarizados..... 57
- Figura 5.7: Matriz de confusión conseguida al evaluar en el grupo de test conjunto (72 pacientes adultos y 72 niños), el modelo de 18 capas convolucionales entrenado también con registros de ambas edades..... 58
- Figura 5.8: Matrices de confusión conseguidas al evaluar con registros de la base de datos CHAT (derecha) y con registros de la base SHHS (izquierda) , el modelo entrenado con el grupo de entrenamiento conjunto (18 capas y sin dropout). La evaluación se realiza usando los sujetos que no han formado parte de grupo de entrenamiento ni validación. (Lo que implica que se evalúa en 72 sujetos para la base CHAT y 5340 para la base SHHS). 59

ÍNDICE DE TABLAS

CAPÍTULO 1: INTRODUCCIÓN

Tabla 1.1: Fases del sueño y sus principales características en la señal de EEG.....	12
--	----

CAPÍTULO 2: DETECCIÓN AUTOMÁTICA DE LAS FASES DEL SUEÑO

Tabla 2.1: Métodos inicialmente seleccionados como posibles estudios que tomar como base para el resto del TFG.	23
--	----

CAPÍTULO 3: SUJETOS Y SEÑALES

Tabla 3.1: Características de los 5793 pacientes que forman el grupo SHHS-1 en la base de datos SHHS.	30
--	----

Tabla 3.2: Características de los 453 pacientes que forman el grupo baseline en la base de datos CHAT.	31
---	----

CAPÍTULO 5: RESULTADOS

Tabla 5.1: Comparación de los resultados conseguidos en este TFG y los conseguidos por [16] al aplicar el mismo modelo.	52
--	----

Tabla 5.2: Tabla comparativa con los valores del coeficiente Kappa y precisión obtenidos al aplicar el modelo original a las bases de datos originales, re-referenciadas o re-referenciadas y estandarizadas.	55
--	----

Tabla 5.3: Resultados obtenidos al entrenar un modelo de CNN que consiga un alto rendimiento al clasificar los EEG pediátricos.	56
--	----

Tabla 5.4: Resultados obtenidos al entrenar un modelo de CNN con registros de adultos y niños, provenientes de las bases de datos de SHHS y CHAT, respectivamente. La evaluación se realiza usando un grupo de validación con registros de ambas bases de datos, concretamente incluye a 65 pacientes de cada una.....	58
--	----

Tabla 5.5: Resultados de aplicar a las bases de datos por separado el modelo conjunto de 18 capas convolucionales. La evaluación se realiza usando los sujetos que no han formado parte de grupo de entrenamiento ni validación. (Lo que implica que se evalúa en 72 sujetos para la base CHAT y 5340 para la base SHHS).....	59
---	----

CAPÍTULO 6: DISCUSIÓN

Tabla 6.1: comparación con otros métodos de detección automática de las fases del sueño.
Parte I 65

Tabla 6.2: comparación con otros métodos de detección automática de las fases del sueño.
Parte II 66

Capítulo 1 : INTRODUCCIÓN

1.1. PROCESADO DE SEÑALES BIOMÉDICAS

El procesado de señales biomédicas constituye un área de investigación multidisciplinar que engloba campos como la medicina, ingeniería o la biología y que busca analizar las diferentes señales generadas por el cuerpo humano [1]. Muchas de las propiedades de los sistemas biológicos están reflejadas en las señales que generan, pero gran parte de la información que éstas llevan no está disponible a simple vista, sino que es necesario estudiar su estructura en profundidad.

Gracias al análisis de dichas señales se han podido realizar grandes avances al identificar y explicar distintas enfermedades, facilitando su diagnóstico y, en algunas ocasiones, hasta automatizándolo [2]. Además, la información obtenida mediante el preprocesado de las señales biológicas ayuda al personal sanitario a seleccionar las terapias más adecuadas y a seguir de manera más fiable la evolución de los pacientes. Las mejoras tecnológicas de los últimos años han permitido que las técnicas de la Ingeniería Biomédica sean cada vez más útiles, económicas y eficientes, convirtiéndose en un elemento muy presente en el panorama médico actual [1].

Es a la hora de aplicar estos avances tecnológicos en el ámbito de la salud donde se unen la ingeniería y la medicina, dando lugar a la Ingeniería Biomédica previamente mencionada. Una de las principales ventajas de que los conocimientos científicos y tecnológicos cobren importancia en el ámbito de la salud es que se gana objetividad [2]. Hasta hace poco, la gran mayoría del análisis de señales biológicas se realizaba con técnicas manuales o de manera visual, lo que provocaba una alta discordancia entre diagnósticos realizados por distintos médicos, e incluso entre diagnósticos realizados por un mismo médico en momentos diferentes [2]. Esto introduce un nivel de subjetividad que se minimizaría si se utilizasen algoritmos de apoyo en el diagnóstico, que asegurasen el seguimiento de unas pautas comunes en todos los casos. Otra aplicación muy útil del procesado de señales en la medicina es el mejorar la calidad de las medidas, eliminando el efecto de posibles interferencias, ruido, o artefactos que existan al realizar las grabaciones [2]. Estas interferencias pueden ser causadas por otros procesos biológicos ajenos al que se está estudiando, por errores al colocar los equipos de medida o por la señal eléctrica de alimentación, entre otros.

La clasificación de las distintas señales biomédicas se puede llevar a cabo de acuerdo a diversos criterios, diferenciándolas en función de las técnicas de adquisición usadas, del origen de las señales o de sus características principales. Atendiendo a su descripción matemática se distingue entre **señales deterministas** y **señales aleatorias**. Las primeras pueden ser descritas de manera exacta mediante una ecuación matemática, por lo que se pueden calcular valores futuros de manera fiable a partir de valores ya existentes [3]. Las segundas, por el contrario, no permiten predecir su comportamiento a partir de muestras pasadas ni mediante el uso de

expresiones matemáticas [3], por lo que hay que usar otras herramientas de análisis estadístico. En realidad, al tratar con señales biomédicas, se tiene una mezcla de los dos tipos, pero dependiendo de cuál predomine y de la finalidad del estudio se pueden aproximar a uno, otro o ambos [3].

Otra posible clasificación de las señales sería atendiendo a su origen y a sus características principales, en este ámbito destacan las siguientes:

- **Señales bioacústicas.** Son muchos los procesos biológicos que generan señales acústicas, como la circulación de sangre en el corazón o el movimiento de aire en las vías respiratorias al inhalar o exhalar. Dado que el sonido se propaga a través de los tejidos biológicos, estas señales no son muy complicadas de conseguir desde la superficie mediante métodos poco invasivos, y permiten ayudar a diagnosticar los diversos problemas que puedan existir en los sistemas que las generan [1].
- **Señales bioeléctricas.** Son causadas por los potenciales de membrana de las células nerviosas y musculares. Para detectar estas señales lo más habitual es utilizar electrodos que, colocados en la superficie corporal, actúan como sensores y detectan la actividad eléctrica generada al combinarse la acción de varias células. También existen técnicas más invasivas que permiten medir la actividad eléctrica de manera más localizada, pero su uso está menos extendido. Las señales bioeléctricas son las más estudiadas dentro del conjunto de señales biológicas, dado que están presentes en los sistemas fisiológicos más importantes. Este tipo de señal es el que se tratará con más detalle a lo largo de este Trabajo Fin de Grado [2].
- **Señales de bioimpedancias.** La impedancia de los tejidos puede aportar mucha información, por ejemplo datos acerca de su composición, o de su actividad sanguínea, nerviosa y endocrina. Para acceder a esta información se analizan estas señales de bioimpedancia que se generan al introducir en el tejido corrientes sinusoidales [1].
- **Señales biomecánicas.** Éstas son las señales que se generan a causa de los diferentes procesos mecánicos que tienen lugar en el cuerpo, procesos que generan algún tipo de desplazamiento, fuerza tensión o presión. Debido a que estas señales no se propagan, a diferencia de lo que ocurría por ejemplo con las señales bioeléctricas, su adquisición se debe realizar en el punto en el que se generan y esto provoca que haya que usar procedimientos invasivos [1].
- **Señales ópticas.** Las señales bioópticas pueden ser generadas de manera natural, o provocadas al realizar las medidas, y resultan de las funciones ópticas del sistema biológico [1].
- **Señales biomagnéticas.** Este tipo de señales proviene de los campos magnéticos que se generan en diferentes órganos, como el cerebro, los pulmones o el corazón. Pueden aportar información no presente en los otros tipos de señales, pero llevan asociada la

desventaja de que los campos magnéticos del cuerpo humano son muy débiles, lo que provoca que la obtención de estas señales sea bastante complicada [1].

- **Señales bioquímicas.** Estas señales surgen tras realizar en laboratorios especializados pruebas químicas a partir de los distintos tejidos orgánicos. Estas pruebas incluyen medir las concentraciones iónicas o la presión de distintos gases en el sistema sanguíneo [1].

Como ya se ha mencionado, dentro del conjunto de las señales biomédicas en este trabajo se va a indagar más en las señales bioeléctricas. Concretamente, en el electroencefalograma (EEG) o actividad eléctrica del cerebro.

1.1.1. OBTENCIÓN DE SEÑALES BIOELÉCTRICAS

La generación de las señales bioeléctricas se puede estudiar en diferentes ámbitos, desde estudios a nivel celular hasta centrándose únicamente en las señales que se consiguen en la superficie del cuerpo humano. A nivel celular, las características de permeabilidad de la membrana plasmática provocan que en condiciones de descanso el interior de las células posea una mayor concentración de iones negativos que el exterior, lo que produce un potencial transmembrana negativo. Cuando es estimulada con corriente suficiente, la membrana cambia su permeabilidad abriendo canales a través de los que se da la difusión de iones, modificando así su potencial transmembrana. Este efecto causa la propagación de la corriente entre células adyacentes en un mismo tejido y, repetido en multitud de células, es lo que acaba generando las señales bioeléctricas que luego se detectan [3].

La obtención de estas señales hoy en día se puede llevar a cabo con equipos de precio no muy elevado que consiguen amplificarlas y digitalizarlas de manera adecuada, lo que ha ayudado a que sean la base de muchos de los procedimientos clínicos que se realizan en los hospitales de todo el mundo. Los sistemas de adquisición de señales biomédicas usan frecuencias de muestreo bajas, sin superar 1kHz, e incluyen los siguientes componentes principales [2]: uno o más sensores junto con el instrumental necesario para colocárselos al paciente, etapas de amplificación, una etapa de conversión analógica/digital (en este proceso suele valer con usar 12-14 bits por muestra) y algún tipo de software que permita analizar las señales obtenidas. En algunos casos se requiere que este análisis se lleve a cabo en tiempo real, mientras que en otros las señales se almacenan en alguna memoria y se procesan *a posteriori*, éste último es el caso de este estudio. En la *Figura 1.1* se muestra el diagrama de bloques que refleja este proceso. Es importante amplificar las señales nada más obtenerlas porque, por lo general, las señales biomédicas son muy débiles y de lo contrario se perdería mucha de la información que contienen.

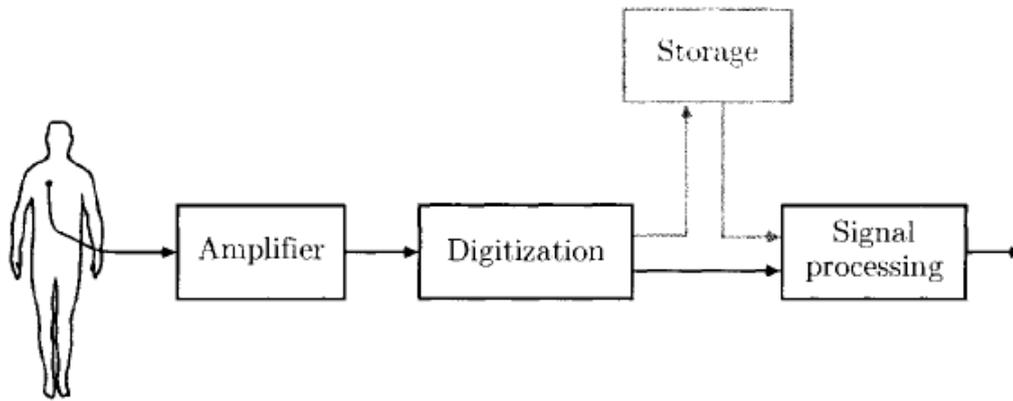


Figura 1.1: Diagrama de bloques describiendo los pasos principales en el análisis de señales biomédicas [2].

En el caso de que se usen electrodos a la hora de grabar las señales, se necesita colocar al menos dos, de manera que el primero se sitúe cerca de la zona dónde se genera la señal de interés y que el segundo se pueda usar como referencia. En muchas ocasiones se prefiere obtener las señales usando un mayor número de electrodos debido a que así se consigue también información espacial de la señal a medir.

1.1.2. TIPOS DE SEÑALES BIOELÉCTRICAS

Dependiendo de la zona en la que se generan se distinguen diferentes tipos de señales bioeléctricas, algunas de ellas se muestran en la *Figura 1.2*.

El **electroencefalograma** (EEG) refleja la actividad eléctrica del cerebro [2]. Se consigue mediante diversos electrodos situados en la cabeza y cuando dichos electrodos se sitúan directamente en el cerebro se denomina **electrocortigrama** (ECoG).

Los **potenciales evocados** (EPs) reflejan actividad cerebral que ha sido causada por un estímulo sensorial, como pueden ser los estímulos visuales o acústicos [2]. Son señales muy usadas en la detección de posibles alteraciones de las vías sensitivas y se consiguen mediante electrodos, de manera similar a la que se usa con el EEG.

El **electrocardiograma** (ECG) muestra la actividad eléctrica del corazón y se obtiene también mediante electrodos, pero en este caso colocándolos en el pecho, brazos y piernas. Estas señales son de gran utilidad a la hora de tratar enfermedades cardiovasculares, cuando se consiguen con electrodos situados directamente sobre el corazón se les denomina **electrogramas** (EG) [2].

El **electromiograma** (EMG) registra la actividad eléctrica de los músculos, actividad que suele ser proporcional al nivel de actividad al que están siendo sometidos. Su uso es frecuente a la hora de detectar inflamaciones fuera de lo normal o lesiones en los nervios de las extremidades. Se diferencia entre EMG extramuscular o intramuscular, dependiendo de si los electrodos se sitúan sobre la piel que cubre el músculo o, por medio de agujas, en el propio músculo [2].

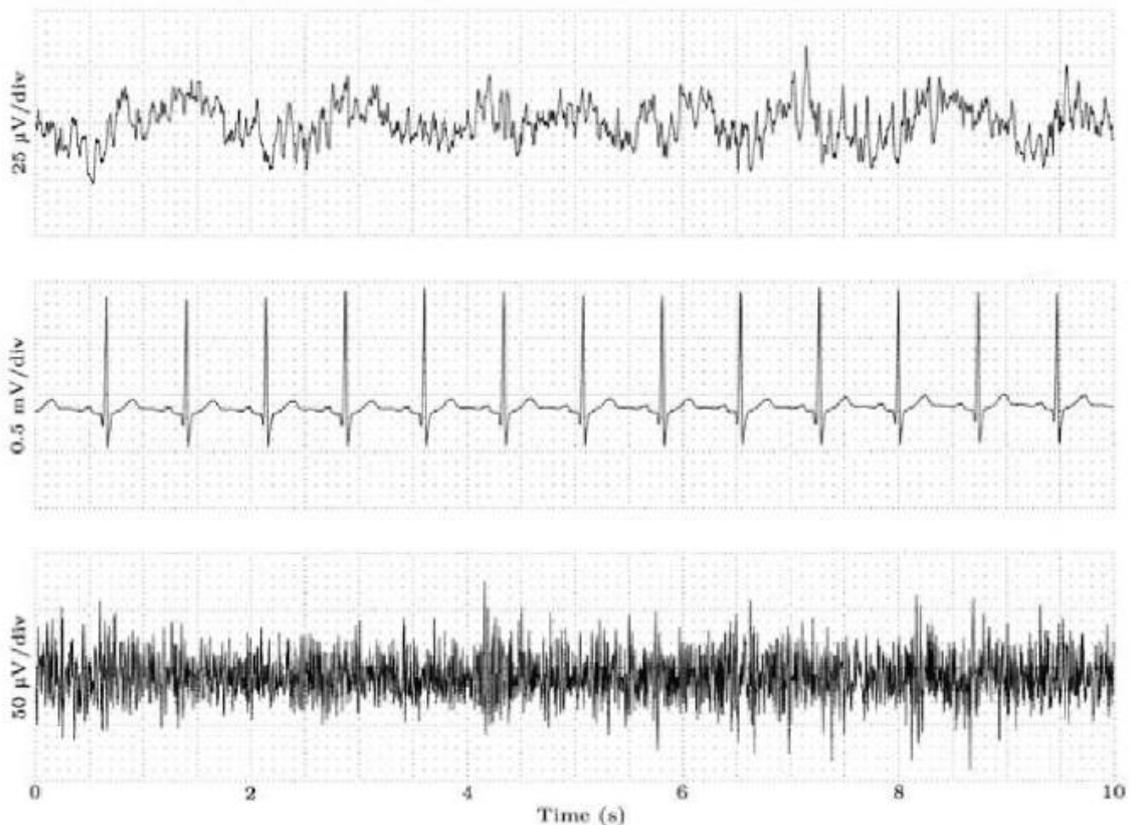


Figura 1.2: Empezando desde arriba se muestra la representaciones de un electroencefalograma (EEG), un electrocardiograma (ECG) y un electromiograma (EMG). Las tres sacadas de la base de datos MIT-BIH [5].

El **electroneurograma** (ENG) resulta tras estimular los nervios periféricos con impulsos eléctricos, de manera que la respuesta del nervio a dicho estímulo pueda ser estudiada y se puedan medir distintas magnitudes, como la velocidad de conducción [2].

El **electroretinograma** (ERG) se usa para estudiar los potenciales eléctricos que genera la retina bajo estimulación luminosa. Se consigue mediante lentes de contacto especiales que llevan un electrodo integrado. También relacionado con la actividad eléctrica del ojo, pero en este caso reflejando sus movimientos horizontales y verticales, está el **electrooculograma** (EOG) [2]. El EOG tiene mucho interés a la hora de tratar enfermedades del sueño ya que, como se verá en más profundidad en capítulos posteriores, la presencia de los llamados movimientos rápidos oculares (REM por *rapid eye movements*) es una de las características que permite diferenciar las fases del sueño [4]. Otro ámbito en el que se usa esta señal es en las aplicaciones de realidad virtual, ya que permite identificar hacia dónde está mirando el usuario [2].

El **electrogastrograma** (EGG) refleja los impulsos que se propagan a través de los músculos del estómago y que controlan sus movimientos al realizar la digestión [2].

En el siguiente apartado se describe con más detalle la señal bioeléctrica del EEG y sus características, ya que es la señal en la que se va a basar el resto del trabajo.

1.2. ELECTROENCEFALOGRAMA (EEG)

El funcionamiento del cuerpo humano está asociado con multitud de señales bioeléctricas que, adecuadamente procesadas y decodificadas, aportan gran cantidad de información sobre los sistemas biológicos que las generaron. Una de estas señales es el EEG, que refleja la actividad eléctrica del cerebro, actividad que engloba las diferentes señales eléctricas que se generan en millones de neuronas en cada instante. El EEG, como se ha comentado antes brevemente, se obtiene de forma no invasiva a través de varios electrodos colocados sobre el cuero cabelludo y es de gran ayuda para el diagnóstico y el estudio de patologías como la epilepsia, la enfermedad del Alzheimer o diferentes trastornos de sueño [2].

El estudio de la actividad cerebral ha interesado al hombre desde hace muchos años, en 1875 Richard Caton demostró que en la corteza cerebral de los animales se podían conseguir señales eléctricas [6]. Años más tarde (1929), Hans Berger desarrolló el primer EEG [6], apuntando que las ondas cerebrales no solo difieren entre pacientes sanos y enfermos, sino que también dependen del estado de atención, relajación o sueño. Desde entonces la interpretación clínica del EEG, y más concretamente la caracterización de las diferentes fases de sueño, se ha convertido en una disciplina en sí misma.

La forma tradicional de obtener información del EEG es a través de la observación visual de su frecuencia, amplitud, o distribución espacial. Sin embargo, limitaciones inherentes a esta metodología, como la existencia de información no observable visualmente o el consumo de tiempo de los especialistas, han llevado a sustituirla por técnicas de procesado de señal mediante algoritmos automáticos que se apoyan en las nuevas tecnologías. Además, la utilización de dichos algoritmos permite realizar un análisis del EEG no dependiente del observador y, por tanto, más objetivo, minimizándose el riesgo de errores y aumentando las opciones de conseguir un diagnóstico más apropiado.

La mayoría de estudios en la línea del actual (relacionados con el diagnóstico de las distintas enfermedades del sueño) utilizan la técnica de la polisomnografía (*Polysomnography*, PSG). La PSG es una prueba en la que se registran múltiples señales biomédicas simultáneamente, normalmente durante el transcurso de una noche. Entre estas señales obtenidas se encuentran, además del EEG, otras como el ECG, EMG y EOG mencionados anteriormente, el flujo aéreo, la SpO₂ o la frecuencia cardíaca [2]. En algunas ocasiones se graban también vídeos de la estancia nocturna de los pacientes, con el fin de usarlos luego para identificar, por ejemplo, posibles movimientos. En la *Figura 1.3* se muestra una representación de esta técnica [2]. De entre todas las señales que forman la PSG, para detectar las fases del sueño interesan sobre todo el EEG, el EOG y los vídeos, aunque en este trabajo se decide utilizar solamente un único canal de EEG (*single channel*).

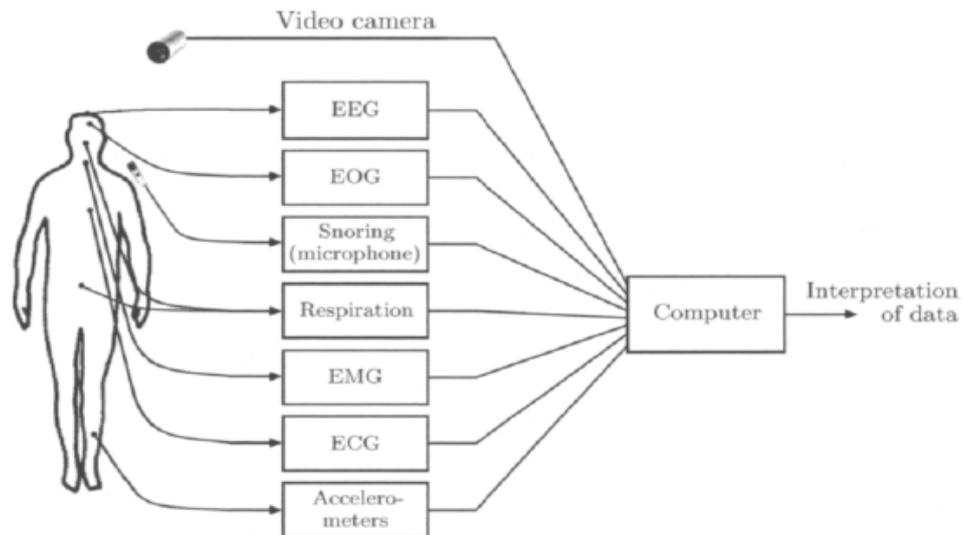


Figura 1.3: Representación de todas las medidas realizadas durante una polisomnografía [2].

El centrarse en métodos que utilicen un solo canal de EEG en vez de varios simplifica bastante el proceso por el que se obtienen los registros, pues se necesita un menor número de señales y se requiere menos instrumental para conseguirlas. A mayores, esta técnica reduciría también el número de artefactos presentes en la señal causados por los diferentes elementos del equipamiento, pues éste es mucho más simple. El conseguir un método que detecte de forma eficiente las fases del sueño basándose únicamente en el EEG de un canal es especialmente conveniente a la hora de desarrollar dispositivos portátiles y no invasivos que permitan monitorizar el sueño en la propia casa de los pacientes [7]. Este tipo de dispositivos podría ser de gran ayuda en el diagnóstico temprano de trastornos del sueño.

Las principales características visuales de la señal de EEG son su amplitud y frecuencia. La amplitud de esta señal viene dada en gran medida por la sincronización que haya entre las neuronas en ese instante. Cuando este grado de sincronización es elevado las actividades individuales de las neuronas generan interferencias constructivas (la amplitud de la señal EEG resultante será mayor), mientras que en situaciones de baja sincronización predominan las interferencias destructivas (amplitud del EEG menor). Ello provoca que el rango de amplitudes que se pueden observar vaya desde unidades de μV hasta alrededor de $100 \mu\text{V}$ [2].

La actividad frecuencial del EEG durante el sueño se concentra entre 1 y 70Hz, por lo que se trabaja con frecuencias bastante bajas. Dentro de este rango de frecuencias se diferencian 5 sub-bandas [8]: delta ($< 4 \text{ Hz}$) principalmente asociada al sueño profundo, theta (4-8 Hz) durante somnolencia y algunas fases del sueño, alpha (8-13 Hz) mayormente presente en individuos despiertos en momentos de relax, beta 1 (14-19 Hz), beta 2 (19-30 Hz) y gamma ($>30 \text{ Hz}$), las tres últimas sub-bandas, las de mayor frecuencia, se asocian con diferentes grados de consciencia y alta actividad cerebral [2], [9]–[13]. En la *Figura 1.4* se puede ver cómo la estructura del EEG cambia, modificando tanto su frecuencia como su amplitud dependiendo del estado en el que se encuentre el sujeto.

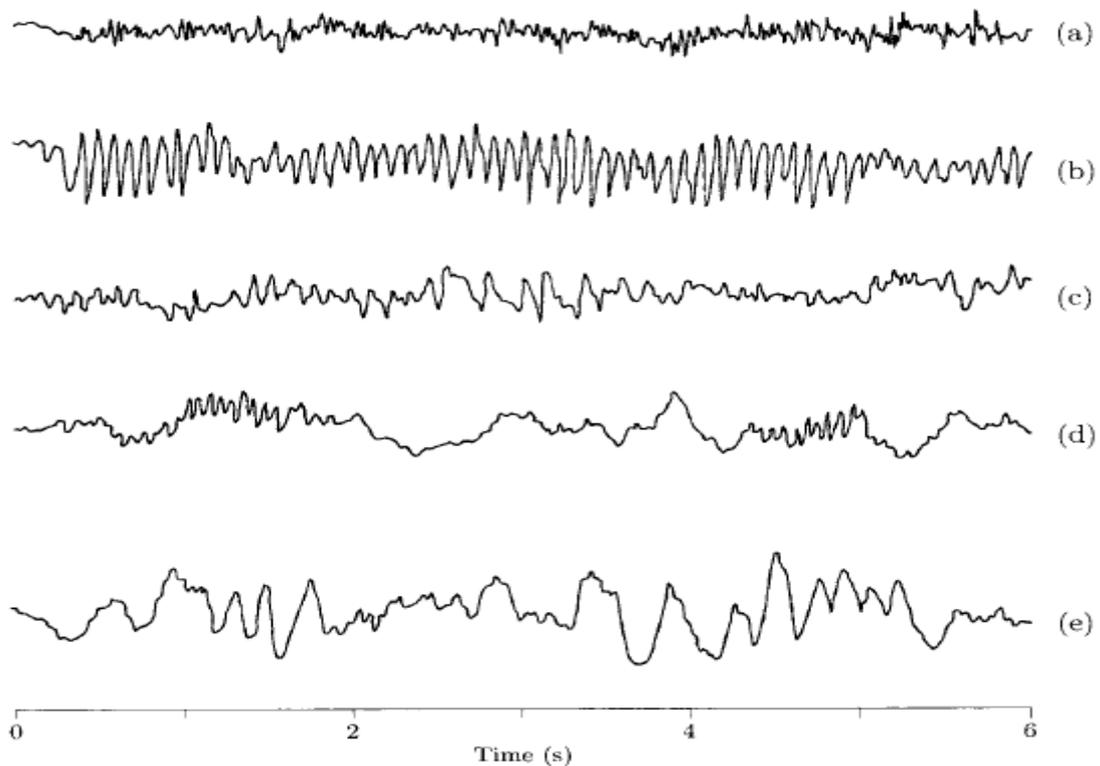


Figura 1.4: Representaciones de la señal de electroencefalograma asociada a diferentes estados: (a) animado, (b) relajado, (c) adormecido, (d) dormido y (e) en sueño profundo [2].

A lo largo del EEG también ocurren diversos micro eventos que ayudan a diferenciar las fases del sueño, pues el modo en el que aparecen es característico para cada fase; entre estos micro eventos destacan los *spindles*, *V waves* y los *k-complexes*, que se explicarán más adelante [5,6].

La señal EEG obtenida puede verse afectada por diferentes fuentes de ruido, tanto externas al cuerpo humano como causadas por otras señales biomédicas. Entre las fuentes externas se encuentran las interferencias con otros equipos de medida y con la red eléctrica, mientras que, entre las señales internas del cuerpo humano, las que más interfieren con el EEG suelen ser los movimientos oculares y musculares.

En los últimos años han surgido nuevas técnicas de adquisición de imágenes médicas cada vez más sofisticadas que han revolucionado el estudio de las funciones cerebrales y han restado protagonismo al EEG. Entre ellas destacan técnicas como la PET (*Positron Emission Tomography*), la SPECT (*Single Photon Emission Computed Tomography*) o la MRI (*Magnetic Resonance Imaging*). Sin embargo, en aplicaciones que necesitan priorizar la resolución temporal respecto a la espacial el EEG sigue dominando, presentando además la ventaja de ser más barato [2]. Entre estas aplicaciones, además del diagnóstico de enfermedades del sueño, destacan el tratamiento de la epilepsia, el estudio de la esquizofrenia y la enfermedad del Alzheimer y la generación de interfaces de comunicación cerebro-máquina (BCI, *Brain Computer Interface*) [2].

1.2.1. POSICIONAMIENTO DE LOS ELECTRODOS

Existen estándares que rigen la posición que han de seguir los electrodos cuando se desea registrar la señal de EEG de manera no invasiva. El estándar más comúnmente usado es el sistema internacional 10/20, aprobado por la Federación Internacional de Sociedades de Electroencefalografía y Neurofisiología Clínica [14]. En él se indican ciertas reglas que buscan optimizar la obtención del EEG, situando los electrodos distribuidos a lo largo de la cabeza y buscando que las señales finalmente conseguidas optimicen la aportación de información, tanto espacial como temporal, de la actividad cerebral.

En este estándar particular se utilizan 21 electrodos que se distribuyen en puntos específicos del cuero cabelludo y se diferencian dependiendo del hemisferio y la zona del córtex en la que se colocan [14]. De esta manera, cada electrodo se reconoce mediante una o dos letras (la primera en mayúscula) y un subíndice numérico. El número indica si dicho electrodo se encuentra en el lado izquierdo de la cabeza (impares) en el lado derecho (pares), o en la zona central, siendo cero el número en este último caso y representándose con una "z". La letra indica la sección en la que se sitúa el electrodo, pudiendo ser una de las seis zonas en las que se divide el cráneo (frontopolar (Fp), frontal (F), central (C), parietal (P), occipital (O) y temporal (T)) o también los lóbulos de las orejas (A1 (izquierdo) y A2 (derecho)), estos últimos se suelen usar para ubicar los electrodos referencia, ya que son puntos que presentan un voltaje relativamente constante [4].

En la *Figura 1.5* se muestra la distribución que se acaba de detallar, los puntos de referencia usados son la zona superior de la nariz (*nasion*) y la zona del final del cráneo (*inion*) [2].

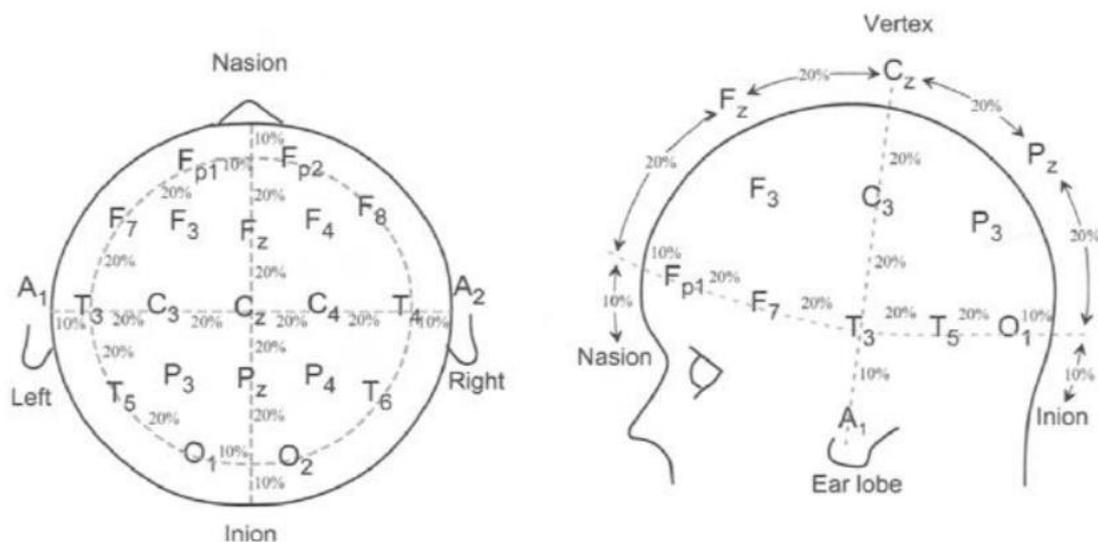


Figura 1.5: Distribución de los electrodos según el sistema internacional 10/20 [2].

1.3. FASES DEL SUEÑO

Uno de los principales factores en una vida saludable es tener una buena calidad de sueño. Éste es un estado físico y mental que lleva asociada una gran actividad, pues incluye movimientos cerebrales, musculares, oculares, del corazón y respiratorios [15]. Aproximadamente un tercio de la vida humana se desarrolla durmiendo, momento en el que se llevan a cabo procesos fisiológicos como el metabolismo o la consolidación de la memoria [16], sin embargo, pese a su gran importancia, el estilo de vida cada vez más estresante de la sociedad actual, y la incorporación de las nuevas tecnologías, han provocado que la calidad y cantidad del sueño hayan decaído y que cada vez más personas sufran de trastornos del mismo [16]. A mayores, el aumento de la esperanza de vida humana también ha incrementado el número de personas que sufren estas enfermedades, pues la calidad del sueño disminuye notablemente con la edad. Los trastornos del sueño, entre los que encontramos insomnio, síndrome de la apnea-hipopnea del sueño, síndrome de las piernas inquietas (RLS) o narcolepsia, pueden provocar cansancio, depresión o ansiedad entre otros efectos negativos [17], afectando enormemente al día a día de la persona que los padece, dificultando tanto sus relaciones sociales como su productividad laboral.

De acuerdo con estudios previos, alrededor de un 25% de la población adulta mundial sufre de enfermedades del sueño, y en un alto porcentaje de casos las personas afectadas recurren a ayuda médica para superarlas [18]. Por ello, el conseguir un diagnóstico efectivo de estas enfermedades que pueda facilitar su posterior tratamiento, se ha convertido en un tema de gran importancia en la comunidad científica. En este contexto, el primer paso para un buen diagnóstico es la correcta detección de las diferentes fases del sueño, y es en este primer paso en el que se centra este trabajo.

Entre las fases del sueño existen dos bien diferenciadas: el sueño de movimientos oculares rápidos, conocido como sueño REM (*Rapid Eye Movement*), y el sueño de ondas lentas, también llamado sueño No-REM; a lo largo de una noche se pasa varias veces por cada uno de estos dos tipos de sueño. La fase No-REM, a su vez, la componen cuatro estadios que se hallan relacionados con los distintos grados de profundidad del sueño. Aparecen así, sucesivamente, estadios 1 y 2, o de sueño ligero, y estadios 3 y 4, que corresponden al sueño profundo [4]. Conocer el patrón formado por las diferentes fases y estadios mientras dormimos resulta fundamental para establecer la calidad del sueño y detectar posibles patologías asociadas.

La estructura de la señal EEG presenta formas de onda características dependiendo de la fase del sueño en la que se encuentra el individuo en cada instante. Entre estas formas de onda destacan [4], [16]:

- **Vertex waves o V waves.** Son ondas con picos muy pronunciados que duran algo menos de 0.5 s y son fácilmente distinguibles del resto de actividad cerebral. Suelen darse en las primeras fases del sueño en respuesta a estímulos externos, como por ejemplo sonidos.
- **Sleep spindles.** Se refieren a que duran más de 0.5 s y tienen una frecuencia de entre 11 y 16 Hz [4].
- **K complexes.** Formas de onda que surgen de la fusión de las dos anteriores.

- **Sawtooth waves.** Secuencia de ondas de apariencia triangular o aserrada y frecuencia de 2-6 Hz. Se dan en la región central de la cabeza.

Atendiendo tanto a estas formas de onda como a la actividad frecuencial de las señales, comentada en la sección 1.2, las características de la señal EEG que definen las diferentes fases del sueño en adultos se resumen en la *Tabla 1.1* [4], [16].

Los típicos ciclos de sueño incluyen secuencias específicas de estas fases, lo que ayuda a la hora de clasificarlas. Por ejemplo, es común encontrarse las transiciones de fases 1-2-3 o 3-2-1, mientras que otras, como pasar directamente de estar despierto a sueño profundo, son mucho menos probables [19].

Al tratar con pacientes pediátricos las características de las fases del sueño varían algo respecto a las vistas anteriormente. Esto se debe a que en niños la actividad en la banda alpha no está presente de la misma manera que en adultos, sino que en su lugar aparecen los ritmos básicos posteriores (*Dominant posterior rhythm*, DPR) [4]. Por ello, a la hora de detectar la fase W y la fase 1 se busca en el EEG la presencia de estos ritmos. Los DPRs se dan sobre la zona occipital del cerebro y en niños presentan una amplitud de más de 50 μ V. La frecuencia de los DPRs depende de la edad del sujeto: 3.5-4.5 Hz en bebés de hasta 4 meses, 5-6 Hz entre los 4 meses y el año de edad, en niños de 3 años se centra entre 7.5-9.5 Hz y así va aumentando hasta que en sujetos adultos estos ritmos DPRs derivan en el ritmo alpha (una frecuencia de 8.5-13 Hz) [4]. Además, también se ha demostrado que en adultos el porcentaje de sueño REM es algo menor que en niños [20].

Otra de las diferencias entre el EEG adulto y pediátrico reside en que, a lo largo de la adolescencia, disminuye la amplitud de la actividad cerebral en la banda delta durante las fases no REM [21]. Estudios han mostrado que la amplitud de este tipo de ondas (delta) incrementa mucho en los primeros años de vida, alcanzando un máximo en la infancia, y descendiendo al llegar a la adolescencia [22]. En [21] este hecho se relaciona con el desarrollo neuronal, que en la infancia es muy amplio y al acercarse a la edad adulta se estabiliza. Esto provocará que el aspecto de las fases 3 y 4, pues son en las que la presencia de ondas delta es máxima, cambie entre los registros adultos y pediátricos. Siguiendo en esta línea, en [23] se discute cómo la potencia espectral del EEG disminuye con la edad.

En las últimas décadas la investigación en la materia del sueño ha ido en aumento, en 1968 Allan Rechtschaffen y Anthony Kales publicaron los primeros estándares para métodos de caracterización del sueño normal [24]. Posteriormente, la necesidad de métricas que también tuviesen en cuenta casos de sueño anómalos provocó que en 2003 la AASM (*American Academy of Sleep Medicine*) formará un comité para desarrollar un nuevo manual de métodos: *The AASM Manual for the Scoring of Sleep and Associated Events*, publicado en 2007 y posteriormente revisado en 2012 y 2017 [4], [25] y [26]. Entre sus diferentes recomendaciones, la AASM propone fusionar las fases del sueño 3 y 4 en una única fase. Dado que aún no se han encontrado métodos efectivos con los que realizar un diagnóstico lo suficientemente eficiente de los diversos trastornos de sueño existentes, este es un tema de investigación en el que se considera que se debe seguir profundizando.

Tabla 1.1: Fases del sueño y sus principales características en la señal de EEG.

Fase	Profundidad de sueño	Características de la señal
W (wake)	El sujeto está despierto en esta fase	Bandas de frecuencia alpha o superiores presentes durante más de la mitad de la época
1	Somnolencia	Presencia de <i>Vertex waves</i> junto con frecuencia alpha.
2	Sueño ligero	Presencia de <i>K complexes</i> y <i>sleep spindles</i>
3	Sueño profundo	Actividad frecuencial en la banda delta, puede seguir dándose algún <i>K-complex</i> o <i>spindle</i>
4	Sueño muy profundo	La actividad en la banda delta se incrementa
REM	Sueño de movimientos oculares rápidos	Presencia de ondas de forma serrada (<i>saw-tooth waves</i>) de baja amplitud

Hoy en día, la detección de las fases del sueño es realizada mediante inspección visual de las señales biomédicas por médicos especialistas, acarreando un importante consumo de tiempo y recursos. Además, diferentes expertos pueden tener una tasa de acuerdo a la hora de clasificar las fases del sueño de menos del 85%, por lo que la subjetividad también es un inconveniente [16]. Estas limitaciones podrían reducirse con el desarrollo de métodos lo suficientemente efectivos para la automatización de la detección de las fases, que consigan una mayor objetividad y rapidez que los especialistas humanos a la hora de analizar un gran número de señales.

Diferentes propuestas de técnicas para la detección automática de las fases del sueño han sido publicadas y son conocidas como sistemas ASSC (*Automatic Sleep Stage Classification*). La mayoría siguen los siguientes pasos: preparación y preprocesado de las señales biomédicas, extracción y selección de las características más relevantes de dichas señales y clasificación en las diferentes fases del sueño, dependiendo de las características previamente seleccionadas. En los últimos años se han empezado a aplicar también técnicas de *Machine Learning* y *Deep Learning* para optimizar esta tarea, y esa es la línea que sigue el estudio actual.

1.4. HIPÓTESIS Y OBJETIVOS

En este TFG se parte de la hipótesis de que es posible detectar de manera automática las distintas fases del sueño mediante métodos basados en el procesado del EEG. Siguiendo la anterior hipótesis, el objetivo general de este estudio es desarrollar y evaluar nuevos algoritmos de detección automática de las fases del sueño basados en técnicas de *Deep learning*. Para ello se analizarán registros de EEG nocturno de adultos y niños, incluidos en dos bases de datos públicas provenientes de los estudios *Sleep Heart Health Study* (SHHS) y *Childhood Adenotonsillectomy Trial* (CHAT).

Una vez detallado el objetivo general, a continuación se indican los objetivos específicos del estudio:

1. Realizar una revisión del estado del arte sobre los algoritmos publicados en revistas científicas de detección automática de la fase del sueño.
2. Seleccionar el algoritmo más adecuado para la detección de fases del sueño e implementarlo.
3. Aplicar el algoritmo seleccionado a nuestras bases de datos y verificar su precisión en la detección de las diferentes fases del sueño.
4. Estudiar a fondo el algoritmo implementado para proponer posibles mejoras de los mismos.
5. Implementar las diferentes mejoras sugeridas y aplicarlas a las bases de datos disponibles para validar los nuevos algoritmos y verificar si se aumenta el nivel de precisión en la detección de las diferentes fases del sueño.
6. Extraer las conclusiones finales del estudio.

Para alcanzar el objetivo general y los diferentes objetivos específicos del proyecto, se ha seguido el siguiente plan de trabajo dividido en 3 fases principales:

Fase 1: revisión bibliográfica e implementación de los métodos seleccionados

- a) Búsqueda bibliográfica de los algoritmos publicados de detección automática de las fases del sueño, valorando sus ventajas e inconvenientes.
- b) Selección de uno de los algoritmos publicados para analizarlo y estudiarlo con mayor detalle.

Fase 2: aplicación del algoritmo y propuestas de mejora

- a) Aplicación del algoritmo seleccionado a las bases de datos disponibles para verificar su nivel de precisión en la detección de las diferentes fases del sueño.
- b) Propuestas de mejoras del algoritmo implementado y su re-programación.
- c) Aplicación del nuevo algoritmo propuesto a las bases de datos disponibles.

Fase 3: conclusiones e informe final

- a) Validación y verificación del nuevo algoritmo para comprobar si ha aumentado el rendimiento en la detección de las diferentes fases del sueño.
- b) Extracción de las principales conclusiones a partir de todas las pruebas realizadas.
- c) Redacción del informe final del estudio.

1.5. ESTRUCTURA DEL TFG

El Trabajo Fin de Grado se divide en 8 capítulos: introducción, detección automática de las fases del sueño, sujetos y señales, metodología, resultados, discusión, conclusiones y líneas futuras.

En este capítulo de introducción, en primer lugar se ha explicado brevemente en que consiste la Ingeniería Biomédica y las diferentes señales que en ella se estudian. De todas estas señales, se ha comentado con más profundidad la señal de EEG, pues es en la que se basa el resto del TFG. A continuación, se han detallado los conceptos principales relacionados con las fases del sueño y su estudio. Finalmente, se han descrito la hipótesis y los objetivos del trabajo, así como la estructura de lo que resta de documento.

El segundo capítulo se centra en la revisión del estado del arte en el contexto de la detección automática de las fases del sueño. Para ello, inicialmente se comentan los antecedentes de esta rama de estudio. A continuación, se detallan las distintas fases en las que se dividen la mayoría de los artículos revisados: obtención de datos, preprocesado, selección y extracción de características y clasificación. Una vez vistos los diferentes métodos existentes, se comparan atendiendo a diversos criterios y se elige uno como base para el resto del TFG. El siguiente capítulo se dedica a presentar las bases de datos que se usan en el TFG y a explicar sus principales características. Además, se comentan los pasos que se han llevado a cabo en el preprocesado de las señales y la organización de los pacientes en grupos de entrenamiento, validación y test.

En el capítulo 4 se comienza explicando la arquitectura *Deep learning* que se ha usado en el TFG, las redes neuronales convolucionales. En primer lugar, se detallan las características generales de este tipo de redes, para luego centrarse en el modelo concreto que se usa en este trabajo. También se explican en este capítulo los parámetros estadísticos que se han utilizado para evaluar el comportamiento de los modelos conseguidos. Para terminar, se incluyen los cambios que ha habido que introducir al intentar clasificar los registros pediátricos.

Los resultados obtenidos se exponen en el quinto capítulo y se discuten y analizan en el sexto. Tanto a la hora de mostrar los resultados como en la discusión, se diferencia entre los modelos entrenados solo con pacientes adultos, solo con niños, o los modelos en los que se prueba un entrenamiento conjunto. Además, se comparan los resultados de este trabajo con los de estudios publicados anteriormente con el fin de analizar los puntos en común y aquellos en los que se diferencian.

El TFG se concluye con un séptimo capítulo en el que se exponen las conclusiones alcanzadas junto con posibles líneas futuras de investigación. Seguidamente se incluye la bibliografía, donde se recogen todas las referencias consultadas durante el transcurso del Trabajo Fin de Grado.

Capítulo 2 DETECCIÓN AUTOMÁTICA DE LAS FASES DEL SUEÑO

En todo proyecto de investigación uno de los primeros pasos es realizar una extensa revisión del estado del arte, con el fin de comprender los avances que se han dado hasta la fecha en el área de investigación que corresponda. Por ello, tras revisar los diferentes artículos publicados en los últimos años, en este capítulo se presentan los conceptos principales relacionados con los métodos de detección de las fases del sueño. En primer lugar se realiza una pequeña introducción en la que se mencionan los antecedentes. A continuación, se pasa a explicar los cuatro bloques principales en los que se estructuran la mayoría de los artículos revisados (obtención de los datos, preprocesado de las señales, extracción/selección de características y clasificación). Por último, se detalla el proceso seguido al elegir uno de los estudios como base para el resto del TFG.

2.1. ANTECEDENTES

El desarrollo de la tecnología ha provocado que sea durante la última década cuando los métodos de detección automática de las fases del sueño propuestos han conseguido unas prestaciones suficientes para igualar y superar los resultados de la clasificación visual. Sin embargo, este tema lleva intrigando a la comunidad científica especializada desde hace mucho más tiempo. Así, ya en los años ochenta y noventa, se publicaron los primeros artículos al respecto [27]–[29], siendo precursores de este área de estudio del sueño que hoy tiene tanta importancia y en el que se basa este proyecto.

En un primer momento empezaron a surgir técnicas para detectar en la señal EEG las diferentes bandas de frecuencia y las formas de onda más comunes durante el sueño, sin aplicar aún *software* de apoyo, usando *hardware* analógico y digital como contadores, registros y comparadores. [5] La diferenciación de las fases del sueño era análoga a la que se realiza hoy en día, aunque se acoplaban las fases 1 y 2 en sueño ligero y las fases 3 y 4 en sueño profundo.

Algunos de los métodos inicialmente usados para la extracción de características eran algoritmos de detección de los pasos de la señal por cero y de detección de picos. También se aplicaban técnicas de lógica difusa (*fuzzy logic*) para representar mediante gráficos las principales características de las señales [30]. En 1989 se desarrolló uno de los primeros modelos para la detección automática de las fases del sueño: una máquina de estados, en la que cada estado representaba una fase del sueño y, dependiendo de la combinación de formas de onda que se detectase, se pasaba de un estado a otro. Las transiciones entre estados se realizaban aplicando teoría de la evidencia de Dempster-Shafer [27].

Por lo general, el número de registros usados para validar los diferentes métodos de detección ha ido incrementándose con los años. Esto se debe a las mejoras en el almacenamiento y procesado de grandes cantidades de datos, que han permitido poder crear bases de registros mucho más completas. Debido a este aumento de los datos disponibles y a los avances en los algoritmos usados, los métodos que han surgido en los últimos años son lo más fiables. Es por ello que, al realizar la revisión del estado del arte para este TFG, se han tenido en cuenta únicamente los estudios recientes.

La clasificación visual de las fases del sueño se puede realizar siguiendo diferentes reglas que se indican en diversos manuales técnicos. Las reglas de clasificación seguidas son un detalle importante, pues en muchos estudios los resultados varían dependiendo del manual que se siga al definir dichas reglas. Los dos manuales comúnmente utilizados ya se mencionaron en el capítulo anterior; el primero fue generado por R&K (Rechtchaffen y Kales) en 1968 [24] y el segundo por la AASM (American Academy of Sleep Medicine) en 2007 [4] con posteriores revisiones en 2012 [25] y 2017 [26]. R&K, el más usado durante años, tiene las desventajas de que da lugar a bastante subjetividad y de que se desarrolló pensando en adultos jóvenes y sanos, por lo que sus reglas no se adaptan bien a grupos de sujetos de más edad [24]. Las reglas publicadas por la AASM se diferencian de las anteriores en diversos detalles, como la retirada de la fase de movimiento (movement time) presente en R&K, la fusión de las fases 3 y 4 en una sola o que se añaden recomendaciones para la frecuencia de muestreo y el filtrado [4].

Los bloques principales en los que se estructuran la mayoría de los métodos de detección automática de las fases del sueño son: obtención de los datos, preprocesado de las señales, extracción/selección de características y clasificación. En los siguientes subapartados se revisarán las diferentes técnicas usadas para la realización de estos bloques.

2.2. OBTENCIÓN DE DATOS

El desarrollo de algoritmos de detección automática de fases del sueño requiere de un elevado número de señales biomédicas provenientes de diferentes sujetos. En muchos estudios, estas señales se consiguen utilizando bases de datos ya existentes. Cada base de datos indica el número de sujetos de los que vienen las señales, sus características (sexo, edad o posibilidad de enfermedades del sueño), la frecuencia de muestreo aplicada y la técnica de adquisición de las señales. En el caso del EEG se indican los canales utilizados. Entre las bases de datos más usadas en los últimos años se encuentran las siguientes: SleepEDF y SleepEDF expandida, MASS, DREAMS, ISRUC, SHHS (Sleep Heart Health Study) y la base de datos *Sleep Apnea Dataset* de las universidades de St. Vicents y Dublin (UCD) [7], [31]–[35]. No obstante, existen estudios en los que se registran y analizan sus propias señales.

El hecho de que no siempre se usen las mismas bases de datos ni el mismo número de señales provoca que la comparación entre métodos sea difícil, pues una misma técnica dará diferentes resultados dependiendo de si tiene que clasificar las fases del sueño basándose en registros de un tamaño u otro, o de que los sujetos padezcan o no enfermedades del sueño.

2.3. PREPROCESADO

Las señales biomédicas pueden contener diferentes artefactos que lleven a una mala interpretación de las mismas. Por ello, es necesaria una etapa inicial de preprocesado que los atenúe o elimine, dando así más importancia a la componente informativa de la señal. Estos artefactos son causados por diversos factores ajenos a la actividad cerebral que se desea medir, como el ruido generado por el instrumental de medida o las interacciones con otras señales biomédicas diferentes al EEG, por ejemplo movimientos musculares (EMG) o respiratorios [2].

Los métodos más utilizados al atenuar los efectos de artefactos son los siguientes [16].

- **Filtrado digital.** Se realiza para potenciar las bandas de frecuencia donde va la información de la señal EEG. La mayoría de métodos aplican un filtro Notch, que elimine la frecuencia de alimentación de 50 Hz (Europa) o 60 Hz (Estados Unidos), y un filtro paso banda, normalmente tipo *Butterworth*, entre las bandas delta y gamma.
- **Regresión.** Método que busca obtener un modelo que imite el comportamiento del ruido, para luego restárselo a la señal original y que dicho ruido quede atenuado.
- **Transformadas wavelet (*Wavelet Transform, WT*).** Se divide la señal en las diferentes sub-bandas frecuenciales, lo que ayuda a diferenciar la actividad cerebral real de los diferentes artefactos. Posteriormente se elimina la energía presente en las sub-bandas no deseadas, que se corresponderá con el ruido.
- **Separación ciega de fuentes (*Blind Source Separation, BSS*).** Que busca separar la señal original del ruido, basándose en que cada una proviene de una fuente.

La normalización y segmentación de la señal de EEG en varias épocas es otro paso común en el preprocesado de la mayoría de los métodos, consiguiendo así un conjunto de señales mucho más manejables con las que trabajar. En [4] se recomienda asignar a cada época una duración de 30 segundos. A mayores, en muchas ocasiones se aplica un *oversampling* de la señal, con el fin de que el número de muestras sea similar para todas las fases del sueño, y que no haya alguna con mucha menos presencia. Algunas de las técnicas usadas para realizar esta tarea son SMOTE (*Synthetic Minority Over-sampling Technique*) [36] o ADASYN (*adaptive synthetic sampling approach*) [37].

2.4. SELECCIÓN DE CARACTERÍSTICAS

La actividad cerebral conseguida durante una fase del sueño concreta presenta una serie de características que definen a dicha fase y la diferencian del resto. Estas características hacen posible que los métodos automáticos puedan diferenciar entre las épocas de las señales que pertenecen a unas fases u otras. La elección de un número óptimo de características independientes entre sí es un aspecto de gran importancia que afecta en gran medida al funcionamiento de los algoritmos. Un conjunto demasiado bajo podría no ser suficiente para llevar a cabo una correcta clasificación. Sin embargo, demasiadas características muy

dependientes entre sí podrían llevar a utilizar información redundante, que incluso disminuiría la capacidad de detección del método correspondiente.

Para conseguir estas características diferenciamos cuatro tipos de análisis: temporal, frecuencial, combinación temporal-frecuencial y no lineal. El último es en realidad un subtipo de análisis temporal, pero se diferencia del resto porque aplica métodos no lineales. A continuación se va a realizar un breve resumen de estos cuatro tipos de análisis.

Análisis temporal. Es el método más sencillo, claro, intuitivo y el primero que se usó. Se basa en fijarse en características morfológicas de la señal EEG como la amplitud, la duración, los pasos por cero o histogramas, entre otras. También es frecuente trabajar con parámetros estadísticos (variancia, desviación estándar, media, mediana o kurtosis). Los parámetros de Hjorth, *Activity* (Ha), *Mobility* (Hm) y *Complexity* (Hc) también entran en este tipo de análisis [16].

Análisis frecuencial. El EEG contiene mucha información en el dominio de la frecuencia que, con un análisis que se limite únicamente al dominio temporal, no se tendría en cuenta. Esto se soluciona implementando también un análisis frecuencial, que se basa en estimar la potencia espectral de la señal en las diferentes bandas de frecuencia, usando herramientas como la FFT o la STFT [16]. En este tipo de análisis se incluyen también espectros de orden superior, (*HOS High Order Spectra*), que tienen en cuenta no sólo la amplitud de la señal sino también su fase [38].

Análisis temporal – frecuencial. Las características frecuenciales del electroencefalograma no permanecen siempre constantes, van cambiando con el tiempo, por lo que un análisis que tenga en cuenta los dos dominios es de gran utilidad. Algunas técnicas usadas para llevar a cabo este análisis son *Wigner-Ville Distribution* (WVD), o *Hilbert Huang Transform* (HHT), que muestran la distribución de la energía de la señal a lo largo del tiempo. También es común el uso de *wavelets* para este tipo de análisis [16].

Análisis no lineal. La naturaleza caótica, no lineal y no estacionaria del EEG provoca que las técnicas anteriores sean insuficientes para una buena caracterización de la señal, y que haga falta prestar atención también a parámetros no lineales. Los métodos no lineales se pueden agrupar en dos grandes grupos, en primer lugar, los basados en la entropía y complejidad de la señal, que muestran los niveles de impureza e irregularidad y engloban, entre otras, la entropía de Reyi y la de Tallis. En segundo lugar los métodos basados en las características fractales, que muestran el comportamiento aleatorio de la señal y buscan similitudes, en este grupo se encuentran los exponentes de Lypanov y de Hurst [16].

Las dimensiones del conjunto de características conseguido se reducen mediante diferentes algoritmos, como PCA (*Principal Component Analysis*), SFS o SBS (*Sequential forward/backward selection*) [16]. De esta manera se consiguen subconjuntos óptimos para realizar la tarea de detección de fases del sueño.

2.5. CLASIFICACIÓN

En esta etapa se identifican, por medio de algoritmos automáticos, propiedades y comportamientos subyacentes de las señales que permitan clasificarlas en las distintas fases del sueño. Para ello, en la mayoría de métodos se aplican algoritmos de aprendizaje automático (*machine learning*).

En la última década, el aprendizaje automático ha permitido la aparición de vehículos autónomos, dispositivos que reconocen la voz o algoritmos efectivos para búsquedas y recomendaciones online, entre otros [39]. Dentro del campo del *machine learning* hay diversas técnicas disponibles, y muchas de ellas se pueden aplicar a la detección automática de las fases del sueño. A continuación, se hace un breve resumen de las técnicas que mejores resultados han dado en los últimos años, aunque en posteriores capítulos de este TFG se va a estudiar más a fondo solo una de ellas, las redes neuronales convolucionales.

2.5.1. K-MEANS

Método de clasificación que aplica *machine learning* para dividir el conjunto de observaciones en varios grupos o *clusters*, el número de grupos viene dado por el parámetro k [40], [41]. Inicialmente, el centro de cada uno de los k grupos se selecciona de manera aleatoria. Una vez inicializados los centros, cada observación de los datos se asigna a un grupo u otro dependiendo de la cercanía espacial, usando como métrica más común la distancia euclídea [40]. A continuación, se calcula la media de todas las muestras asignadas a un mismo grupo, llegando a k nuevos valores que pasarán a ser los nuevos centros de los respectivos grupos. Este proceso de establecer los grupos, clasificar en función de la distancia las observaciones y actualizar en función de esta clasificación los organización de los grupos, se repite hasta que el valor de los centros de cada grupo se estabiliza, momento en el que el algoritmo considera que la clasificación de los datos es óptima [40].

Este método propone una técnica de clasificación simple que minimiza la distancia entre las observaciones y los centros de los grupos. Sin embargo, presenta la desventaja de que no hay garantías de que el proceso se estabilice en el mínimo global, y no en un mínimo local, lo que implicaría que hay otras opciones de clasificación mejores. Además, la organización final de los grupos depende en gran medida de los valores de inicialización de los centros, por ello han surgido técnicas como *k-means ++*, que buscan optimizar esta inicialización [40].

En [42] y [43] Diykh et al aplican el algoritmo de clasificación *k-means* en el área de la detección de las fases del sueño, utilizando para ello la señal de EEG y diferenciando entre las 6 fases del sueño. Los estudios difieren en cómo se consigue a partir de la señal inicial el conjunto de datos a clasificar: en [42] aplican la teoría de grafos para extraer en el dominio del tiempo las características estadísticas del EEG, en [43] se sigue la misma técnica, pero añadiendo una etapa a mayores en la que se pasan los datos a formato complejo.

2.5.2. LDA (LINEAR DISCRIMINANT ANALYSIS)

Modelo de clasificación que trabaja con regiones de decisión, cuyas fronteras se definen mediante el uso de funciones lineales, y que busca minimizar la varianza entre los elementos pertenecientes a cada una de las regiones. Toma como entradas vectores de características, la dimensión de estos vectores y el cómo se generan a partir de los datos iniciales dependen de la técnica utilizada en cada estudio [44]. Este tipo de clasificador asume una distribución normal de los datos, lo que, en caso de que no exista esta normalidad, puede ser una desventaja. En estudios como [45] y [46] se usa esta técnica para intentar encontrar un buen método de detección de las fases del sueño. A veces se sigue el clasificador LDA de una época de *smoothing*, con el propósito de tener en cuenta las interacciones con las secuencias temporales cercanas.

2.5.3. SVM (SUPPORT VECTOR MACHINES)

Algoritmo de aprendizaje supervisado que aplica *machine learning* para encontrar vectores de soporte que definan fronteras que separen, de manera óptima, diferentes conjuntos de datos. Presenta la ventaja de una buena generalización, lo que implica que si se aplican nuevos conjuntos de datos al algoritmo estos van a ser correctamente clasificados [44]. En muchas ocasiones se utiliza junto a funciones *kernel*, que se encargan de transformar el conjunto de datos de entrada en un conjunto de características con las dimensiones y propiedades adecuadas para el algoritmo de clasificación [44]. Entre las distintas opciones de funciones *kernel* se distinguen funciones lineales, polinómicas, funciones de base radial (*Radial Basis Function*, RBF) o la función *sigmoid* [44], [47]. Inicialmente este algoritmo se diseñó pensando en problemas de clasificación binaria, pero se puede extender a problemas multiclase aplicando métodos como *One-Against-ALL* (OAA) o *One-Against-One* (OAO) [48].

El uso de SVM, o variantes, en estudios de detección automática de las fases del sueño es muy amplio. En los artículos [47] y [48] se utiliza LIBSVM (*A Library for SVM*) [49], que es un *software* especial que permite aplicar este tipo de algoritmos a problemas de clasificación multiclase usando el método de *One-Against-One*; además, ambos aplican la función RBF para transformar el conjunto de características no lineales extraído del EEG a un conjunto lineal, antes de pasárselo como entrada al algoritmo. Los estudios de [50] y [51] también se basan en algoritmos SVM al clasificar los datos y, a mayores, añaden una etapa de ajuste (también denominada *smoothing*) para tener en cuenta la relación con épocas anteriores y posteriores. Este ajuste se realiza de manera que si en tres épocas consecutivas del EEG la primera y la tercera se clasifican como una misma fase, se considera que la época intermedia pertenecerá también a esa fase y, en caso de que se haya clasificado en alguna otra, se modifica.

2.5.4. ANN (ARTIFICIAL NEURAL NETWORKS) CLÁSICAS

Modelos de procesamiento de información inspirados en las interacciones neuronales. Consisten en varios procesadores interconectados, formando una red, que combinan sus resultados con ciertos pesos para ir pasando la información de capa a capa. Dentro de este tipo de algoritmo existen muchas variantes, que permiten dar a la red unas funcionalidades u otras. Por ejemplo,

en [52] se consigue una buena clasificación de las fases del sueño aplicando un subtipo de estas redes neuronales, CVANN (*Complex-valued ANN*).

2.5.5. ENSEMBLE LEARNING

Muchos métodos aplican la idea de que la combinación de varios clasificadores simples y poco eficientes, bien entrenados y con valores de parámetros adecuados, puede dar lugar a métodos globales de clasificación con muy buenas prestaciones. Sin embargo, tienen la desventaja de que son bastante difíciles de analizar, pues al componerse de un elevado número de modelos individuales no es intuitivo el distinguir que factores contribuyen a mejorar o empeorar las decisiones de clasificación. Los métodos de *ensemble learning* generan, a partir del conjunto de entrenamiento, los diferentes modelos de clasificación cuyas decisiones luego se combinan para dar lugar a la clasificación global. Dado que en muchas ocasiones el conjunto de entrenamiento no es lo suficientemente extenso como para dividirlo en tantos sub-conjuntos diferentes como modelos a entrenar, se recurre a técnicas como el sobre-muestreo, que permitan generar nuevos datos a partir de los que se tienen [40].

Dentro de estos métodos se diferencian principalmente tres subgrupos de algoritmos, *bagging*, *boosting* y *stacking*. En los métodos de ***bagging*** (*bootstrap aggregating*) se entrenan los modelos individuales de manera independiente (usando subgrupos de los datos de entrenamientos generados a partir de los originales replicando algunas de las muestras y eliminando otras) [40]. La decisión global de a qué fase del sueño pertenece cada época del EEG se equivale con aquella fase que más ‘votos’ haya recibido en las clasificaciones de los modelos individuales. Entre los estudios que aplican *bagging* en la detección automática de las fases del sueño destaca [7] por los buenos resultados obtenidos.

Un método de *bagging* muy presente en la literatura de detección de las fases del sueño es el método de *Random Forest*. Este clasificador consiste en múltiples *decision trees* (árboles de decisión) y va eligiendo como resultado la predicción más votada por los diferentes árboles. El número de *decision trees* que se aplica no es fijo, sino que se busca la opción óptima dependiendo del problema en cuestión. En el estado del arte se encuentran varios estudios que aplican esta técnica; en [53] se decide que el número óptimo de *trees* sea 35, mientras que [54] fija el número óptimo en 400 y en [55] se escogen 120. Para tener en cuenta las relaciones temporales entre secuencias, muchos métodos combinan el uso de RF con modelos como HMM (*Hidden Markov Model*) [56].

La idea detrás de los métodos de ***boosting*** es similar a la de *bagging*, pues en ambas se combinan varios modelos y se toma como decisión final la clasificación que más se repite entre los modelos individuales. Sin embargo, en *boosting*, al contrario que en *bagging*, los diferentes modelos no se entrenan por separado, sino que el comportamiento de un modelo influye en la generación del modelo siguiente [40]. Con esto se pretende conseguir que la contribución de cada modelo a la clasificación global complemente al resto, dando más importancia a los detalles en los que el resto de modelos fallan. Entre los estudios que aplican *boosting* para la detección de las fases del sueño encontramos los siguientes: en [35] se usa AdaBoost (*adaptive boosting*), subtipo de *boosting* diseñado especialmente para problemas de

clasificación, y [57] aplica RUSBoost (*Random under sampling boosting*) para solucionar el problema de tener en los datos demasiada presencia de algunas fases y demasiado poca de otras.

Por último, los métodos de **stacking** difieren de los anteriores en que en vez de decidir la clasificación final por un proceso de ‘votación’ equitativo entre las clasificaciones de los modelos individuales, se introduce una etapa extra con un algoritmo que se encarga de combinar las clasificaciones individuales y dar una global [40]. De esta manera se pretende aprender cuáles de los modelos individuales son los más fiables, y darles un mayor peso en la decisión final. Este tipo de métodos se suelen usar cuando los modelos que se combinan siguen diferentes algoritmos de machine learning, pues en esos casos es muy probable que unos se comporten mejor que otros [40].

2.5.6. LSTM (LONG SHORT-TERM MEMORY)

Un subtipo de ANN que se ha utilizado mucho en los últimos años para todo tipo de problemas. Este método usa una arquitectura RNN (*recurrent neural network*) que se adapta mejor al procesado de secuencias de datos, pues se tienen en cuenta las dependencias con datos anteriores a la hora de dar un resultado. Esto es útil en los casos en los que se trabaja con señales dependientes del tiempo, como el EEG. En el ámbito de las fases del sueño se encuentran estudios como [58], que realiza la clasificación utilizando dos LSTM en cascada, o [59], que aplica la variante BLSTM, que también permite tener en cuenta la información de épocas futuras.

2.5.7. CNN (CONVOLUTIONAL NEURAL NETWORK)

Las redes neuronales convolucionales son otro subtipo de ANN que se ha empezado a utilizar en alguno de los últimos estudios publicados sobre detección automática de las fases del sueño. Dado que éste es el método en el que se va a profundizar en el resto del TFG, se reserva el detallado de sus características para los capítulos posteriores. Algunos de los estudios que usan estas redes para resolver el problema de detección de las fases del sueño son [60] y [61]. En [62] se aplica una OCNN (CNN ortogonal) para optimizar la extracción de características y [63] combina una CNN con una RNN para tener en cuenta las transiciones entre secuencias temporales.

2.6. COMPARACIÓN Y ELECCIÓN DEL MÉTODO A IMPLEMENTAR

El primer paso realizado en el TFG fue revisar el estado del arte en relación a la detección automática de las fases del sueño. La información obtenida a lo largo de esta revisión es la que se ha resumido hasta ahora en este capítulo 2. Una vez entendido todo lo que se había realizado hasta la fecha, el siguiente paso fue elegir un estudio en concreto que tomar como base para las siguientes etapas del Trabajo Fin de Grado.

El método que será implementado en futuras fases de este estudio es seleccionado del conjunto de técnicas revisadas en el apartado anterior. Atendiendo a los resultados, en un primer momento prestando especial atención a la precisión y al coeficiente Kappa (*Accuracy* y *Kappa coeficient*), se realiza una selección de 12 métodos que presentan valores bastante altos para ambos parámetros (*Tabla 2.1*).

La precisión mide el ratio del número de épocas correctamente clasificadas respecto al número total de épocas, mientras que el coeficiente Kappa mide la concordancia entre la clasificación hecha por el algoritmo y la misma clasificación realizada por un experto de manera visual. Se representan en tanto por ciento y tanto por uno, respectivamente, con valores más altos implicando una mejor clasificación.

Tabla 2.1: Métodos inicialmente seleccionados como posibles estudios que tomar como base para el resto del TFG.

Autor	Clasificación	Acc	Kappa	Base de datos	Citas
Musa Peker [52]. 2015	CANN	92%	0.91	8 adultos Sleep-EDF	23
Diykh et al. [42]. 2016	K-means	95%	0.80	16 adultos Sleep-EDF y Sleep Spindles	46
Hassan et al. [7]. 2017	Bagging	93.6% (78%)	0.85 (0.82)	8 adultos (20 adultos) Sleep-EDF (DREAMS)	59
Hassan et al. [35]. 2017	AdaBoost	91.3% (81%)	0.86 (0.81)	8 adultos (25 adultos) Sleep-EDF (UCD)	40
Sors et al. [61]. 2018	CNN	87%	0.81	5793 adultos SHHS	49
Seifpour et al. [48]. 2018	LIBSVM	91% (83%)	0.87 (0.77)	8 adultos (20 adultos) Sleep-EDF (DREAMS)	11
Ghasemzadeh et al. [51]. 2019	SVM	93% (82%)	0.87 (0.77)	13 adultos (10 adultos) Sleep-EDF expandida (ISRUC)	2
Mousavi et al. [60]. 2019	CNN	92.9%	0.86	8 adultos Sleep-EDF	1
Ghimatgar et al. [56]. 2019	RF y HMM	78%	0.69	28 adultos Sleep-EDF y Sleep-EDF expandida,	0
Jiang et al. [55]. 2019	RF y HMM	89%	0.83	28 adultos Sleep-EDF y expandida	6
Liang et al. [45] 2012	LDA	88%	0.81	20 adultos Siesta 802 PSG (Compumedics, Inc.)	207
Alickovic et al. [64]. 2018	RotSVM	91%	0.88	20 adultos Sleep-EDF expandida	22

Los dos parámetros previamente explicados (precisión y coeficiente Kappa) pueden no mostrar de manera certera el funcionamiento de los métodos. Por ejemplo, puede que algún algoritmo clasifique muy mal alguna de las fases, pero si el número de épocas correspondientes con esa fase dentro de los datos es muy bajo, la precisión seguirá saliendo muy alta. A causa de esto, se ha decidido considerar también en la comparación la cantidad de datos que usa cada uno de los 12 métodos y el número de citas asociadas a los estudios correspondientes, pues permiten ver el impacto científico. Toda la información se resume en la *Tabla 2.1*. En aquellos estudios que evalúan los métodos en dos bases de datos, se incluyen los resultados de la segunda evaluación entre paréntesis, para diferenciarlos de los de la primera.

En la tabla se pueden comparar las características de los 12 métodos preseleccionados. La diferencia entre el número de citas solo es relevante si se consideran artículos no muy recientes, ya que aquellos publicados en el último año no han tenido tiempo de ser muy referenciados. A continuación se realiza un breve repaso de cada uno de los estudios considerados, para finalizar detallando el que se ha decidido tomar como base para el resto del trabajo y las razones detrás de dicha decisión.

En el estudio de Musa Peker [52] se examina la influencia de usar clasificadores que trabajen con datos de formato complejo en la detección automática de las fases del sueño, en concreto redes neuronales de valores complejos (*complex-valued artificial neural network*). La razón que les lleva a usar este tipo de clasificadores reside en las ventajas que ofrece trabajar con números complejos respecto a tener en cuenta únicamente números reales, sobretodo en ámbitos como el procesado de señales. Para ello en primer lugar se seleccionan 9 características no lineales de la señal EEG y posteriormente, antes de pasárselas a la red neuronal como entradas, se transforman a datos complejos mediante un método de codificación de fase. Este estudio consigue un rendimiento de clasificación bastante alto, y la idea de trabajar con datos complejos no se ha visto en muchos artículos dentro de la literatura en cuestión; sin embargo, utiliza una base de datos muy reducida a la hora de comprobar sus hipótesis.

Diykh et al [42] desarrollan su método de detección automática de las fases del sueño partiendo de una selección de características temporales de la señal EEG. Para ello, dividen cada época de 30 segundos de la señal en 75 sub-segmentos, de manera que cada uno de ellos se pueda considerar como una señal estacionaria. El número óptimo de subsegmentos en los que se divide cada época se ha calculado de manera empírica durante el experimento. A continuación, aplican la teoría de grafos al conjunto de características extraído, las propiedades gráficas resultantes se clasifican en las distintas fases del sueño mediante el algoritmo *K-means*, anteriormente explicado. Aunque los resultados de este estudio muestran que se consigue una clasificación bastante precisa, se siguen manteniendo el problema de utilización de una base de datos reducida visto en el estudio anterior.

En [7] Hassan et al aplican un método de *ensemble learning*, *bagging* en concreto, para llevar a cabo la clasificación de las fases del sueño. En primer lugar, cada época de la señal EEG se divide en sub-bandas usando un tipo de transformada wavelet denominado *tunable-Q wavelet transform* (TQWT) que permite ir variando los parámetros para adaptarse a los datos, ganando así flexibilidad; de cada una de las sub-bandas se extraen 4 momentos estadísticos. A

continuación, se implementa una etapa de selección de características utilizando Kruskal-Wallis *one-way analysis of variance* (Kruskal-Wallis ANOVA) para asegurar que las características elegidas difieren de manera significativa entre las distintas fases del sueño, eliminando así aquellas características que no cumplan el criterio de varianza. Una vez conseguido un conjunto de características óptimo estas se clasifican usando un método de *bagging*, explicados en la sección anterior, optando por usar árboles de decisión como modelos individuales a combinar. El modelo propuesto se evalúa al clasificar de 2 a 6 fases del sueño, los resultados muestran cómo el rendimiento de clasificación aumenta al distinguir menos fases, en la tabla se han incluido los valores correspondientes con la detección de 5 fases distintas. Además de obtener un alto rendimiento de clasificación, en la tabla se puede ver que el impacto científico de este estudio es bastante grande, ha sido citado en 59 ocasiones. Sin embargo, el número de sujetos que forman las bases utilizadas es algo reducido.

En el estudio [35] de Hassan et al también se aplica una técnica de *ensemble learning*, en concreto un tipo de método *boosting* denominado *AdaBoost*, que propone que a medida que se avanza en la clasificación los nuevos modelos entrenados se vayan centrando en las entradas más difíciles de predecir. En primer lugar, se descompone la señal EEG en sub-bandas usando, al igual que en el estudio anterior, la TQWT. A continuación, cada una de las sub-bandas se modela mediante la función de densidad de probabilidad *Normal Inverse Gaussian* (NIG pdf), y los parámetros de dicho modelo son los que forman el conjunto de características a clasificar. Este conjunto pasa, también de la misma forma que en el estudio anterior, por una etapa en la que se aplica Kruskal-Wallis ANOVA para quedarse con aquellas que permiten una buena discriminación entre fases. Una vez conseguido el conjunto de características óptimo, éste se pasa como entrada a la etapa de *AdaBoost*, en la que se opta por usar árboles de decisión como clasificadores individuales.

El estudio [61], publicado por Sors et al en 2018, busca solucionar mediante el uso de redes neuronales convolucionales el problema de la clasificación de señales EEG de un solo canal en las distintas fases del sueño. Al usar un algoritmo *Deep learning* este método no necesita fases específicas para la extracción y selección de características, sino que se pasa como entrada a la red neuronal las épocas del EEG preprocesadas. Los resultados obtenidos muestran un rendimiento de clasificación elevado y en este caso la base de datos usada en los experimentos sí que es bastante extensa. En [60] Mousavi et al también deciden usar una CNN para llevar a cabo la clasificación, consiguiendo altos valores tanto de precisión como de coeficiente Kappa. Sin embargo, la cantidad de sujetos con los que desarrollan el método es muy baja, únicamente 8.

En [48] Seifpour et al clasifican el EEG de un solo canal en las distintas fases del sueño utilizando, para extraer el conjunto de características, una nueva técnica de análisis simbólico denominada *Statistical Behavior of Local Extrema* (SBLE). Esta técnica busca detectar las fases del sueño basándose en cambios morfológicos del EEG, de manera análoga a los criterios que usan los expertos de sueño al realizar la clasificación visual. Para ello, en primer lugar descompone cada época del EEG en 6 bandas frecuenciales y luego se extraen de forma independiente las características de cada una de las bandas. Para seleccionar las características óptimas se aplica el método de selección *MultiClass/Cluster Feature Selection* (MCFS). A continuación, la clasificación se lleva a cabo mediante un algoritmo SVM, usando el *software*

LIBSVM ya mencionado en el apartado anterior. El rendimiento de clasificación conseguido en este estudio es bastante elevado, aunque baja algo en la clasificación de sujetos con sospecha de trastornos de sueño (DREAMS) respecto al caso de sujetos sanos (Sleep EDF). Además, realizan pruebas en las que distinguen desde 2 hasta 6 fases del sueño, aunque en la tabla únicamente se incluyen los resultados correspondientes con 5 fases. Sin embargo, al igual que en varios de los estudios previamente analizados, el número de sujetos que forman las bases de datos que utilizan no es muy elevado.

Otro estudio en el que también se recurre a SVM para la clasificación es en el de Ghasemzadeh et al [51]; también utilizan otras técnicas de clasificación, como métodos de *ensemble*, pero los mejores resultados los consiguen aplicando SVM. En primer lugar, cada época de la señal EEG se descompone en 14 sub-bandas de tiempo-frecuencia usando una técnica basada en la combinación de dos tipos de transformadas *wavelets* denominada *double-density dual-tree discrete wavelet transform* (D3TDWT). A continuación cada una de estas sub-bandas se modela usando una variante a los modelos auto regresivos denominada *logistic smooth transition autoregressive* (LSTAR), que busca adaptar los modelos auto regresivos al caso de señales no lineales, pues éste es el caso del EEG. Se propone usar los pesos del modelo LSTAR como el conjunto de características a partir del cual clasificar las fases del sueño. Dicho conjunto, tras pasar por una etapa en la que se reduce su dimensión, se usa como entrada de la etapa de clasificación. Al aplicar SVM prueban usando diferentes funciones *kernel*, y concluyen que los mejores resultados se obtienen al usar una función cuadrática. Finalmente se aplica una etapa de ajuste o *smoothing* para intentar arreglar posibles errores de clasificación.

Ghimatgar et al proponen en [56] un método de detección automática de las fases del sueño basado en el uso de *random forest*, que ya se ha visto previamente que es un algoritmo de *ensemble learning* tipo *bagging*. En primer lugar, se extraen 136 características de cada época del EEG. Del total de características extraídas se consigue un conjunto óptimo mediante un método de selección denominado MGACO (*Modified Graph Clustering Ant Colony Optimization*) que busca mantener las características más relevantes y eliminar redundancias. A continuación, dicho conjunto se usa como entrada del clasificador *random forest*. Tras la clasificación, se incluye una etapa en la que se aplica *Hidden Markov Model* (HMM) para tener en cuenta las transiciones temporales entre fases del sueño y así reducir el número de predicciones erróneas. Un aspecto de este estudio que marca la diferencia respecto al resto de estudios revisados es que, a la hora de evaluar el método, prueban a usar bases de datos diferentes en el grupo de entrenamiento y en el de test, para así asegurar una buena generalización.

Siguiendo con la técnica de aplicar algoritmos *random forest* seguidos de una etapa de refinamiento HMM se encuentra el estudio [55] de Jiang et al. La principal diferencia entre este estudio y el anterior reside en la etapa de extracción y selección de características, pues aquí las características se consiguen mediante descomposición multimodal de las épocas EEG. Multimodal ya que las señales se analizan tanto en el dominio de la frecuencia, descomponiendo en las distintas sub-bandas, como en el dominio del tiempo, utilizando en este caso la técnica de *Empirical Mode Decomposition* (EMD) para descomponer las épocas EEG originales en funciones de tiempo más simples de las que posteriormente extraer las características.

En el estudio [45] de Liang et al la detección de las fases del sueño se lleva a cabo usando un algoritmo LDA. La extracción de características se realiza mediante dos técnicas: la primera de ellas consiste en analizar la señal EEG aplicando *multiscale entropy* (MSE), que se basa en calcular la entropía (concretamente la entropía muestral o *Sample Entropy*) en distintas escalas de tiempo; la segunda consiste en aplicar modelos autor regresivos a la sub-banda theta de las épocas EEG. Así se consigue para cada época un conjunto de 21 características en función de las cuales el algoritmo LDA clasifica entre 5 posibles fases del sueño. Finalmente se aplica también una etapa de *smoothing* para, teniendo en cuenta las transiciones temporales, corregir posibles errores. En la *Tabla 2.1* se puede apreciar que, entre todos los estudios en ella mostrados, éste es el que mayor impacto científico ha tenido pues ha sido citado más veces. Esto en parte se debe a que es el estudio más antiguo entre los 12 que se comparan.

Por último, en [64] Alickovic et al proponen usar un método de *bagging* denominado *rotational support vector machine* (RotSVM) para la detección de las fases del sueño. En este método de *bagging* los modelos individuales, que se entrenan de manera independiente con diferentes subconjuntos de los datos, siguen el algoritmo SVM. De esta manera se busca fusionar las ventajas de las técnicas de *ensemble learning* y de los clasificadores SVM. En el pre-procesado de las señales se utiliza la técnica de MSPCA (*Multiscale principal component analysis*) para mejorar la relación señal a ruido. A continuación, en la etapa de selección de características se aplica la transformada wavelet discreta (DWT) para dividir cada época de la señal EEG en sub-bandas. El conjunto características obtenido se usa como entrada al clasificador RotSVM.

Tras haber detallado brevemente cada estudio, se concluye este capítulo 2 seleccionando uno de ellos. En concreto, se elige usar como base del TFG el artículo de Sors [61] en el que se consiguen altos resultados aplicando redes convolucionales; ya que se considera que éste es un método bastante moderno que está resultando muy eficiente en muchas áreas de estudio. Además, de entre todos los estudios revisados, es el que usa una base de datos más extensa, un punto a favor a la hora de alcanzar modelos generalizables.

Capítulo 3 : SUJETOS Y SEÑALES

En este capítulo se presentan las dos bases de datos utilizadas a lo largo del trabajo, SHHS y CHAT, describiendo para ambas tanto las características de la población bajo estudio como los detalles de las señales EEG que contienen. Además, también se especifican las diferentes etapas de preprocesado necesarias para cada una. En ambas bases de datos los ficheros de registros siguen el formato europeo (EDF, *European Data Format*) y los ficheros de anotaciones el formato XML (*Extensible Markup Language*). Una vez descritas las bases de datos utilizadas se detallan los criterios seguidos a la hora de dividir sus registros en los grupos de entrenamiento, validación y test.

3.1. SHHS (*SLEEP HEART HEALTH STUDY*)

3.1.1. POBLACIÓN BAJO ESTUDIO

El SHHS (*Sleep Heart Health Study*) [65] es un estudio iniciado por el NHLBI (*American National Heart Lung and Blood Institute*) para determinar si el síndrome de la apnea-hipopnea del sueño (SAHS) está relacionado con un mayor riesgo de padecer diversas enfermedades cardiovasculares. Los sujetos que participaron en el estudio tenían que cumplir los siguientes requisitos: 40 años de edad o mayores, sin historial de tratamientos previos para el SAHS, sin traqueotomías y sin terapia de oxígeno. En el estudio se incluyen dos grupos de registros grabados en diferentes rondas mediante polisomnografía nocturna, los grupos SHHS-1 y SHHS-2. El primero se grabó entre Noviembre de 1995 y Enero de 1998 e inicialmente participaron 6441 individuos, aunque este número se acabó rebajando a 5804 para adecuarse a la legislación sobre protección de datos vigente. El segundo se consiguió tras realizar una nueva ronda de polisomnografías a 3295 de los individuos anteriores entre Enero de 2001 y Junio de 2003. Las polisomnografías se realizaron en las propias casas de los sujetos bajo las atenciones de técnicos especializados.

En este trabajo se va a utilizar una base de datos de SHHS-1 en la que están disponibles los registros de 5793 sujetos. Se utiliza esta base porque es la misma que se usó en el estudio utilizado como principal referencia del TFG y porque es la más grande encontrada en la literatura científica relacionada con la clasificación de fases del sueño [61]. En la *Tabla 3.1* se muestran las características de los sujetos que componen la base de datos utilizada, a la hora de decidir si uno de los pacientes adultos tenía sobrepeso se ha seguido el criterio de considerar con sobrepeso siempre que el BMI sea mayor que 30.

Tabla 3.1: Características de los 5793 pacientes que forman el grupo SHHS-1 en la base de datos SHHS.

Características	
Edad, media (SD)	63.14 (11.23)
Nº de mujeres (%)	3033 (52.35%)
Sobrepeso (BMI > 30)	1725 (29.78%)
Raza	
Blanca (%)	4898 (84.55%)
Negra (%)	515 (8.89%)
Otra (%)	380 (6.56%)

3.1.2. CARACTERÍSTICAS DE LAS SEÑALES EEG

Las señales de EEG se grabaron usando dos canales bipolares, $C_4 - A_1$ y $C_3 - A_2$, que como se ha visto antes se consiguen colocando el electrodo de referencia en el lóbulo de la oreja y el electrodo que capta la señal en la zona central de la cabeza (oreja derecha y zona central izquierda en el caso de $C_3 - A_2$ y lo contrario para $C_4 - A_1$). Entre ambos se ha seleccionado el canal $C_4 - A_1$ debido a que es el que se recomienda el manual AASM [5]. En el caso de SHHS-1, para todos los registros se usaron 125Hz como frecuencia de muestreo. Los registros se dividieron en épocas de 30 segundos y cada una de estas épocas se asoció con una etiqueta correspondiente con cada una de las distintas fases del sueño: Despierto (W), fase 1 (S1), fase 2 (S2), fase 3 (S3), fase 4 (S4) y sueño REM. Esta clasificación se realizó por expertos siguiendo las reglas de Rechtschaffen y Kales [24]. El único filtrado que se llevó a cabo en la adquisición de las señales fue un filtrado paso alto con frecuencia de corte 0.15Hz [34].

Como el clasificador que se va a usar es una CNN, incluido en el paradigma *Deep learning*, no se necesita realizar un preprocesado intenso de las señales extraídas de los EDFs, ya que la red neuronal debería ser capaz de clasificar las fases del sueño a partir de las señales originales o 'en crudo'. A *priori*, únicamente se realizan tres etapas de preprocesado, las mismas que se realizaban en [61]. En la primera se eliminan épocas clasificadas como *wake* (despierto) del inicio y el final del registro en los casos en los que los sujetos permaneciesen un largo rato despiertos antes de dormirse o se despertasen durante un tiempo significativo antes de finalizar la grabación. En la segunda etapa de preprocesado se fusionan las fases de sueño 3 y 4 en una única fase 3&4 para adaptar las reglas de Rechtschaffen y Kales a las actualizaciones de la AASM. De esta manera, la tarea de clasificación se centrará en diferenciar entre 5 fases del sueño. En la tercera etapa de preprocesado se descartan los pacientes en cuyos registros no haya ninguna etapa para alguna de las 5 fases a clasificar, porque se considera que puedan ser *outliers* que estropeen el modelo entrenado.

3.2. CHAT (*CHILDHOOD ADENOTONSILLECTOMY TRIAL*)

3.2.1. POBLACIÓN BAJO ESTUDIO

La base de datos CHAT (*Childhood Adenotonsillectomy Trial*) se inició en la Universidad de Pensilvania como parte de un estudio diseñado con el fin de determinar las diferencias que existen entre tratar la apnea del sueño obstructiva en niños mediante una operación de adenotonsilectomía (eAT) o simplemente vigilar la evolución de la enfermedad sin ningún tipo de cirugía (*watchful waiting plus supportive care, WWSC*) [66]. La apnea del sueño se trata de un trastorno en el que se obstruyen las vías respiratorias repetidamente durante el sueño del paciente. Mediante la eAT, se busca liberar esas vías para facilitar el flujo del aire, extrayendo tanto las amígdalas como los adenoides. Por otro lado, mediante WWSC simplemente se monitoriza la evolución de la enfermedad. Los niños que participaron en el estudio se encontraban en rango de edad entre 5 y 9.9 años, y en todos los casos se obtuvo en consentimiento previo tanto de padres como de niños.

Los registros se obtuvieron en el centro especializado en sueño *Brigham and Women's Sleep Reading Center* en dos rondas, una al inicio del estudio y otra 7 meses más tarde, para poder analizar la evolución. En total se realizó la PSG a 1447 niños, pero para el estudio se seleccionaron sólo 453, que son los que forman la base de datos utilizada, ya que el resto de sujetos no cumplían todos los criterios necesarios, los cuales se enumeran en [66]. En la *Tabla 3.2* se pueden ver las características generales de estos 453; a la hora de evaluar si un paciente tenía sobrepeso se usó el criterio de tener un BMI estandarizado (*Body Mass Index*) en un intervalo superior al percentil 95. La base de datos contiene 3 subgrupos [66]:

1. **Baseline.** Se corresponde con los registros de la ronda inicial en la que se realiza la PSG a los 453 niños seleccionados. Tras esta ronda se dividió el grupo en dos, 226 de los niños pasarían por la operación de eAT y los 227 restantes no.
2. **Follow-up.** Se corresponde con la segunda ronda 7 meses más tarde, en ella se repitió la PSG a 194 de los 226 niños que habían sido operados y a 203 de los 227 que se habían mantenido únicamente en observación. La diferencia en el número de sujetos se debe a que algunos de ellos abandonaron el estudio antes de esta segunda ronda.
3. **Nonrandomized.** Incluye los registros de los niños que no fueron seleccionados para el estudio.

Tabla 3.2: Características de los 453 pacientes que forman el grupo baseline en la base de datos CHAT [67].

Características	
Edad, media (SD)	7.0 (1.4)
Nº de niñas (%)	234 (51.7%)
Nº de afroamericanos (%)	249 (55%)
Sobrepeso (BMI > 95 th)	151 (33.3%)

3.2.2. CARACTERÍSTICAS DE LAS SEÑALES EEG

Las señales de EEG se grabaron usando los canales F_3 , F_4 , C_3 , C_4 , T_3 , T_4 , O_1 , O_2 , M_1 y M_2 tomando el electrodo frontopolar central F_p_z como referencia. La frecuencia de muestreo usada al conseguir las señales no es fija, por lo general es de 200 Hz, pero en algunos pacientes se usaron valores algo mayores (algunos de los registros usan frecuencias de muestreo de 256 Hz o de 512 Hz). Para la obtención de las señales y su clasificación se siguieron las pautas dadas por la AASM [4], dividiendo la señal EEG en épocas de 30 segundos y asignando una fase del sueño a cada una de esas épocas. Dado que el fin de este TFG es poder comparar los resultados de aplicar modelos de clasificación a las dos bases de datos con sujetos de diferentes edades, se necesita que en ambas la frecuencia de muestreo sea la misma, ya que de lo contrario no se podría aplicar un mismo modelo. Para solucionar este problema se realiza un re-muestreo a los 125 Hz que se usaban en SHHS, homogeneizando así todos los registros.

Para re-muestrear se empleó la función `resample()` del módulo `scipy.signal` en Python [68]. A mayores, en alguno de los registros hizo falta eliminar algunas muestras del final, pues el número de épocas de 30 segundos no era exacto.

3.3. ENTRENAMIENTO, VALIDACIÓN Y TEST

A la hora de conseguir los modelos de *Deep learning*, las bases de datos utilizadas se dividen en tres grupos: entrenamiento, validación y test. Estos grupos se forman de manera que cada uno de los sujetos pertenezca únicamente a un grupo, por lo que los datos que contienen cada uno de los tres son independientes.

El grupo de entrenamiento se utiliza para conseguir los diferentes modelos de clasificación. A medida que avanzan las épocas, el modelo irá aprendiendo las características principales de los registros que forman este grupo, con el fin de que luego esa información permita al modelo realizar buenas predicciones al clasificar registros independientes no involucrados en el entrenamiento. El grupo de validación sirve para poder ir controlando la evolución de los modelos entrenados, ayudando al programador a decidir el valor óptimo de los diferentes hiperparámetros de la red, como el número de capas o el nivel de *dropout*. En el caso concreto de este TFG, el grupo de validación se usa para decidir el número óptimo de épocas de entrenamiento. Se impondrá un número de épocas máximo (300000 en el caso del modelo original y 100000 al entrenar los modelos de niños y los modelos conjuntos, como se detallará más adelante) y a lo largo del entrenamiento se irá comprobando en este grupo de validación en qué momento la precisión de la clasificación es mayor. El modelo que mejor se comporte en este grupo de validación será el que se guarde. Cabe destacar que los parámetros de la red únicamente se actualizan de acuerdo a la información del grupo de entrenamiento, nunca el de validación.

Finalmente, el modelo que ha presentado el mejor rendimiento al clasificar el grupo de validación se evalúa en el grupo de test, con el fin de asegurar que el modelo entrenado generaliza lo suficientemente bien. Los resultados obtenidos al clasificar este grupo de test son

los que proporcionan la medida más fiable acerca del rendimiento del modelo. Al igual que con el grupo de validación, aquí tampoco se actualizan los pesos de la CNN.

En el estudio [61] tomado como base, la organización de la base de datos SHHS se realizó de manera que un 50% de los sujetos formaban parte del grupo de entrenamiento, un 20% formaban el grupo de validación y el restante 30% formaba el grupo de test. Con la idea de adaptarse lo más posible al método original, en la primera parte del TFG se han usado estos mismos porcentajes al dividir los registros de adultos antes de entrenar el modelo.

Por otro lado, debido a que el número de sujetos que contiene la base CHAT es mucho menor que el de SHHS (453 en la primera respecto a los 5793 de la segunda), la organización de los sujetos al entrenar los modelos específicos para niños se modifica ligeramente. Aquí se decide asignar un 60% de los pacientes al grupo de entrenamiento, un 20% al grupo de validación y otro 20% al grupo de test. Mediante este cambio se busca que el número de registros usados para entrenar sea lo suficientemente grande como para conseguir modelos de clasificación que consigan detectar con precisión las diferentes fases del sueño.

En la última fase del TFG, en la que se intenta conseguir un modelo conjunto, se vuelve a modificar el contenido de los grupos utilizados. En este caso se decide incluir en el grupo de entrenamiento 316 pacientes de la base de datos de niños y otros 316 de la de adultos, sumando un total de 632 pacientes. El grupo de validación lo forman 130 sujetos, 65 provenientes de una de las bases y otros 65 de la otra. Por último, el grupo de test se forma con los 72 sujetos de la base CHAT restantes (de los 453 que hay en total) junto con otros 72 de SHHS, sumando 144 sujetos para test. Al igual que en los casos anteriores, al organizar los pacientes en grupos se asegura que no hay un mismo paciente que pertenezca a dos grupos diferentes.

Al realizar la organización de grupos para el modelo conjunto se ha buscado balancear el número de registros que se tiene de cada rango de edad, para que así el modelo reciba como entradas épocas de niños y de adultos con la misma probabilidad, y pueda aprender las características que permitan clasificar ambas. Por eso, aunque en realidad se tienen muchos registros de adultos, únicamente se han usado los primeros 453 (siguiendo el orden en el que están numerados), igualando así el número de registros pediátricos de los que se dispone.

Capítulo 4 : METODOLOGÍA

Las redes neuronales no son un concepto nuevo. Ya en los años sesenta se publicaban artículos basados en esta técnica de aprendizaje automático inspirada en el funcionamiento del cerebro, como los estudios de Rosenblatt en [5]. Sin embargo, recientemente el uso de redes neuronales como herramienta de resolución de problemas se ha extendido, impulsado por las mejoras en la capacidad computacional de las máquinas y por el aumento de los datos disponibles para el entrenamiento de las redes. Un subtipo de este tipo de redes son las redes neuronales de *Deep learning* [69], y es dentro de este grupo donde se localiza el método en el que se va a profundizar a lo largo de este capítulo.

Los clasificadores basados en redes neuronales de *Deep learning* se entrenan usando la señal original (*end to end classifiers*) lo que ahorra tener que realizar una etapa previa de extracción de características. También se simplifica el preprocesado que se debe realizar a las señales que entran en la red, pues ella misma es capaz de diferenciar que aspectos de la señal son relevantes en la clasificación.

4.1. REDES NEURONALES CONVOLUCIONALES

Las redes neuronales convolucionales (CNN) representan una alternativa a las redes neuronales convencionales. Se inspiran en el funcionamiento de la corteza visual del cerebro y en sus inicios se pensaron con el fin de ayudar en el procesado de señales 2-D, como imágenes [69]. En 1980 ya se empezaba a hablar de este tipo de redes en estudios como el de Kunihiko Fukushima [28], y con los años las aportaciones de nuevos investigadores fueron mejorando los modelos, aportaciones como la de Yann LeCun en 1998 [70] al introducir en el entrenamiento del sistema el método de aprendizaje *backpropagation*. En los últimos años, el uso de redes convolucionales se ha extendido más allá del procesado de imágenes, demostrando su utilidad en un gran abanico de áreas [69].

Una CNN, como toda red neuronal, está formada por multitud de unidades de procesamiento o neuronas asociadas con ciertos pesos (filtros) y con un factor de corrección o *bias*, el valor óptimo de estos elementos es lo que se busca al entrenar este tipo de redes. Cada neurona recibe varias entradas, de las que hace una suma ponderada de acuerdo con el valor de los pesos, y esa suma se pasa por una función de activación generando una salida.

Existen dos etapas diferenciadas a la hora de entrenar una red neuronal en general, y una CNN en particular. La primera es la etapa en la que se introduce la señal a clasificar en la red y ésta, tras realizar una serie de multiplicaciones o convoluciones, produce una salida. En el caso concreto de uso de una CNN para clasificación de las fases del sueño, la salida será la probabilidad *a posteriori* de que la observación de entrada pertenezca a cada fase.

La segunda es la etapa de propagación hacia atrás, en la que los pesos de la red se actualizan para que el error de clasificación cometido vaya decreciendo al avanzar en el entrenamiento. Esto se consigue seleccionando funciones de coste adecuadas que vayan comparando las salidas de la red con los resultados esperados. Estas dos etapas se suceden secuencialmente a lo largo de la fase de entrenamiento hasta conseguir un modelo óptimo, cuyos pesos permitan, en este caso, una correcta clasificación de las fases del sueño. En la *Figura 4.1* se muestra un esquema de cómo se aplican estas etapas.

Lo que caracteriza a las redes convolucionales es que, en vez de una multiplicación matricial, en alguna de sus capas se usa la operación matemática de la convolución [71]. En la *Figura 4.2* se muestra gráficamente el efecto de esta operación, en la que se aplica el filtro a distintas zonas de la señal de entrada para dar lugar a una salida, que se puede entender como una mezcla entre dicha entrada y el filtro utilizado. De esta manera el filtro recorre todas las posiciones de la matriz de entrada, que se corresponderá con la información proveniente de la capa anterior de la CNN, y genera una matriz de salida, que acabará dando lugar a la siguiente capa de la red. La idea detrás de implementar redes que realicen convoluciones es utilizar características locales en pequeños grupos de la señal de entrada, que acaben dando al modelo la información suficiente para realizar una buena clasificación.

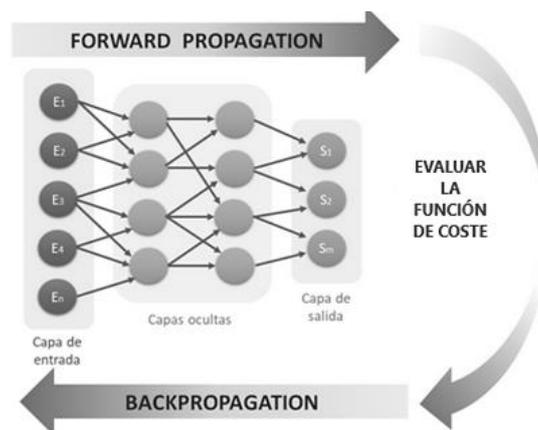


Figura 4.1: Esquema de las etapas principales en las que se basa el entrenamiento de una red neuronal y como se suceden entre sí.

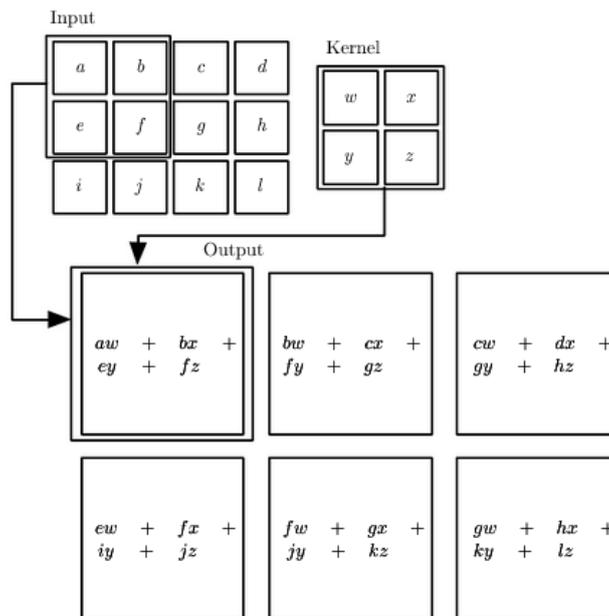


Figura 4.2: Ejemplo de convolución 2-D [71].

Las CNNs han sido usadas en diferentes dominios dentro del procesamiento *end-to-end* de señales continuas, mostrando una elevada capacidad de clasificación. Como ya se ha comentado, empezaron a usarse en el área de visión artificial [72], [73], y poco después se aplicaron también en estudios con otros fines, como el procesamiento de lenguaje [74]. Recientemente se ha comenzado a usar CNNs en investigaciones que se basan en la señal de EEG para desarrollar diversas aplicaciones, entre las que se encuentran sistemas de interfaz cerebro-máquina (BCI) [75], sistemas de detección de convulsiones [76], o sistemas de seguimiento de la mirada [77].

4.1.1. ESTRUCTURA GENERAL DE UNA CNN

Una CNN completa suele estar compuesta por un número variable de capas convolucionales, cada una de ellas seguida por una capa de *pooling*, y terminando con una o dos capas *fully connected* que dan cómo salida la probabilidad de pertenecer a cada clase. La salida con mayor probabilidad se decide como la clase seleccionada por el modelo para esa entrada.

4.1.1.1. CAPA CONVOLUCIONAL

La estructura de una capa convolucional para señales de una dimensión se muestra en la *Figura 4.3*. Cada una de las capas l recibe cómo entrada una matriz $X^{(l-1)}$ correspondiente con la salida de la capa anterior, salvo para la primera capa de la red, cuya entrada será directamente la señal a clasificar. Esa entrada se convoluciona con un filtro $W^{(l)}$, que consiste en una matriz formada por los pesos a entrenar. Siendo $n^{(l-1)}$ la profundidad de la capa $l-1$, equivalente al ancho de la matriz de entrada (para la primera capa este número será 1, ya que se trabaja con señales 1D) y $n^{(l)}$ la profundidad de la capa actual; entonces el filtro tendrá dimensiones $(k^{(l)}, n^{(l-1)}, n^{(l)})$, donde $k^{(l)}$ denota el tamaño de dicho filtro. Para la capa representada en la figura $n^{(l-1)}$ sería de 3, $n^{(l)}$ de 4 y $k^{(l)}$ de 3 [61], [71], [78].

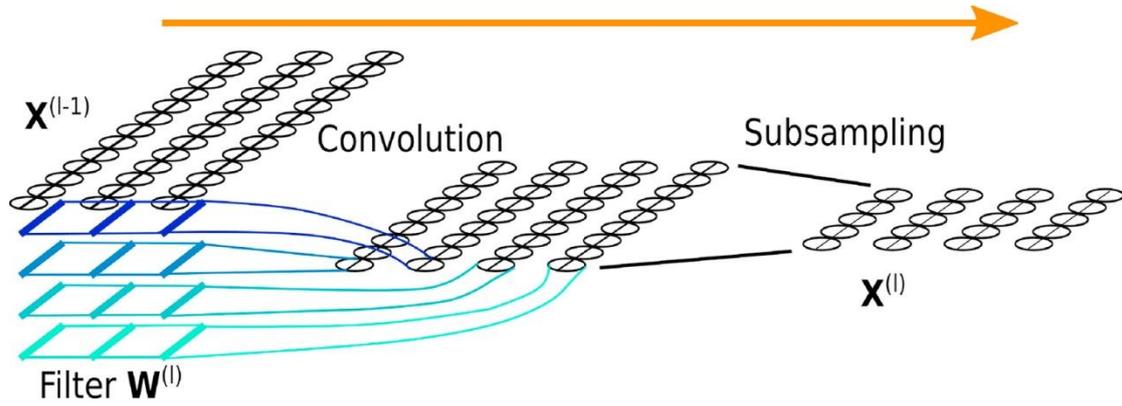


Figura 4.3: Composición de una capa de convolución 1D, incluyendo tanto la fase de convolución como el submuestreo. La función de activación no se representa [61].

En la ecuación (1) se puede ver la expresión que modela cómo se consiguen las salidas en este tipo de capas [61].

$$x_j^{(l)} = \sigma \circ g_{p^{(l)}} \left(\sum_{i=1}^{n^{(l-1)}} x_i^{(l-1)} * w_{ij}^{(l)} + b_j^{(l)} \right) \quad (1)$$

A mayores de lo anteriormente mencionado, en esta ecuación aparecen el elemento σ que representa la función de activación no lineal que se aplica a las salidas, el operador $g_{p^{(l)}}$ que representa el submuestreo realizado, el operador de convolución 1-D '*' y el factor de corrección o *bias* $b_j^{(l)}$ que se aplica a cada salida [61].

El **factor de corrección o bias** es un valor propio de cada unidad o neurona de la red, que se va modificando a lo largo de la fase de entrenamiento, a la vez que los pesos, para que el modelo se vaya adaptando cada vez mejor a los datos. Al resultado de convolucionar la entrada con los filtros se le suma este *bias*, y es eso lo que luego se pasa a la función de activación [71].

El uso de **funciones de activación no lineales** al trabajar con redes neuronales permite a la red ser capaz de modelar funciones de gran complejidad, garantizando que las salidas no son simplemente combinación lineal de las entradas. Esto se asemeja al funcionamiento biológico del cerebro, dónde las neuronas son activadas cuando el estímulo que reciben reúne ciertas características [79].

En la *Figura 4.4* se muestran algunas de las funciones de activación más utilizadas [78], [79]. Las funciones sigmoid y tanh están acotadas, pero tienen la desventaja de que si las salidas son muy elevadas sus derivadas son casi nulas, por lo que el aprendizaje del modelo se ralentiza. Las funciones de activación tipo ReLU (*Rectified Linear Unit*), típicas de las redes *Deep Learning*, solucionan este problema, pues sus derivadas son constantes para toda salida positiva. La función ReLU normal manda los valores negativos a cero, lo que puede acabar anulando muchas de las neuronas de la red. Por ello se exploran variantes como la *Leaky ReLU* o la *Randomized Leaky ReLU*, que asignen cierta pendiente a los valores negativos, en vez de cero, consiguiéndose así un mejor funcionamiento.

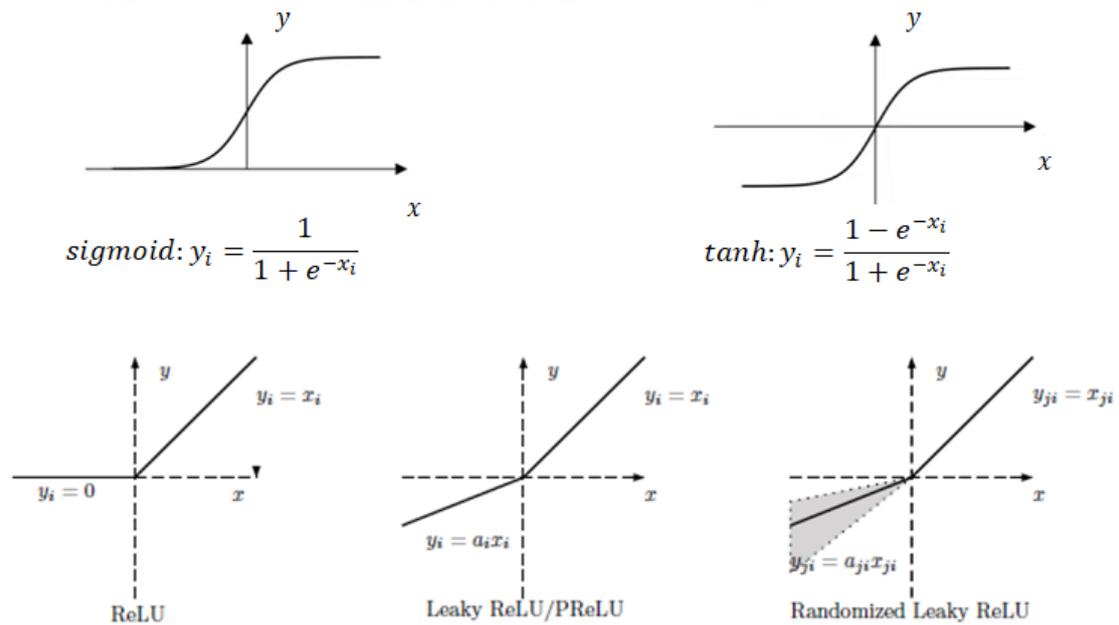


Figura 4.4: Ejemplos de funciones de activación no lineales [78], [79].

El **submuestreo** sirve para reducir la dimensión de la matriz de salida, controlando así que el coste computacional no sea demasiado elevado, a expensas de renunciar a un porcentaje de la información extraída. El efecto del submuestreo es equivalente a indicar que las convoluciones se realicen con un cierto **stride**. El *stride* hace referencia al desplazamiento que hacen los filtros al ir convolucionando con las diferentes regiones de los datos de entrada. Cuanto mayor sea ese desplazamiento menos tardará el filtro en recorrer toda la matriz de entrada y se tendrán que calcular un menor número de convoluciones entre ambos, lo que dará lugar a una disminución en la dimensión de la matriz de salida [71]. En la Figura 4.5 se muestra cómo ambas técnicas provocan que el número de salidas se reduzca, en la imagen de la izquierda se usa un filtro de tamaño 3 con un *stride* de 2, mientras que en la imagen de la derecha se consigue el mismo número de salidas usando también filtro de tamaño 3, pero con un *stride* de 1 e incluyendo una etapa de submuestreo al final.

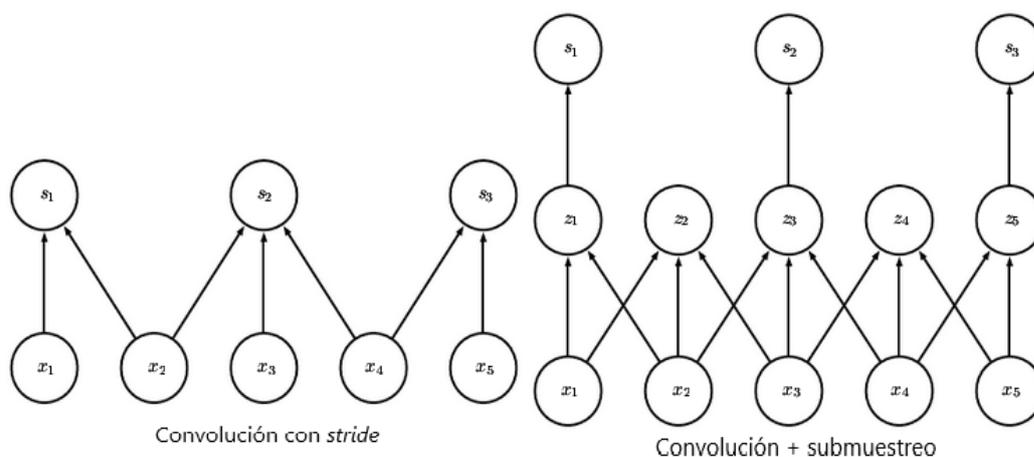


Figura 4.5: Efecto equivalente de submuestrear o aplicar un stride [71].

4.1.1.2. CAPA DE *POOLING*

La capa de agrupación o *pooling* busca que las salidas conseguidas tras cada capa de convolución no se vean afectadas por pequeñas variaciones en las entradas. Esto se consigue sustituyendo el valor de cada unidad de salida de la capa convolucional por otro que tenga también en cuenta el valor de las salidas más próximas. Dependiendo del tipo de *pooling* esto se realiza de una manera u otra. Por ejemplo, en *max pooling* cada unidad toma el valor máximo entre todas las que la rodean, mientras que en *average pooling* se realiza la media. En ocasiones el submuestreo se incluye en esta capa, dividiendo la matriz de salida de la capa convolucional en segmentos del mismo tamaño y tomando de cada segmento únicamente el valor máximo.

En la *Figura 4.6* se muestra un ejemplo de esta operación, en la que se combinan las etapas de *pooling* (*max-pooling* a la derecha y *average-pooling* a la izquierda) y de submuestreo.

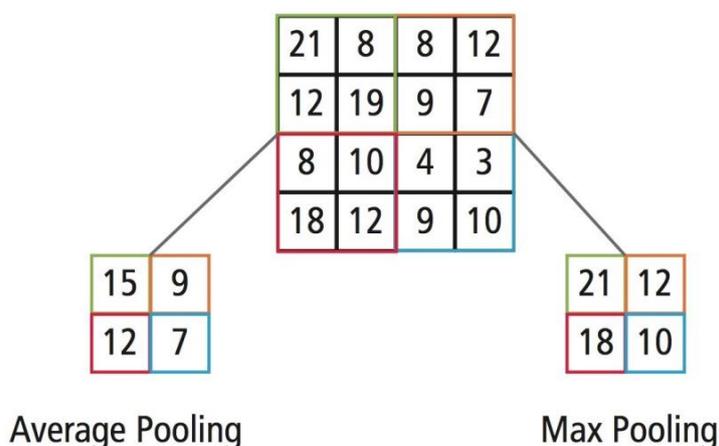


Figura 4.6: Submuestreo con *max pooling* y *average pooling* [80].

4.1.1.3. CAPA FULLY CONNECTED

Las capas *fully connected* o completamente conectadas son equivalentes a una capa de una red neuronal convencional, como se muestra en la parte inferior de la *Figura 4.7*, en la que todos los resultados de la capa anterior están conectados a todos los nodos de la capa siguiente. Normalmente van al final de la red, y, en el caso de estar buscando modelos de clasificación, la última de ellas contiene tantos nodos como clases entre las que clasificar existan. Estas capas, al igual que las convolucionales, también hacen uso de funciones de activación no lineales [78].

4.1.2. VENTAJAS FRENTE A ANN CONVENCIONALES

El motivo por el que se decide usar redes convolucionales es que estas ofrecen diversas ventajas respecto a las redes convencionales, a continuación se detallan brevemente algunas de ellas.

La primera de estas ventajas es que las interacciones entre las distintas unidades de una CNN son dispersas (*sparse interactions*), lo que implica que para generar una salida solo se tienen en cuenta cierto número de unidades de entrada, número que dependerá del tamaño de filtro usado. Por el contrario, en las redes neuronales tradicionales, que usan la operación de multiplicación matricial en vez de la convolución, cada unidad de salida de cierta capa depende de todas las entradas que llegan a esa capa. Esto provoca que el cómputo de las salidas en una CNN requiera de menos operaciones, mejorando así la eficiencia de los algoritmos. En la *Figura 4.7* se muestra este efecto, se ve que el número de unidades de salida (s_i) que dependen de cada unidad de entrada (x_i) es mucho menor si se trabaja con una CNN [71].

Una segunda ventaja que presentan las CNN es que permiten una reducción del número de parámetros a entrenar por capa utilizada. En las redes neuronales tradicionales, para cada una de las unidades de salida de una capa hay tantos pesos diferentes como número de unidades de entrada. Por el contrario, en redes convolucionales las salidas se generan convolucionando un mismo filtro con todas las diferentes zonas de la entrada. Así, en una capa de red neuronal tradicional que reciba m entradas y genere n salidas, la red necesita aprender el valor óptimo de los $m \times n$ pesos distintos que generan las salidas. Sin embargo, en una capa convolucional con el mismo número de entradas y salidas, la red únicamente tiene que aprender el valor óptimo de $k \times n$ pesos, siendo k la anchura de los filtros. Dado que se cumple que $k \ll m$, la cantidad de parámetros en la capa convolucional es mucho menor [71].

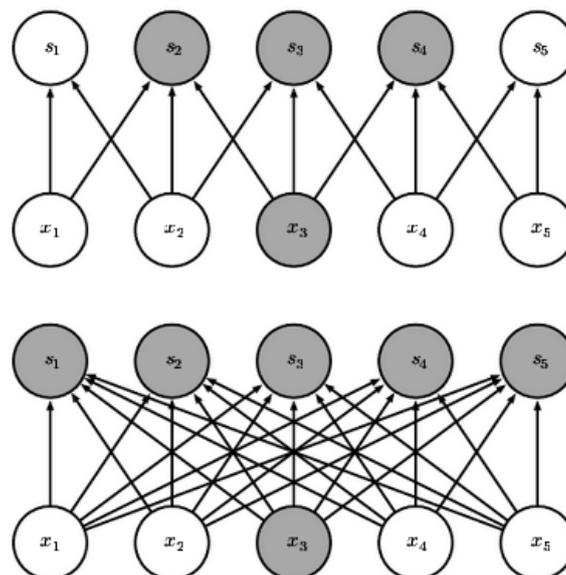


Figura 4.7: Se muestra en sombreado una unidad de entrada x_3 , y las unidades de salida s_i que dependen de dicha unidad de entrada. Arriba las salidas se forman convolucionando con un filtro de tamaño 3 (caso de una CNN), por lo que solo 3 salidas se ven afectadas por una misma entrada. Abajo las salidas se forman por multiplicación matricial (caso de red neuronal), por lo que todas las salidas se ven afectadas [71].

Una última ventaja de las CNN que se va a mencionar es la equivarianza a traslaciones de las entradas. Esta propiedad implica que si se traslada una zona concreta de la señal de entrada a otra localización, dentro de la misma entrada, las salidas experimentarán una traslación equivalente. Esto se debe a que cada salida depende solo de ciertas unidades de entrada cercanas entre sí, por lo que si un grupo concreto de las entradas cambia su posición, seguirá generando las mismas salidas que antes del cambio, pero re-localizadas. Al procesar señales que varían con el tiempo, como es el caso del EEG, esto implica que la red puede diferenciar eventos que ocurren en diferentes instantes de tiempo, y utilizar esta información para generar un modelo de clasificación óptimo [71].

4.2. MODELO ORIGINAL

El sueño se puede clasificar en diferente número de fases, desde dos, diferenciando únicamente entre despierto o dormido, hasta las seis fases definidas en la introducción. Este TFG se ha centrado en la clasificación de 5 fases diferentes, fusionando las fases 3 y 4 en una única fase 3&4 de sueño profundo u ondas lentas, ya que es el criterio que se sigue en la mayoría de estudios y el que adoptaron Sors et al en el método tomado como referencia [61].

Posteriormente, al trabajar con los registros pediátricos, se mantendrá este criterio de fusionar las fases 3 y 4 y clasificar así cada época EEG entre 5 fases del sueño. Dicho método se implementó en Python usando TensorFlow, plataforma de código abierto que facilita el desarrollo de algoritmos de *machine learning*, por ello en este trabajo se han utilizado esas mismas herramientas. A continuación, se explican las principales características del modelo de referencia, empezando por los detalles de su arquitectura y siguiendo con las técnicas de optimización que usa.

4.2.1. ARQUITECTURA

La arquitectura de la CNN usada en el estudio [61] tomado como base para este TFG se muestra en la *Figura 4.8*. La entrada a la red está compuesta por la señal EEG de la época a clasificar concatenada con las señales correspondientes a las dos épocas inmediatamente anteriores y la época posterior. Estas épocas anteriores y posteriores se incluyen para modelar mejor las reglas de clasificación de fases del sueño existentes, que se fijan en las épocas colindantes cuando la actual da lugar a dudas. Cada época de 30 segundos, al usar una frecuencia de muestreo de 125 Hz, tiene un total de 3750 muestras. Por ello la entrada a la red es de $3750 * 4 = 15000$ muestras. Dado que se clasifican todas las épocas, las entradas a la red se superponen, pues una misma época entrará una vez como época actual y otras tres de predecesora o antecesora.

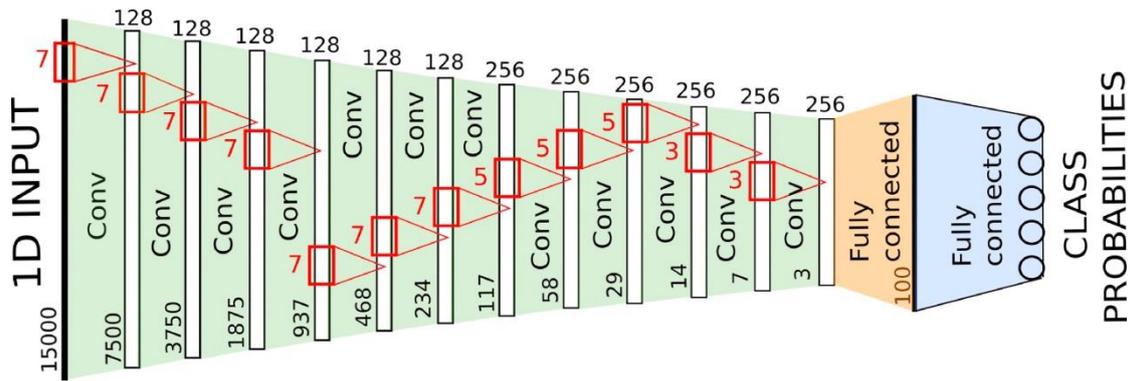


Figura 4.8: Arquitectura de la red convolucional 1D usada en Sors [61].

La arquitectura consta de 14 capas en total, compuestas por 12 capas convolucionales seguidas de una *fully connected* de tamaño 256 y otra *fully connected* de tamaño 5, ya que al ser la capa final el tamaño tiene que coincidir con las 5 clases distintas entre las que clasificar cada época.

La función de activación no lineal usada en todas las capas excepto en la última es una *leaky rectified linear unit* [79], que como se vio en el punto anterior asigna cierta pendiente a los valores negativos, en este caso la pendiente es de 0.1. En la última capa se usa la función de activación *softmax*, también conocida como *multinomial logistic regression*, cuya expresión se muestra en (2). Esta función, acotada entre 0 y 1, transforma las salidas a una representación en forma de probabilidades, de manera que la suma de todas las salidas sea 1 [81]. Anteriormente se ha visto que la última capa de la red tiene una salida por cada clase a clasificar, 5 en este caso, por lo que la salida con mayor probabilidad se corresponderá con la fase del sueño que se ha asignado a la época de entrada.

$$y_i = \frac{e^{x_i}}{\sum_{k=1}^K e^{x_k}} \quad (2)$$

Dado que el tamaño de la entrada a la red es siempre el mismo, la salida de la zona convolucional (tras las 12 primeras capas se tiene como salida una matriz 3x256) está directamente relacionada con dicho tamaño de entrada, con el número de capas de convolución y con sus *strides* (se ha visto que es equivalente a una etapa de submuestreo). Se busca que esta salida no sea demasiado grande para que el número de neuronas en la capa *fully connected* no sea elevado y así no se incremente demasiado el número de pesos de la red, pues, como se ha visto antes, las capas *fully connected* requieren de una cantidad de parámetros mucho más elevada que las capas convolucionales. Teniendo esto en cuenta, el modelo que [61] propone usar *strides* de 2 para las 12 capas de convolución, escogiendo una profundidad de capa de 128 para cada una de las 6 primeras y de 256 para cada una de las siguientes 6. En cuanto a las dimensiones de los filtros, en las 7 primeras capas se decide asignar tamaño 7, en las 3 siguientes tamaño 5, y en las 2 últimas capas, que preceden a las *fully connected*, se utilizan filtros de tamaño 3.

4.2.2. OPTIMIZACIÓN

La función de coste calcula el error entre las clasificaciones estimadas y las reales, con el fin de optimizar los parámetros de la red. En [61] se utiliza como función de coste la entropía cruzada multiclase, una métrica vinculada a la estimación por máxima verosimilitud que refleja la precisión del pronóstico realizado por la red. Además, se realiza un entrenamiento estocástico en pequeños grupos (*mini-batches*) de tamaño 128. Esto implica que, cada 128 entradas del grupo de entrenamiento a la red, los parámetros a optimizar, concretamente los pesos y las *bias*, se actualizan.

La expresión de esta función de coste se muestra en la ecuación (3), donde \mathbf{w} representa el conjunto de parámetros a optimizar, m el tamaño del *batch* (128 en este caso), \mathbf{B} el conjunto de m entradas que forman un *batch*: $\mathbf{B} = \{\mathbf{x}_i^{(0)}, i \in [1, m]\}$, los elementos $\{\mathbf{y}_i, i \in [1, m]\}$ representan la clasificación estimada por la red para cada entrada de \mathbf{B} , y $\{\mathbf{t}_i, i \in [1, m]\}$ representa, en formato *one-hot encoding*, las clasificaciones reales que deberían obtenerse. El que las clasificaciones sigan el formato *one-hot encoding* implica que, para cada entrada \mathbf{x}_i , su \mathbf{t}_i correspondiente es un vector binario de 5 elementos, un elemento por cada fase del sueño, de los cuales 4 tienen valor de 0 y el quinto valor 1. La posición del elemento de valor 1 dentro del vector indica la fase del sueño real de dicha entrada i .

$$l(\mathbf{w}, \mathbf{B}) = -\sum_{i=1}^m \mathbf{t}_i^T \log \mathbf{y}_i(\mathbf{w}) \quad (3)$$

El algoritmo usado para la actualización de los parámetros, una vez cada 128 entradas, sigue el optimizador Adam [82] con parámetros ($lr = 3 \times 10^{-5}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$).

Este optimizador se basa en que se vayan modificando los parámetros de manera que, a medida que se avanza en la fase de entrenamiento, los resultados obtenidos se vayan aproximando a un mínimo de la función de coste. Para ello, cada 128 iteraciones (un *mini-batch*) se da un “paso” en la dirección de este mínimo. Dar este “paso” equivale a actualizar el valor de los parámetros atendiendo al valor del gradiente de la función de coste, pues indica en cada momento la dirección más rápida para acercarse al mínimo. En concreto, Adam tiene en cuenta los momentos de primer y segundo orden de estos gradientes [82].

Existen otros tipos de optimizadores de la función de coste, como el de AdaGrad (*Adaptive Gradient Algorithm*), RMSProp (*Root Mean Square Propagation*) o descenso estocástico del gradiente (SGD, *Stochastic Gradient Descent*). La razón de elegir el optimizador Adam y no otro es que resulta computacionalmente eficiente y ha demostrado su utilidad en problemas en los que la cantidad de datos y parámetros es muy elevada [61].

4.3. SELECCIÓN DE LAS ENTRADAS A LA RED

Las bases de datos usadas en este trabajo son demasiado grandes como para poder cargar todo el conjunto de datos antes de iniciar el entrenamiento de la CNN. Por ello, los diferentes registros se van cargando de la memoria a medida que se van necesitando.

La técnica seguida es hacer uso de 50 colas simultáneamente, cada una asociada a un sujeto, que vayan pasando aleatoriamente a la red las épocas del registro de dicho sujeto. De esta manera, primero se coge aleatoriamente una entrada (formada por la época a clasificar, las dos anteriores y la posterior) de la primera de estas 50 colas, luego una entrada de la segunda cola. Así sucesivamente hasta llegar a la cola en la posición 50, momento en el que se vuelve a pasar a la cola número 1. Una vez que se tienen las 128 entradas que forman un *mini-batch* estas se pasan a la red para ser clasificadas y que, en función de los resultados, se actualicen los parámetros. Junto con las 128 entradas a la red también se le pasa la clasificación real de cada una de ellas, en forma de vector binario como se ha explicado anteriormente.

Cuando se vacía una de estas colas, pues ya se han pasado a la red todas las épocas del registro de ese paciente, se asocia dicha cola a otro paciente diferente y se sigue con el mismo procedimiento. Una vez que todos los sujetos del grupo de entrenamiento han pasado por las colas, se informa por consola de que ya se ha iterado todo el grupo, y se procede a una segunda ronda, en la que volverán a pasar por la red todas las épocas de cada sujeto. A lo largo del entrenamiento se suceden varias iteraciones completas de todo el conjunto de registros. En concreto, para el estudio realizado en [61] se entrena a la red durante 300000 *mini-batches*, lo que implica que se le pasan $300000 \times 128 = 38400000$ entradas a clasificar.

En el código esto se consigue mediante el uso de un conjunto de funciones que manejan objetos de tipo *iterators* [83]. Este tipo de objetos facilitan la tarea de iterar sobre un conjunto de elementos, épocas de registros de EEG en este caso, ya que no se necesita almacenar inicialmente todos estos elementos en una lista o *array*, lo que saturaría la memoria de la máquina utilizada. Cuando el conjunto sobre el que se está iterando finaliza se genera una excepción de tipo `StopIteration`. Los sucesos de vaciado de una cola o de que se acabe de iterar todo el grupo de entrenamiento se controlan por medio de estas excepciones `StopIteration`. El pasar la clasificación real de cada época a formato *one-hot encoding* también se lleva a cabo en estas funciones.

4.4. MEDIDAS DEL RENDIMIENTO DE CLASIFICACIÓN

A la hora de medir la capacidad de clasificación de un modelo concreto normalmente se usan una serie de medidas estadísticas. Estos estadísticos se suelen obtener a partir de la **matriz de confusión** (*confusion matrix*), una representación en forma de tabla que permite observar la clasificación realizada por el modelo automático frente a la clasificación real. Cada columna de la matriz representa el número de épocas que la red ha predicho como pertenecientes a cada una de las fases, mientras que las filas representan el número de épocas que había realmente en cada fase.

En caso de clasificación binaria, únicamente hay que diferenciar entre dos clases posibles (de cara a la explicación se les llamará clases A y B), los elementos de la matriz de confusión son los siguientes:

- a) Verdaderos positivos (TP, *True Positives*). Número de épocas pertenecientes a la clase A que han sido correctamente clasificadas como clase A por el modelo a valorar.

- b) Falsos negativos (FN, *False Negatives*). Número de épocas pertenecientes a la clase A que han sido erróneamente clasificadas como clase B por el modelo a valorar.
- c) Verdaderos negativos (TN, *True Negatives*). Número de épocas pertenecientes a la clase B que han sido correctamente clasificadas como clase B por el modelo a valorar.
- d) Falsos positivos (FP, *False Positives*). Número de épocas pertenecientes a la clase B que han sido erróneamente clasificadas como clase A por el modelo a valorar.

Para problemas multiclase, como es el caso de la detección de las fases del sueño, se pueden definir también estos cuatro elementos, pero centrándose solo en una de las fases cada vez. Así, para el estado despierto (*wake*) los verdaderos positivos son todas las épocas clasificadas por la CNN como *wake*, que de verdad eran épocas *wake*. Los falsos negativos son todas las épocas que, siendo *wake*, han sido clasificadas por el modelo como cualquiera de las otras cuatro fases (1, 2, 3&4 o REM). Los verdaderos negativos son el conjunto de épocas que no son de fase *wake* y no han sido clasificadas como tal. Por último, los falsos positivos engloban a las épocas que, perteneciendo en realidad a alguna de las otras cuatro fases, han sido clasificadas por la red como *wake*.

De esta manera se pueden sacar los valores de TP, FN, TN y FP para cada una de las cinco fases del sueño. Posteriormente, a partir de esos cuatro valores, se pueden sacar los diferentes estadísticos que se definen a continuación. De todos ellos, en este TFG se van a usar sobre todo la precisión y en la kappa (k) de Cohen, pues son los que se usan en el método tomado como base [61].

Precisión o exactitud (*Accuracy, Acc*). Determina, para una cierta fase, el porcentaje de clasificaciones acertadas respecto al total. En caso de trabajar con clases desbalanceadas, que haya muchas más épocas de algunas que de otras, esta medida puede tomar valores altos incluso cuando el modelo de clasificación subyacente no es bueno. Viene dada por:

$$Acc = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

Otro criterio que se ha tenido en cuenta a la hora de evaluar los modelos es el **coeficiente Kappa de Cohen, k** . Este coeficiente mide el nivel de concordancia que hay entre las clasificaciones reales realizadas por un experto en sueño y las clasificaciones dadas por el algoritmo, sin tener en cuenta los casos en los que ambas coincidan solo por casualidad. Viene determinado por [84]:

$$k = \frac{p_o - p_c}{1 - p_o} \quad (5)$$

Donde p_o es el ratio de concordancia observado entre las clasificaciones reales y las conseguidas por la red, y p_c es la probabilidad que existe de que ambas coincidan solo por casualidad. El valor máximo que puede tomar k es +1, caso en el que la concordancia entre las fases reales y las asignadas por la red es perfecta. Si el valor de k es igual a 0 significa que los casos en los que hay coincidencia se deben solo a casualidad, si k es -1 implica que nunca coinciden las predicciones de la red con las clasificaciones reales.

4.5. APLICACIÓN DEL MODELO A NIÑOS

A la hora de evaluar el rendimiento del modelo original en los registros pediátricos se tuvieron que realizar, a mayores del re-muestreo comentado en el capítulo 3, etapas de re-referenciado y de estandarizado para ambas bases de datos. A continuación, se explican ambas etapas.

4.5.1. RE-REFERENCIADO

La señal de EEG se mide de acuerdo a la diferencia de potencial que existe entre dos electrodos, uno de ellos colocado en la zona de la cabeza donde se desea medir la actividad y el otro en un punto de referencia. Idealmente, la actividad eléctrica en el punto de referencia debería mantenerse constante a lo largo del tiempo, para así no afectar a la señal medida. Por ello, se suele situar el electrodo de referencia en lugares como la nariz, los lóbulos de las orejas o los mastoides, que al estar alejados de la zona de actividad neuronal se asumen de potencial constante. Sin embargo, ya se ha demostrado en varias ocasiones que esta asunción es incorrecta, pues en ningún punto del cuerpo humano se puede asumir actividad eléctrica totalmente constante [85]. Ésta es la razón por la que, antes de poder aplicar el mismo modelo a ambas bases, se decide re-referenciarlas. Ya que, aunque en ambas se va a trabajar con el canal C_4 , en SHHS la referencia es A_1 mientras que en CHAT la referencia es Fp_z .

Al revisar el estado del arte se ha visto que existen varias técnicas con las que llevar a cabo este re-referenciado, como el generar una referencia promedio, el usar la transformada Laplaciana de superficie [86] o el aplicar el método REST (*Reference Electrode Standardization Technique*) que consiste en referenciar a un punto en el infinito, de manera que no se vea afectado por la actividad neuronal [87]. En este TFG se ha elegido la primera, referenciado promedio, ya que es la que soluciona el problema de una manera más sencilla. Si en el futuro se viese que esta técnica no da buenos resultados, se podría probar a aplicar otra y comparar lo conseguido con cada una.

Para aplicar el re-referenciado promedio se resta, a cada una de las muestras de la señal EEG, el promedio del valor de las muestras de todos los canales en ese mismo instante. Dado que en la base de datos SHHS se tienen solo los canales C_3 y C_4 , en la base CHAT se han considerado también solo esos dos, para que así las señales resultantes fuesen parecidas. Así, el re-referenciado del canal bipolar $C_4 - Fp_z$ de los registros de CHAT quedaría como se muestra en (11), observándose que la señal resultante ya no depende del electrodo Fp_z usado como referencia.

$$C_4 - Fp_z - \text{promedio} = C_4 - Fp_z - \frac{C_4 - Fp_z + C_3 - Fp_z}{2} = \frac{C_4 - C_3}{2} \quad (6)$$

El caso de las señales de SHHS es algo diferente, pues, como se vio en el capítulo 3, la referencia no es la misma para ambos canales. En uno se usa el lóbulo de la oreja derecha ($C_3 - A_2$), mientras que en el otro el de la izquierda ($C_4 - A_1$). Sin embargo, dado que la actividad en ambos lóbulos es parecida, y para poder seguir avanzando con el TFG en esta línea, se re-referencia considerando que ($A_1 \approx A_2$). En (12) se muestra cómo quedarían las señales de SHHS.

$$C_4 - A_1 - \text{promedio} = C_4 - A_1 - \frac{C_4 - A_1 + C_3 - A_2}{2} = \frac{C_4 - A_1 - C_3 + A_2}{2} \approx \frac{C_4 - C_3}{2} \quad (7)$$

Este paso se aplica en el preprocesado de las señales, antes de que se guarden los registros en memoria y, en el caso de los EEG de niños, justo después de realizar el re-muestreo.

4.5.2. ESTANDARIZADO

Al analizar las señales de ambas bases de datos (adultos y niños) se notó que el rango dinámico variaba mucho de una a otra, posiblemente debido a los diferentes equipos de adquisición utilizados y a las características intrínsecas de las poblaciones bajo estudio. Para solucionar este problema se estandariza cada una de las entradas (15000 muestras formadas por la época a clasificar, las dos épocas anteriores y la posterior) a medida que se van pasando a la CNN.

La función usada para llevar a cabo el estandarizado es `zscore()` [88], perteneciente al conjunto de funciones estadísticas del módulo `scipy.stats`. Con esto se consigue que todas las entradas a la red, independientemente de la base de datos con la que se trabaje, tengan media nula y desviación típica 1.

4.6. MODELO ESPECÍFICO PARA NIÑOS

En la segunda fase del TFG se intenta entrenar, partiendo de la estructura del modelo original, un modelo específico para la detección de las fases del sueño en niños. Para ello, se prueba a variar el número de capas de la red y el valor del parámetro de regularización utilizado (*dropout*). Además, para que el coste computacional de las pruebas no fuese excesivo y para que los tiempos de entrenamiento no se alargasen demasiado, se decide bajar el número de *mini-batches* (conjuntos de 128 entradas a la red tras los que se actualizan los pesos) de los 300000 usados en la sección anterior a 100000. La reducción de este número de *batches* también se debe a que se dispone de menos registros de niños que los que se tiene en el caso de adultos.

4.6.1. DROPOUT

Los modelos de redes de *Deep learning*, con su gran número de parámetros, son algoritmos que ofrecen grandes posibilidades. Sin embargo, el sobreentrenamiento es un problema bastante presente en este tipo de sistemas. El que un modelo se sobreentrene implica que se adapta demasiado al grupo de entrenamiento, lo que provoca que no generalice, pues solo clasificará bien aquellos registros con los que ha sido entrenado. A la hora de buscar un algoritmo que permita la detección automática de las fases del sueño, éste es un fallo que se ha de evitar, ya que el modelo debería dar buenos resultados al clasificar las señales EEG de cualquier paciente. Idealmente se busca que los parámetros de la red se optimicen para ser capaces de distinguir las características de la señal EEG en las distintas fases del sueño, pero ignorando aquellas características que solo son propias de algunos registros del grupo de entrenamiento [29].

Para evitar sobreentrenar los modelos, en este TFG se decide incluir un parámetro de regularización basado en *dropout*. La idea de aplicar esta técnica es que se vayan cancelando algunas de las unidades o neuronas a medida que se entrena la red, de manera que la optimización de los parámetros se ralentiza. Al eliminar una neurona se anulan también las conexiones de entrada y salida asociadas a esa neurona, como se muestra en la *Figura 4.9*, que representa una red neuronal a la que en un caso se le aplica *dropout* y en otro no [29].

La elección de las neuronas a anular es aleatoria. En el método de *dropout* usado en este trabajo cada neurona tiene una probabilidad p de anularse, independientemente de lo que ocurra con el resto de neuronas. El valor de p es fijo durante toda la fase de entrenamiento de un modelo, y viene dado por el índice de *dropout* que se asigna. Un índice de 0.1 ($p = 0.1$) implicará que cada neurona tiene un 10% de probabilidad de ser anulada. Este proceso se aplica únicamente durante la fase de entrenamiento, al evaluar en los grupos de test los modelos previamente conseguidos no se elimina ninguna neurona de la red.

La función usada para llevar a cabo la técnica de *dropout* en este TFG ha sido `tf.nn.dropout()`, que se incluye dentro de las funciones que ofrece Tensor Flow [89]. Esta función recibe como entrada un vector, que representa el conjunto de neuronas que forman cierta capa de la red, y les asigna valor cero con probabilidad p . Además, para que el *dropout* no afecte al valor medio de las salidas de la red, el valor de los elementos asociados a las neuronas no eliminadas se escala por $1/(1 - p)$.

Al revisar el estado del arte se nota que el uso de las técnicas de *dropout* está bastante extendido al trabajar con redes neuronales estándares, pero que su efecto en capas convolucionales no está aún claro [90]. Aunque sí que se han encontrado casos en los que utilizar *dropout* en estas capas ayuda a conseguir mejores modelos [91], en este trabajo se decide usarlo únicamente en las últimas capas de la red, donde ya no se usa la operación de convolución. Se aplica tanto a la salida de la última capa convolucional como a la de la primera capa *fully connected*.

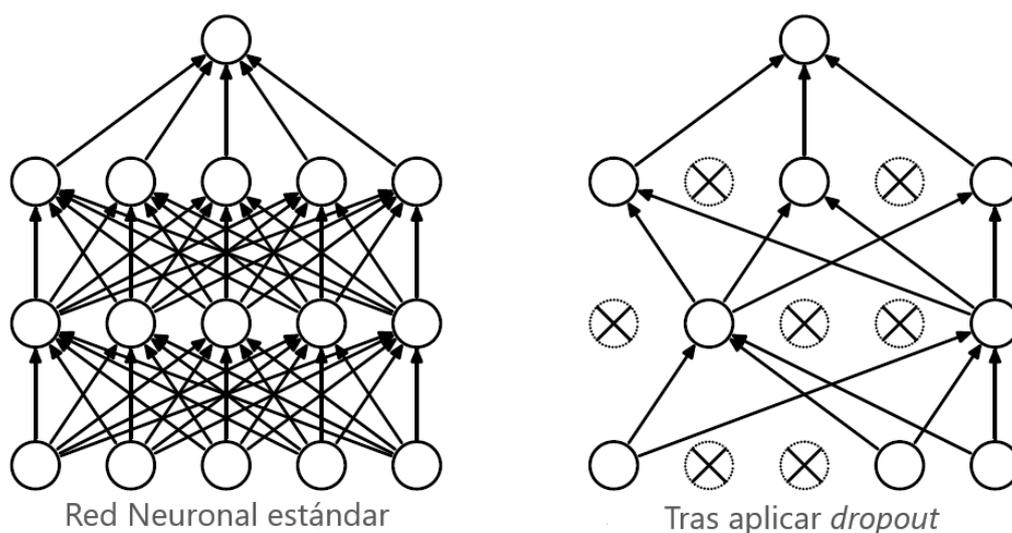


Figura 4.9: Representación de una misma red neuronal con, derecha, y sin dropout, izquierda [29].

A lo largo de la fase de entrenamiento se va a probar con índices de *dropout* de 0.0, 0.1, 0.2, 0.3, 0.4, y 0.5, para luego, en base a los resultados en el grupo de validación, quedarse con el modelo que mejor se comporte al clasificar las fases del sueño.

4.6.2. NÚMERO DE CAPAS

La segunda característica de la red convolucional que se va a modificar a lo largo de la fase de entrenamiento es el número de capas convolucionales. En una red neuronal, el número de capas es uno de los aspectos fundamentales que permiten que el modelo aprenda bien la función que se busca. A mayor número de capas, se podrán modelar funciones cada vez más complejas, aunque hay que tener cuidado, pues en muchas ocasiones esto puede también llevar a sobreentrenamiento.

No existen reglas fijas para elegir el número óptimo de capas que se han de usar en una red neuronal, la técnica que se suele seguir es la de probar con diferentes valores y quedarse con el que mejor comportamiento presente. Entre los estudios previamente realizados que utilizan CNNs como clasificadores *Deep learning* se pueden encontrar arquitecturas de diversos tamaños, dependiendo del tipo de problema que se busque solucionar. En el caso de este TFG, se quiere un modelo lo suficientemente complejo para poder aprender las características de la señal EEG que permiten diferenciar las fases del sueño.

Para adaptarse al formato de la arquitectura original de [61](12 capas convolucionales de las cuales las 6 primeras son de profundidad 128 y las 6 siguientes de profundidad 256) se decide mantener siempre un número de capas par, de manera que la mitad tengan profundidad 128 y la otra mitad 256. Así, se entrenan modelos con 6, 8, 10, 12, 14, 16, 18 y 20 capas convolucionales, a las que siguen los dos niveles *fully connected* de los que ya se ha hablado previamente. El tamaño de los *strides* se mantiene de 2 para todos los niveles, siguiendo el modelo original. Para el tamaño de los filtros también se siguen las pautas de la arquitectura de [61], manteniendo los filtros de tamaño 3 para las 2 últimas capas convolucionales de la red, y los filtros de tamaño 5 para las 3 capas anteriores. En el resto de capas de convolución se usan filtros de tamaño 7.

Capítulo 5 : RESULTADOS

En este capítulo se presentan los resultados obtenidos a lo largo de este TFG. Para ello se evalúan, usando el grupo de registros de test, los diferentes modelos conseguidos. Todos los modelos se basan en una red convolucional, cuyas características se han detallado en el capítulo anterior. En primer lugar se presentan los resultados obtenidos al imitar el modelo original y las variaciones de dicho modelo tras re-referenciar y estandarizar las señales. Además, se muestra cómo se comportan estos modelos al evaluarse con los registros pediátricos. En segundo lugar, se incluyen los resultados obtenidos al entrenar un modelo específico para niños. El capítulo finaliza con la presentación de los resultados obtenidos al generar un modelo conjunto.

5.1. RESULTADOS CON EL MODELO ORIGINAL

El primer paso del TFG fue confirmar que los resultados que se daban en [61] son los que se consiguen al aplicar el modelo original, cuyas características se han detallado en el capítulo 4, a la base de datos SHHS-1. Para ello, usando el grupo de entrenamiento de SHHS-1 (formado por la mitad de los sujetos), se consigue un modelo con parámetros de valor equivalente a los del original.

En las diferentes iteraciones de esta fase de entrenamiento se van guardando la precisión y el coste de las predicciones tanto para el grupo de validación (20% de los sujetos) como para el grupo de entrenamiento, para así poder controlar que no se sobreentrena la red. El que hubiese sobreentrenamiento implicaría que el modelo se adaptaría muy bien al grupo de entrenamiento, pero no conseguiría buenos resultados al clasificar datos independientes, como por ejemplo los registros del grupo de validación. Es importante, a la hora de desarrollar un método de detección de las fases del sueño, asegurarse de que generaliza bien, pues el objetivo ideal es que sea capaz de clasificar las fases a partir de la señal EEF de cualquier paciente.

En la *Figura 5.1* se muestra la evolución de la precisión y el coste en ambos grupos a medida que se avanza en el entrenamiento, el eje de abscisas de ambas gráficas indica el número de *mini-batches* (de 128 entradas cada uno), que ya han pasado por la red en el momento que se realiza cada medida. Se puede ver que los costes tienden a decrecer a medida que se itera la red, mientras que la precisión tiende a aumentar. Se considera que el sobreentrenamiento está minimizado, ya que los resultados obtenidos en el grupo de validación siguen a los obtenidos en el grupo de entrenamiento. El modelo final se consigue, de acuerdo con las indicaciones de [61], tras pasar por la red 300000 *mini-batches*.

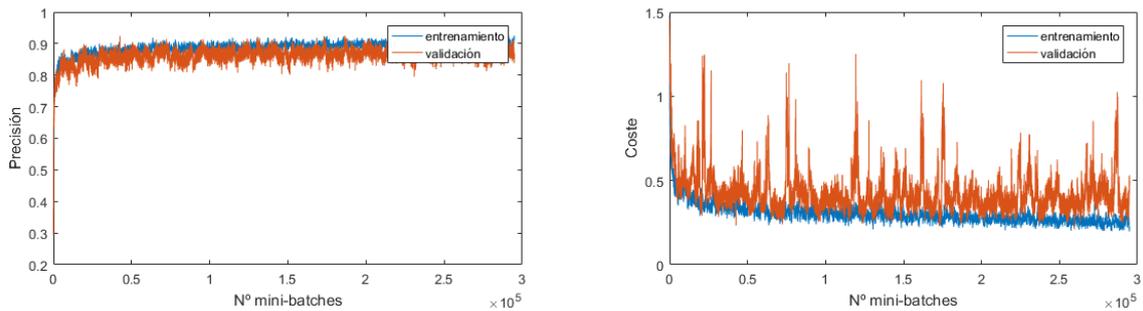


Figura 5.1: Evolución de la precisión y el coste a medida que se entrenaba el modelo, para los grupos de entrenamiento y validación.

A continuación, tras finalizar la fase de entrenamiento se evalúa el modelo conseguido usando el grupo de test, formado por el 30% del total de sujetos en SHHS-1. Los resultados obtenidos son muy similares a los que se esperaban, en la *Tabla 5.1* se puede ver la comparación con [61]. Las pequeñas diferencias existentes entre ambos pueden venir dadas porque los grupos de entrenamiento, validación y test, aunque de igual tamaño que los originales, no estuvieran compuestos por exactamente los mismos sujetos. Ya que en el artículo sólo se especificaba el porcentaje de sujetos que formaban cada uno de los grupos.

La matriz de confusión de la evaluación del modelo en el grupo de test se muestra en la *Figura 5.2*. Se puede ver que la fase que peor clasifica el algoritmo es la fase 1, que se confunde en muchas ocasiones con la fase despierta o con fase 2.

Tabla 5.1: Comparación de los resultados conseguidos en este TFG y los conseguidos por [16] al aplicar el mismo modelo.

	Precisión	Kappa
Sors [Sors]	0.87	0.81
Estudio actual	0.87	0.82

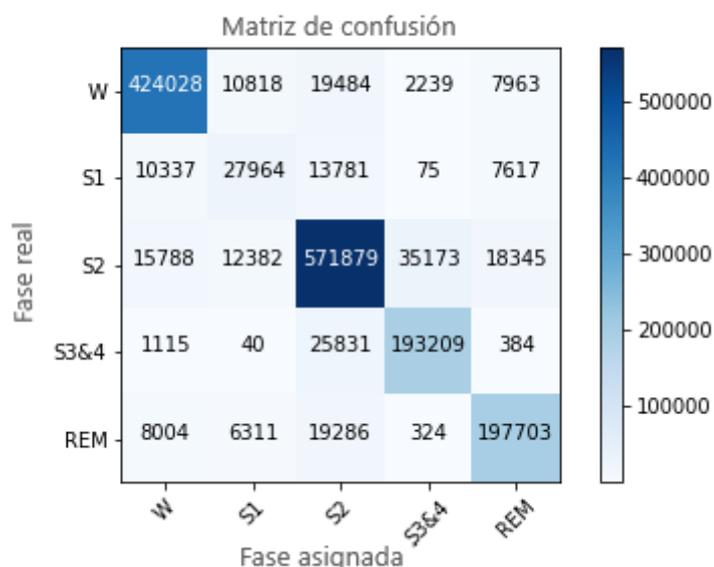


Figura 5.2: Matriz de confusión conseguida al evaluar, con el grupo de test de la base de datos SHHS, el modelo entrenado con el grupo de entrenamiento de SHHS (14 capas y no dropout).

El siguiente paso fue probar el mismo modelo, pero para clasificar los registros de la base de datos CHAT en vez del grupo de test de SHHS. Los resultados de clasificación conseguidos fueron muy bajos, precisión de 0.276 y Kappa de 0.036. En la matriz de confusión, *Figura 5.3*, se ve cómo el algoritmo clasifica, erróneamente, casi todas las épocas como fase despierta. En vista de esto, se pensó que los malos resultados podían deberse a que la referencia era distinta para cada base de datos, A_1 en SHHS-1 y Fp_2 en CHAT.

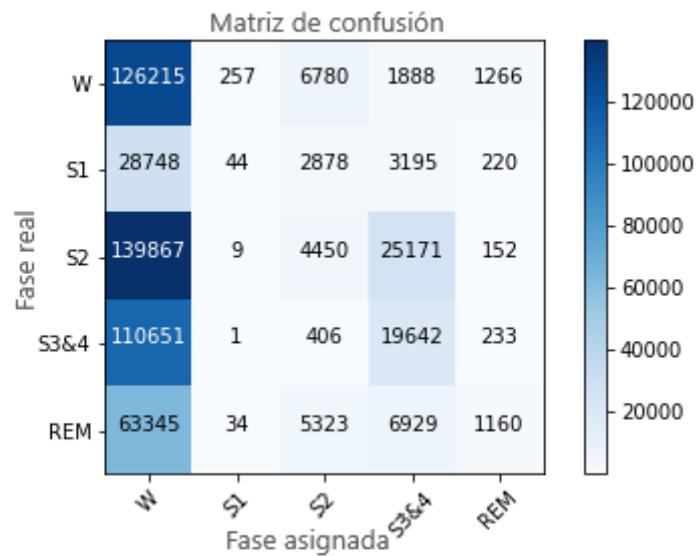


Figura 5.3: Matriz de confusión conseguida al evaluar, con todos los registros de la base de datos CHAT, el modelo entrenado con el grupo de entrenamiento de SHHS (14 capas y no dropout).

5.1.1. APLICANDO RE-REFERENCIADO

Tras observar el bajo rendimiento del modelo al clasificar los registros de la base de datos CHAT, se prueba a re-referenciar las bases, siguiendo la técnica vista en el capítulo 4, para ver si así los resultados mejoran. Con ambas bases ya re-referenciadas, se entrena la arquitectura original usando el grupo de entrenamiento de SHHS. A continuación se evalúa el modelo conseguido tanto en el grupo de test de SHHS como en la base CHAT. Los resultados obtenidos se muestran en la *Tabla 5.2*, y las matrices de confusión en la *Figura 5.4*.

Se puede ver cómo el re-referenciar a provocado que el rendimiento de clasificación del modelo, al evaluarlo con los registros de SHHS, sea algo peor. Sin embargo, no se ha conseguido gran mejoría en la clasificación de los registros de niños. En vista de los resultados, se estudian más a fondo ambas bases de datos y se descubre que el rango dinámico varía considerablemente de una a otra. Los niveles de amplitud que alcanzan las señales de CHAT son bastante mayores que los de los registros de SHHS. Se prueba, a mayores del re-referenciado, a estandarizar ambas bases de datos, confinando así todas las señales a los mismos límites de amplitud.

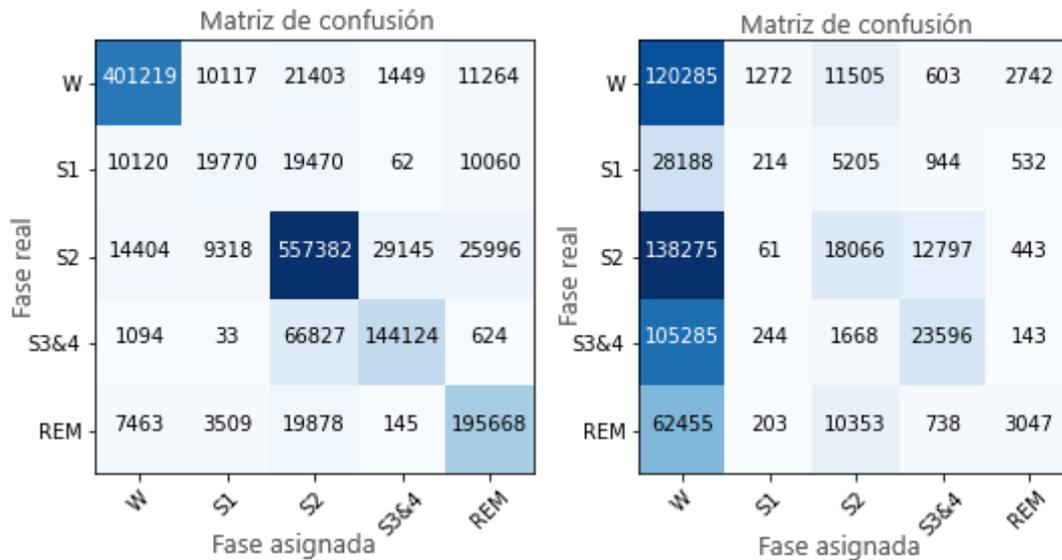


Figura 5.4: Matrices de confusión conseguidas al evaluar, con el grupo de test de SHHS (izquierda) y con todos los registros de la base de datos CHAT (derecha), el modelo entrenado con el grupo de entrenamiento de SHHS (14 capas y no dropout). En este caso todos los registros han sido previamente re-referenciados.

5.1.2. APLICANDO RE-REFERENCIADO Y ESTANDARIZADO

Se prueba a estandarizar las señales ya re-referenciadas antes de pasárselas a la CNN como entradas, siguiendo la técnica vista en el capítulo 4, para ver si así los resultados mejoran. De la misma forma que en los casos anteriores, se entrena la arquitectura original usando el grupo de entrenamiento de SHHS. A continuación se evalúa el modelo conseguido tanto en el grupo de test de SHHS como en la base CHAT. Los resultados obtenidos se muestran en la *Tabla 5.2*, y las matrices de confusión en la *Figura 5.5*.

Se aprecia que, con las bases ya estandarizadas, el rendimiento del modelo al clasificar los registros pediátricos es mucho más alto que en los dos casos anteriores. Además, los resultados para el grupo de test de SHHS se siguen pareciendo a los obtenidos en [61], aunque algo menores. Este modelo conseguido tras estandarizar sí que se aprecia un aumento considerable de rendimiento, pero aún se necesitarían mejores rendimientos en la clasificación de las fases del sueño de los registros pediátricos si se desea que valga como modelo tanto de niños como de adultos.

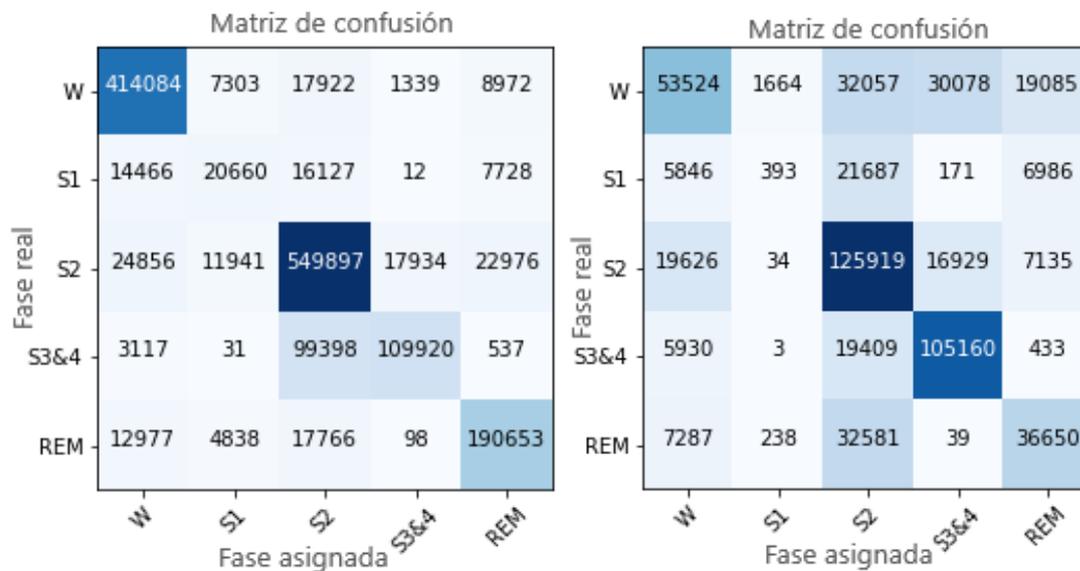


Figura 5.5: Matrices de confusión conseguidas al evaluar, con el grupo de test de SHHS (izquierda) y con todos los registros de la base de datos CHAT (derecha), el modelo entrenado con el grupo de entrenamiento de SHHS (14 capas y no dropout). En este caso todos los registros han sido previamente re-referenciados y estandarizados.

Tabla 5.2: Tabla comparativa con los valores del coeficiente Kappa y precisión obtenidos al aplicar el modelo original a las bases de datos originales, re-referenciadas o re-referenciadas y estandarizadas.

Grupo evaluado	Kappa	Precisión
<i>Registros originales</i>		
SHHS, grupo de test	0.816	0.868
CHAT	0.036	0.276
<i>Registros re-referenciados</i>		
SHHS, grupo de test	0.766	0.834
CHAT	0.067	0.301
<i>Registros re-referenciados y estandarizados</i>		
SHHS, grupo de test	0.738	0.815
CHAT	0.443	0.586

5.2. RESULTADOS CON EL MODELO ENTRENADO CON CHAT

El siguiente paso realizado en este TFG ha sido intentar conseguir un modelo de CNN que clasifique de manera óptima los registros de la base de datos de niños (CHAT). Se opta por usar las señales EEG ya re-referenciadas y estandarizadas para así, una vez conseguido un modelo eficiente a la hora de clasificar los registros pediátricos, se pueda evaluar también clasificando los registros de pacientes adultos.

Para ello se han variado tanto el número de capas convolucionales de la red como el *dropout*, como se detalla en el capítulo 4. En la *Tabla 5.3* se muestran los resultados obtenidos durante el entrenamiento evaluando en el grupo de validación, organizados de manera que para cada número de capas se diferencien los 6 modelos conseguidos, cada uno con un *dropout*.

Tabla 5.3: Resultados obtenidos al entrenar un modelo de CNN que consiga un alto rendimiento al clasificar los EEG pediátricos.

Dropout:	0.0	0.1	0.2	0.3	0.4	0.5
6 capas						
Kappa	0.738	0.764	0.782	0.783	0.767	0.762
Precisión	0.801	0.822	0.835	0.835	0.823	0.819
8 capas						
Kappa	0.790	0.777	0.780	0.796	0.794	0.787
Precisión	0.840	0.831	0.832	0.846	0.844	0.839
10 capas						
Kappa	0.793	0.766	0.801	0.787	0.799	0.792
Precisión	0.843	0.823	0.849	0.839	0.847	0.842
12 capas						
Kappa	0.789	0.778	0.791	0.803	0.799	0.812
Precisión	0.840	0.832	0.841	0.852	0.847	0.858
14 capas						
Kappa	0.808	0.793	0.797	0.800	0.815	0.810
Precisión	0.854	0.843	0.846	0.848	0.859	0.856
16 capas						
Kappa	0.810	0.781	0.798	0.810	0.795	0.794
Precisión	0.856	0.834	0.846	0.856	0.844	0.844
18 capas						
Kappa	0.799	0.810	0.779	0.786	0.803	0.803
Precisión	0.846	0.856	0.833	0.838	0.850	0.850
20 capas						
Kappa	0.784	0.780	0.785	0.798	0.811	0.797
Precisión	0.835	0.834	0.836	0.846	0.856	0.845

En vista de los resultados, el modelo que mejor se comporta a la hora de clasificar los registros de niños que forman el grupo de validación de la base CHAT es el que se resalta en la *Tabla 5.3* de color naranja. Este modelo está compuesto por 14 capas convolucionales y se ha implementado con un coeficiente de *dropout* de 0.4. El comportamiento óptimo del modelo se da cuando ha sido entrenado con 60000 *batches* (128 entradas a la red por cada *batch*), el número de *batches* óptimo se ha monitorizado gracias al grupo de validación, tal y como se detalla en el capítulo 3. Este modelo que mejor rendimiento consigue al clasificar los registros de validación se prueba a continuación en los registros de test, obteniendo como resultados una precisión de 0.840 y un coeficiente Kappa de 0.767.

Una vez conseguido el modelo óptimo, se evalúa en la base de datos SHHS, también referenciada y estandarizada. Con un coeficiente Kappa de 0.25 y una precisión de 0.47, el rendimiento de clasificación del modelo con los registros de adultos no es muy alto. En la *Figura 5.6* se pueden ver las matrices de confusión asociadas a la evaluación tanto en el conjunto de registros de test de CHAT, derecha, como en SHHS, izquierda.

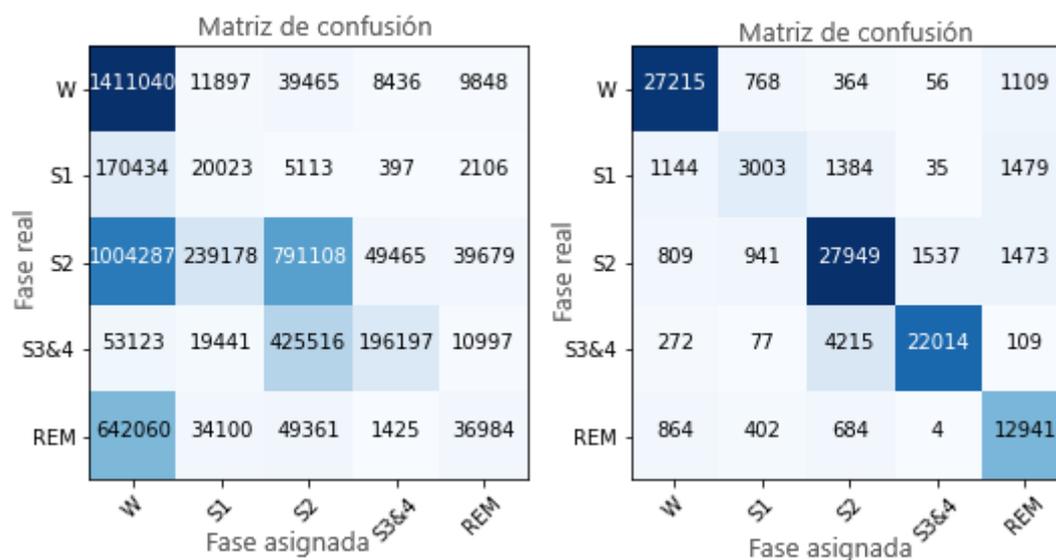


Figura 5.6: Matrices de confusión conseguidas al evaluar, con el grupo de test de CHAT (derecha) y con todos los registros de la base de datos SHHS (izquierda), el modelo entrenado con el grupo de entrenamiento de CHAT (14 capas y un índice de dropout de 0.4). Todos los registros han sido previamente re-referenciados y estandarizados.

Dado que la cantidad de pacientes adultos es más elevada que la de niños (5793 adultos frente a 453 niños), el número total de épocas que muestra la matriz asociada a la base SHHS es mucho mayor. Además, en el caso de SHHS se usan todos los registros de la base para la evaluación, mientras que en CHAT únicamente aquellos pertenecientes al grupo de test, ya que no se quiere evaluar el comportamiento del modelo usando las mismas señales con las que ha sido entrenado o validado.

5.3. RESULTADOS CON EL MODELO ENTRENADO CON CHAT Y SHHS

En esta sección se muestran los resultados obtenidos al entrenar un modelo conjuntamente con registros de niños y de adultos, siguiendo la distribución en grupos de entrenamiento/validación/test que se detalló en el capítulo 3. Aquí las señales EEG que se usan han sido previamente re-referenciadas, y también se lleva a cabo la estandarización de cada entrada a la red. Al entrenar este modelo, por motivos de tiempo y carga computacional, se varía únicamente el número de capas, sin aplicar ningún porcentaje de *dropout*.

En la *Tabla 5.4* se pueden ver los resultados obtenidos al ir evaluando a medida que se avanza en el entrenamiento de la red los modelos en el grupo de validación conjunto, el modelo que mejor comportamiento demuestra al clasificar los registros de validación es el que se corresponde con una arquitectura de 18 capas convolucionales, tras haber sido entrenado con 72000 *batches*. Una vez encontrado el modelo óptimo se evalúa usando el conjunto de registros del grupo de test conjunto, formado por 72 sujetos de cada una de las bases de datos, consiguiendo como resultados una precisión de 0.814 y un coeficiente Kappa de 0.751. La matriz de confusión asociada a la evaluación este modelo óptimo en el grupo de test conjunto se muestra en la *Figura 5.7*.

Tabla 5.4: Resultados obtenidos al entrenar un modelo de CNN con registros de adultos y niños, provenientes de las bases de datos de SHHS y CHAT, respectivamente. La evaluación se realiza usando un grupo de validación con registros de ambas bases de datos, concretamente incluye a 65 pacientes de cada una.

Nº de capas	6	8	10	12	14	16	18	20
Kappa	0.724	0.764	0.767	0.773	0.769	0.764	0.775	0.773
Precisión	0.796	0.825	0.828	0.832	0.829	0.824	0.834	0.832

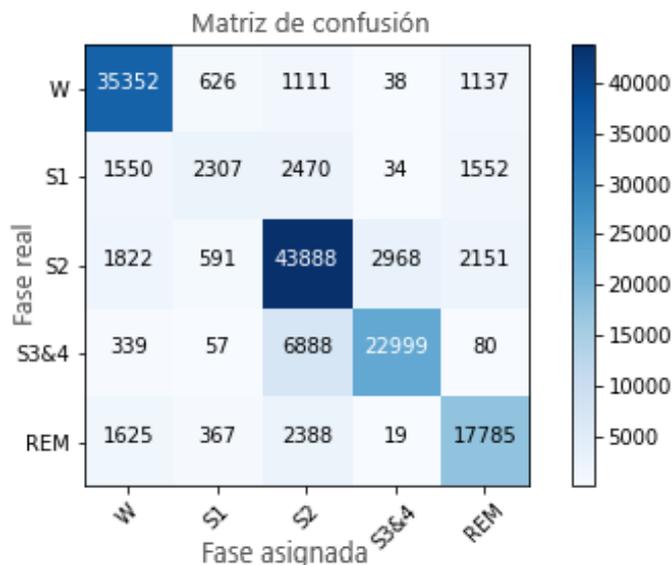


Figura 5.7: Matriz de confusión conseguida al evaluar en el grupo de test conjunto (72 pacientes adultos y 72 niños), el modelo de 18 capas convolucionales entrenado también con registros de ambas edades.

Si se comparan los resultados obtenidos en el modelo conjunto con los de los modelos anteriores, se puede observar que la clasificación conjunta está dando una buena predicción para la mayoría de las épocas. A continuación, para asegurar que este modelo sirve para registros de ambas edades, se evalúa usando registros solo de adultos o solo de niños. Para ello se utilizan los registros de ambas bases de sujetos que no formaban parte ni del grupo de entrenamiento ni del de validación.

Los resultados obtenidos en este paso del TFG se muestran en la *Tabla 5.5* y en la *Figura 5.8*. Se puede ver que ahora, al contrario de lo que ocurría en las fases anteriores, sí que se obtiene un alto rendimiento de clasificación para ambas bases de datos. Al fijarse en las matrices de confusión de la *Figura 5.8* hay que tener en cuenta que al evaluar con la base SHHS se dispone de un número de registros mucho mayor que al evaluar con la base de CHAT.

Tabla 5.5: Resultados de aplicar a las bases de datos por separado el modelo conjunto de 18 capas convolucionales. La evaluación se realiza usando los sujetos que no han formado parte de grupo de entrenamiento ni validación. (Lo que implica que se evalúa en 72 sujetos para la base CHAT y 5340 para la base SHHS).

	SHHS	CHAT
Kappa	0.70	0.79
Precisión	0.79	0.84

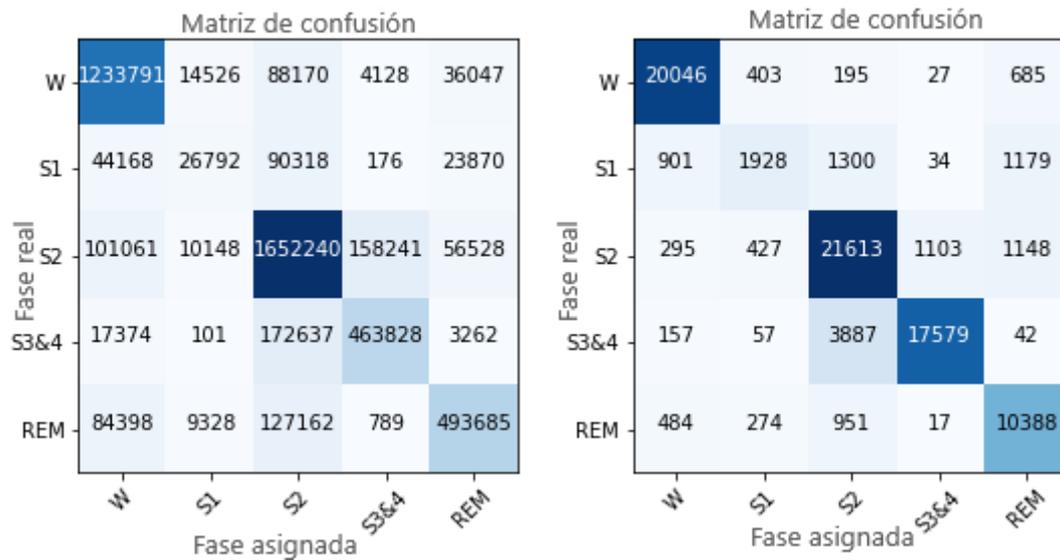


Figura 5.8: Matrices de confusión conseguidas al evaluar con registros de la base de datos CHAT (derecha) y con registros de la base SHHS (izquierda), el modelo entrenado con el grupo de entrenamiento conjunto (18 capas y sin dropout). La evaluación se realiza usando los sujetos que no han formado parte de grupo de entrenamiento ni validación. (Lo que implica que se evalúa en 72 sujetos para la base CHAT y 5340 para la base SHHS).

Capítulo 6 : DISCUSIÓN

En este TFG se ha analizado la opción de utilizar CNNs como métodos de detección automática de las fases del sueño, demostrando que es posible conseguir altos rendimientos de clasificación mediante la aplicación de este tipo de redes a señales EEG de un único canal. Además, se ha indagado en la posibilidad de adquirir un modelo que consiga un alto rendimiento al clasificar registros EEG tanto de adultos como de niños. El usar CNNs permite, además, introducir en la red la señal EEG apenas preprocesada, sin necesidad de implementar etapas previas de extracción y selección de características.

El procedimiento seguido se ha dividido en tres partes bien diferenciadas: replicar el modelo original y probarlo en los registros pediátricos; conseguir un modelo específico para niños y probar a entrenar un modelo combinado, usando registros de pacientes de ambas edades. Para evaluar el rendimiento de los modelos a la hora de clasificar las señales EEG en 5 fases del sueño, se han utilizado grupos independientes de pacientes de test y se ha medido la calidad de la clasificación atendiendo al coeficiente Kappa de Cohen y a la precisión.

La exposición de los resultados obtenidos se realizó en el capítulo anterior, y en el presente se van a analizar e interpretar dichos resultados. En primer lugar, se comentan los resultados obtenidos con el modelo original, después los conseguidos con el modelo entrenado solo con niños y finalmente los del modelo conjunto. A continuación, se lleva a cabo una comparación entre el trabajo actual y estudios previos dentro del contexto de la detección automática de las fases del sueño. Además, al finalizar este capítulo se exponen las limitaciones de este trabajo.

6.1. DISCUSIÓN DEL MODELO ORIGINAL

La primera idea a destacar dentro de este apartado es que, como ya se decía en [61], el modelo original obtiene un alto rendimiento al clasificar las señales EEG de sujetos adultos en 5 fases del sueño, consiguiendo una precisión de 0.868 y un Kappa de 0.816. Al usar los registros re-referenciados, el rendimiento disminuye ligeramente, y cuando además se estandarizan, se disminuye algo más. Aun así, en los tres casos el nivel de clasificación conseguido es elevado, con coeficientes Kappa que no bajan de 0.73.

Es importante destacar que los errores cometidos a la hora de clasificar los registros de pacientes adultos ocurren sobre todo entre épocas adyacentes en las transiciones más comunes que se dan durante un ciclo de sueño. Por ejemplo, las épocas de fase 3&4 mal clasificadas se suelen asignar a fase 2, y casi nunca a fase 1, REM o estado despierto. De la misma manera, la fase 1 se confunde sobre todo con el estado despierto o con fase 2. Otro detalle que resalta al mirar las matrices de confusión (las evaluadas en adultos) es que las fases 1 y 3&4 raramente se confunden. Esto se debe, además de a que no suelen ser fases adyacentes, a que la actividad cerebral es muy diferente entre una y otra. En la fase 1 predomina la actividad en la banda alpha [4], mientras que en las fases 3 y 4 la banda principal

es la delta [4]. También es importante observar que, al estandarizar los registros, el número de épocas que se clasifican erróneamente como *wake* aumenta. Además, la estandarización provoca que los verdaderos positivos de la fase 3&4 decrementsen considerablemente respecto a los de los modelos no estandarizados.

La siguiente idea relevante dentro de este apartado es la importancia de homogeneizar los diferentes registros usados si cada uno proviene de una base de datos distinta. Se aprecia claramente cómo, tras las etapas de re-referenciado y de estandarizado, mejoran los valores de precisión y de coeficiente Kappa obtenidos. Sin embargo, la mejora que se alcanza no es lo suficientemente buena como para poder asumir que el modelo original generaliza bien al aplicarse a sujetos de diferentes edades. La precisión de 0.58 y el Kappa de 0.44 que se consiguen indican que, aún con las bases ya homogeneizadas, la clasificación que predice el algoritmo únicamente concuerda con la dada por el experto en sueño cerca del 60% de las veces, aproximadamente.

Atendiendo a las matrices de confusión conseguidas evaluando el modelo con los registros pediátricos (base de datos de CHAT), en primer lugar, se puede observar que en los modelos sin estandarizar la mayor parte de las épocas se clasifican como fase despierta. Esto deja de ocurrir al fijarse en la matriz conseguida con el modelo estandarizado. De hecho, en esta matriz se pueden ver los detalles comentados anteriormente para el caso de adultos (épocas erróneamente clasificadas se corresponden con fases adyacentes y la fase 1 raramente se confunde con fase 3&4).

Este apartado se puede concluir asegurando que el algoritmo original presenta un alto rendimiento al clasificar registros adultos, pero no generaliza lo bien que se desearía al aplicarlo a registros pediátricos. Esto muestra la necesidad de conseguir un modelo específico para la detección de fases del sueño en niños. Además, también se sugiere en este apartado la utilidad del re-referenciado y especialmente del estandarizado, ya que la mejora, aunque insuficiente, es muy significativa.

6.2. DISCUSIÓN DEL MODELO ESPECÍFICO DE NIÑOS

A la hora de discutir los resultados obtenidos al entrenar un modelo únicamente con los registros de la base CHAT (usando los grupos entrenamiento/validación/test definidos en el capítulo 3), en primer lugar se van a analizar los valores de la *Tabla 5.3*, donde se muestran los resultados obtenidos en el grupo de validación a lo largo del entrenamiento. El primer detalle a destacar es que, para un número de capas fijo, el implementar *dropout* tiende a mejorar la clasificación realizada por el modelo. Esto lleva a pensar que se estaba dando algo de sobreentrenamiento en la red. Por otro lado, para un porcentaje de *dropout* fijo, el aumentar el número de capas de la red provoca inicialmente una mejora en los valores de precisión y Kappa obtenidos, aunque, a partir de cierto punto, estos vuelven a bajar. Esto se debe a que al ir aumentando el número de capas de la red esta puede aprender mejor las características que diferencian una fases del sueño de otras, pero si el número de capas se incrementa en exceso la red empieza a adaptarse demasiado al grupo de entrenamiento y deja de generalizar bien a la hora de clasificar registros que no haya visto nunca.

Es importante destacar que el rendimiento de clasificación conseguido por los modelos entrenados solo con niños se acerca bastante al que mostraba el modelo original. El caso concreto de arquitectura de 14 capas y *dropout* de 0.4 consigue una precisión de 0.84 y un coeficiente Kappa de 0.77, muy similares a la precisión de 0.87 y el Kappa de 0.81 que se detallan en [61]. Esto sugiere que las CNNs son una alternativa válida para tratar con registros provenientes de cualquier grupo de edad, siempre que los modelos se entrenen y se evalúen con sujetos de edades parecidas.

Al fijarse en la matriz de confusión obtenida al evaluar con el grupo de test de CHAT, se aprecia cómo el valor de los verdaderos positivos mejora bastante respecto a lo que se conseguía al probar con CHAT los modelos entrenados solo con adultos. Además, las clasificaciones erróneas siguen las mismas pautas que se comentaron en el caso anterior, se dan sobre todo entre fases adyacentes y raramente entre fases en las que las características de la señal EEG sean muy dispares.

Otra idea que es importante destacar es que, aunque al clasificar los registros pediátricos el rendimiento es alto, este modelo tampoco generaliza. Al evaluarlo en la base SHHS se obtiene un coeficiente Kappa de 0.25 y una precisión de 0.47, valores que indican que las fases predichas por el algoritmo se alejan bastante de las reales.

Teniendo en cuenta lo anteriormente descrito, se puede concluir este apartado resaltando que, si se desea un modelo que consiga buenos resultados tanto para niños como para adultos, se necesitan incluir sujetos de ambas edades tanto en los grupos de entrenamiento como en los de validación.

6.3. DISCUSIÓN DEL MODELO COMBINADO

En este caso se busca un modelo que sea capaz de aprender las características comunes que distinguen las fases del sueño en ambas bases de datos, ignorando las diferencias. En primer lugar se analizan los resultados de la *Tabla 5.4*, conseguidos al evaluar los modelos entrenados con registros de adultos y niños con un grupo de validación también formado por sujetos de ambas edades. En la tabla se ve que, aunque no se llega a los rendimientos de los modelos específicos para cada una de las edades, los valores de precisión y Kappa conseguidos se acercan bastante. El mejor modelo, arquitectura de 18 capas convolucionales, se evalúa a continuación usando un grupo de test independiente a los grupos de validación y entrenamiento anteriores y formado también por pacientes de ambas edades, consiguiendo una precisión de 81% y a un Kappa de 0.75, lo que indica un alto rendimiento en las predicciones que realiza el algoritmo al clasificar las 5 fases del sueño. En la matriz de confusión asociada a este modelo, *Figura 5.7*, se puede ver que para las épocas erróneamente clasificadas se siguen dando las mismas tendencias vistas anteriormente.

Otro detalle a destacar dentro de este apartado es que el modelo conjunto previamente conseguido, el de 18 capas convolucionales, mantiene su alto rendimiento de clasificación cuando se evalúa con grupos únicamente de una de las edades. En la *Tabla 5.5* se puede ver que al clasificar el grupo de registros de CHAT (únicamente los 72 pacientes que formaban

parte del grupo de test conjunto) se consigue un Kappa de 0.79, y al clasificar únicamente los pacientes de SHHS un Kappa de 0.70. Cabe destacar que en el caso de la evaluación con los registros de adultos, se incluyen tanto los registros de los 72 sujetos que formaban parte del grupo de test conjunto como los de los 5340 pacientes restantes, que no se habían incluido en los grupos de entrenamiento, validación o test para que el número de registros de niños y adultos fuese equivalente.

En las matrices de confusión correspondientes con la evaluación del modelo usando las bases de datos por separado, *Figura 5.8*, se aprecia que el comportamiento de la red es similar para ambas bases, teniendo en cuenta que el número de registros que se clasifican al evaluar con SHHS es muy superior a cuando se evalúa con CHAT. Esto puede ser indicativo de que el modelo entrenado ha conseguido aprender las características comunes a ambas bases de datos que permitan diferenciar las fases del sueño, ignorando aquellas otras características específicas de cada una. En esta figura también se aprecia que en ambos casos la fase menos acertada es la fase 1, esto se puede deber a que el número de épocas de esa fase de las que se dispone es menor, por lo que a la red le cuesta más aprender cómo clasificarlas correctamente.

Estos resultados llevan a pensar que la opción de tener un modelo de detección automática de las fases del sueño válido para grupos de diferentes edades es una opción factible, que facilitaría mucho el poder utilizarlo posteriormente en aplicaciones médicas.

6.4. COMPARACIÓN CON OTROS ESTUDIOS

La comparación entre estudios relacionados con la detección automática de las fases del sueño no es tarea fácil, ya que no en todos se usan las mismas bases de datos, número de pacientes o tipo de señales. En este apartado se ha intentado seleccionar estudios de los últimos años que, al igual que el actual, clasifiquen las señales EEG en 5 fases del sueño y usen tanto el coeficiente de Kappa como la precisión a la hora de evaluar los métodos. Es importante asegurarse que el número de fases clasificadas es 5, ya que en los estudios que diferencian entre un número menor (2, 3 o 4 fases) los resultados obtenidos tienden a mejorar al tener menos clases.

Las *Tablas 6.1* y *6.2* muestran algunas de las principales características y resultados de los estudios previos con los que se ha comparado el actual. En concreto, para cada uno de los estudios se indica su año, la base o bases de datos que se usan junto con el número de pacientes que incluyen, el canal de EEG utilizado y su frecuencia de muestreo, el método de clasificación usado, la técnica de validación elegida y los valores de precisión (en tanto por ciento) y coeficiente Kappa (K) obtenidos. Las tablas se han organizado de manera que en *la 6.2* se incluyen los estudios cuyo método de clasificación se basa en redes neuronales (o alguna variante de las mismas), mientras que en *la Tabla 6.1* se incluyen los estudios que utilizan algún otro algoritmo de clasificación.

Cabe destacar que, hasta donde alcanza nuestro conocimiento, no se ha encontrado ningún estudio más que el de [61] que utilice la base de SHHS para métodos de detección automática

de las fases del sueño, por eso no se incluye ninguno en las tablas. De la misma manera, tampoco se han encontrado estudios que detecten las fases de sueño usando la base de CHAT.

En infantes (bebés de menos de 1 año), sí que se han encontrado estudios que presentan técnicas de detección automática de las fases del sueño [92], [93] y [94]. Sin embargo, para niños en el rango de edad de los que forman parte de la base de datos CHAT, entre 5 y 9 años, no se ha encontrado ningún trabajo relacionado.

Tabla 6.1: comparación con otros métodos de detección automática de las fases del sueño. Parte I

Estudio	Base de datos	Canal	Clasificación	Validación	Acc	Kappa
Diykh et al [42] 2016	8 adultos Sleep-EDF 8 adultos Sleep Spindles	P _z -O _z 100Hz C ₃ -A ₁ 200Hz	K-means	10 fold CV	95%	0.80
Hassan et al [7] 2017	8 adultos Sleep-EDF 20 adultos DREAMS	P _z -O _z 100Hz C _z -A ₁ 200Hz	Bagging	LOOCV	93.6% (78%)	0.85 (0.82)
Hassan et al [35] 2017	8 adultos Sleep-EDF 25 adultos UCD	P _z -O _z 100Hz C ₃ -A ₂ 256Hz	AdaBoost	10 fold CV	91.3% (81%)	0.86 (0.81)
Seifpour et al [48] 2018	8 adultos Sleep-EDF 20 adultos DREAMS	Fp _z -C _z 100Hz C _z -A ₁ 200Hz	LIBSVM	50% a train y 50% a test 5 fold CV	91% (83%)	0.87 0.77
Alickovic et al [64] 2018	20 adultos Sleep-EDF extendida	P _z -O _z 100Hz	RotSVM	10 fold CV	91%	0.88
Ghasemzadeh et al [51] 2019	13 adultos Sleep-EDF ext. 10 adultos ISRUC	Fp _z -C _z 100Hz C ₃ -A ₂ 200Hz	SVM	10 fold CV	93% (82%)	0.87 (0.77)
Ghimatgar et al [56] 2019	28 adultos Sleep-EDF y Sleep-EDF ext.	Fp _z -C _z 100Hz	RF y HMM	Cross-dataset	78%	0.69
Jiang et al [55] 2019	20 adultos Sleep-EDF	Fp _z -C _z 100Hz	RF y HMM	2 fold CV	89%	0.83

Tabla 6.2: comparación con otros métodos de detección automática de las fases del sueño. Parte II

Estudio	Base de datos	Canal	Clasificación	Validación	Acc	Kappa
Sors et al [61] 2018	5793 adultos SHHS	C ₄ -A ₁ 125Hz	CNN	0.5/0.2/0.3	87%	0.81
Musa Peker [52] 2015	8 adultos Sleep-EDF	P _z -O _z 100Hz	ANN	LOSO	93%	0.91
Sun et al [59] 2019	20 adultos Sleep-EDF ext. 25 adultos UCD	Fp _z -C _z 100Hz C ₃ -A ₂ 128Hz	BLSTM	LOSO para validación y test.	85% (80%)	0.80 (0.74)
Mousavi et al [60] 2019	8 adultos Sleep-EDF	P _z -O _z 100Hz	CNN	0.45/0.05/0.5	92.9%	0.86
Phan et al [15] 2019	200 adultos MASS	C ₄ -A ₁ 100Hz	RNN	180/10/10 20 fold CV	87.1%	0.815
Zhang et al [62] 2020	25 adultos UCD 16 adultos MIT-BIH	C ₄ -A ₁ 64 Hz	OCNN	LOSO	88.4% (87%)	0.82 (0.81)
Chriskos et al [95] 2019	22 adultos ENVIHAB	Multi-canal	CNN	14 sujetos para entrenar y 8 para test	99%	-
Mousavi et al [63] 2019	8 adultos Sleep-EDF 20 adultos Sleep-EDF ext.	Fp _z -C _z 100Hz	RNN y CNN	20 fold CV	84.2%	0.79
Actual modelo de niños	453 niños CHAT	C ₄ re-ref.	CNN	0.6/0.2/0.2	85.7%	0.81
Actual modelo conjunto	5793 adultos SHHS 453 niños CHAT	C ₄ re-ref.	CNN	0.7/0.14/0.16	82%	0.76

Como ya se ha comentado antes, la mayoría de los estudios incluidos en esta comparación realizan la clasificación en 5 fases del sueño. En aquellos que probaban a clasificar en varias fases, como [48] o [51], únicamente se han cogido los resultados para 5 fases. Se realizan dos excepciones, en [43] que diferencia 6 fases del sueño y en [95] que solo distingue entre 4.

Mirando las tablas comparativas, en primer lugar, se observa que las bases de datos usadas incluyen muy pocos sujetos, y todos de edad adulta. Salvo en [15], que utilizan 200 sujetos, y

en el estudio [61] tomado de partida, ninguna base de datos supera los 25 sujetos. El hecho de que el número de pacientes que se incluyen en este TFG sea tan elevado es algo positivo, pues implica que los modelos se han generado usando una cantidad de datos más diversa y han mostrado su capacidad de generalización en un conjunto más amplio.

Algunos de los estudios utilizan varias bases de datos, con el fin de comprobar si los modelos generalizan bien. En estos casos en la tabla, al poner el valor de la precisión y el coeficiente Kappa, se han incluido entre paréntesis los que se corresponden con la segunda base de datos utilizada. En los estudios [35], [7] y [48] se puede ver cómo los métodos que en una base de datos pequeña (8 sujetos) conseguían muy buenos resultados, al aplicarlos a bases algo mayores presentan rendimientos de clasificación parecidos a los vistos en los algoritmos de este TFG. Cabe destacar que la mayoría de estos estudios no generan un modelo conjunto que luego evalúan con ambas bases de datos, sino que entrenan un modelo diferente con cada una de las bases y luego lo evalúan usando el grupo de test de la base correspondiente. El estudio de Ghimatgar et al [56] es una excepción a esto último, pues entrena el modelo con una base de datos y lo evalúa con otra, buscando así una buena generalización. Elige las bases de manera que los canales, las referencias, y las frecuencias de muestreo coincidan, ahorrándose las dos etapas de re-referenciado y re-muestreo.

Muchos de los estudios usan la técnica de validación *k-fold CV (cross validation)*, que implica que en cada iteración *k* elementos del conjunto total de datos se separan para formar el grupo de validación, y los datos restantes forman el de entrenamiento. Otros usan la técnica *LOSO (leave one subject out)*, que separa un sujeto del grupo de entrenamiento y lo usa para validar los modelos. La división de los datos en grupos de entrenamiento, validación y test se realiza, en muchos de los estudios, a nivel de épocas en vez de a nivel de sujetos [7], [35], [51], [60], [64]. Esta técnica no es muy recomendable, ya que implica que haya distintas épocas de un mismo paciente presentes en los diferentes conjuntos, lo que provoca que los grupos no sean del todo independientes.

La principal diferencia entre los estudios de la *Tabla 6.1* y los de la *Tabla 6.2*, donde se encuentran el modelo de niños y el modelo conjunto desarrollados en este TFG, es que los métodos de la primera tabla requieren fases previas extracción y selección de características, mientras que la mayoría de los estudios de la segunda tabla, al ser basados en redes neuronales de *Deep Learning*, no. Estas fases previas suelen ser bastante complejas, ya que, como se vio en el capítulo 2, las características de la señal EEG no son fáciles de extraer. En algunos casos se aplican las dos técnicas, como en el estudio de Sun et al [59], donde se combinan las características aprendidas por una red neuronal con las extraídas manualmente, y el conjunto es lo que luego se usa para entrenar el clasificador. De esta manera se busca que la información sobre la que realizar la clasificación de las fases del sueño sea mucho más completa.

Además de no requerir fases específicas de selección de características, los métodos basados en redes neuronales *Deep learning* suelen necesitar fases de preprocesado de las señales mucho más simples, ya que, como se ha mostrado en este TFG, la red debería ser capaz de aprender la función de clasificación usando la señal original. Sin embargo, algunos de los estudios de la *Tabla 6.2* sí que realizan preprocesan de manera algo más compleja los datos

antes de entrenar a la red. Este es el caso del estudio de Zhang et al [62], donde convierten previamente las señales EEG a representaciones 2D de tiempo y frecuencia.

Otro punto que marca la diferencia entre los distintos estudios es la utilización de registros provenientes de sujetos con sospecha de padecer trastornos del sueño. La mayoría de estudios se basan en los registros de las bases de Sleep-EDF, que se corresponden únicamente con pacientes sanos. Esto es una desventaja si se prevé acabar utilizando la detección automática de las fases del sueño como herramienta de ayuda en el diagnóstico y monitorización de enfermedades del sueño, ya que en esos casos se requerirá que los métodos presenten altos rendimientos al clasificar registros de pacientes enfermos. Algunos de los estudios sí que usan bases de datos en las que los sujetos podían padecer enfermedades del sueño, como la base de datos UCD, ofrecida por el Hospital universitario de St. Vincent y por la Universidad de Dublin, en la que los 25 pacientes son sospechosos de padecer apnea del sueño. En los artículos [59] o [35] se prueban los métodos tanto en bases de datos de pacientes sanos como de pacientes enfermos, pero no combinadas, sino que se entrena un modelo para cada una y luego se evalúan por separado.

El estudio realizado por Phan et al [15] se acerca bastante a los realizados en este TFG, pues, al igual que en el trabajo actual, clasifica en base a redes neuronales, usa una base de datos moderadamente extensa (formada por 200 sujetos) y la división en los grupos de entrenamiento/validación/test se realiza a nivel de sujetos. Sin embargo utilizan, a mayores del canal C_4 de la señal EEG, la señal de EOG y de EMG, puesto que se ha demostrado que pueden contener información útil a la hora de clasificar las fases del sueño. Esto hace que el método que proponen sea más complejo, ya que obliga a adquirir un mayor número de señales biomédicas. Si se planea adaptar los modelos a dispositivos portátiles que la gente pueda tener en sus casas, la opción de usar un solo canal de EEG es más conveniente. Otra diferencia entre el método propuesto por Phan et al [15] y el método seguido en el estudio actual es que en su modelo se clasifican, en una misma iteración de la red, varias épocas a la vez, en vez de ir de una en una.

Otro estudio que se asemeja al actual es el de Mousavi et al [60], pues utiliza también una CNN con 2 capas *fully connected* al final, pero fijando el número óptimo de capas convolucionales en 9. Además, incluye a mayores un preprocesado que garantiza que el número de épocas por fase del sueño esté balanceado. Éste consiste en repetir parte de las muestras que se tienen para las fases con menos representación consiguiendo así un mayor número de épocas sin variar la información que contienen. Sin embargo, la diferencia principal entre este estudio y el trabajo actual es que Mousavi usa una base de datos con solo 8 sujetos y, además, la división de los registros en los diferentes grupos la realiza a nivel de épocas, no de sujetos, lo que provoca que los grupos no sean del todo independientes.

Muchos de los métodos incluyen una etapa de balanceado de clases, usando diferentes técnicas. Destaca el uso de SMOTE (*Synthetic Minority Oversampling Technique*), que genera nuevas muestras 'sintéticas' para las clases menos representadas, combinando las muestras iniciales. El estudio de Chriskos et al [95] usa esta técnica, y consigue una precisión del 99%, lo que supera por bastante los valores conseguidos por el resto de métodos. Sin embargo, en este estudio [95] se clasifican solo 4 clases, en vez de 5, y no usan grupo de validación.

Además, la clasificación se basa en la señal EEG multicanal, lo que presenta una desventaja si se desea incluir el método de detección automática en algún dispositivo portátil.

Esta sección se puede concluir diciendo que el estudio actual se diferencia de la mayoría de estudios en el ámbito de la detección automática de las fases del sueño por diversos motivos. Entre ellos destacan que utiliza bases de datos con gran número de sujetos, tanto niños como adultos, y que prueba a conseguir un modelo de clasificación conjunta. Aun teniendo en cuenta estas diferencias, los resultados obtenidos se asemejan a los conseguidos por el resto de métodos.

6.5. LIMITACIONES

Para finalizar el capítulo de discusión se van a detallar las limitaciones presentes en el estudio realizado. La primera de estas limitaciones, y una de las más importantes, reside en que el rango de edades de los sujetos en los que se han probado los modelos conjuntos es entre 5 y 9 años, en el caso de los registros pediátricos, y de más de 40 años, en el caso de adultos. Esto implica que no se ha utilizado ningún sujeto de entre 9 y 40 años, un rango de edad de tamaño considerable, lo cual podría llevar a que el rendimiento de clasificación bajase al aplicar el modelo a pacientes entre esas edades. Este efecto no debería de ser muy grande, ya que la señal de EEG cambia sobre todo en los primeros años de vida, por lo que cabe esperar que si en adultos de más de 40 años la clasificación ha sido alta, en adultos de menor edad no sea muy inferior. Sin embargo, para asegurar que esto es así en el futuro se podría incluir una base de datos a mayores que contenga sujetos de entre 9 y 40 años. En el estudio tampoco se han utilizado registros de niños menores a 5 años, pero, dado que en esas edades las señales cerebrales son muy cambiantes, se considera que la tarea de conseguir un modelo que también clasificase correctamente las fases del sueño en niños tan pequeños es ya algo más compleja.

Otra limitación del estudio, también relacionada con las bases de datos utilizadas, es que los registros de CHAT provenían todos de pacientes con sospecha de apnea del sueño. Esto podría sugerir que la clasificación está sesgada debido al mayor número de pacientes con esta patología. No obstante, la apnea del sueño produce cambios en la secuencia natural de las fases del sueño, y no tanto en las características dentro de cada fase, por lo que se considera que los modelos obtenidos también alcanzarían rendimientos elevados en poblaciones no patológicas. Aun así, se necesitarían más pruebas para confirmar esta hipótesis.

La siguiente limitación de la que se va hablar es una que no afecta únicamente a este estudio, sino a la mayoría de estudios en el ámbito de la detección automática de las fases del sueño. Esta limitación reside en que, a la hora de decidir a qué fase corresponde cada época, se toma como valor real la clasificación hecha por los técnicos del sueño. Se ha demostrado en diferentes ocasiones que el nivel de acuerdo entre estos técnicos expertos en sueño no es absoluto. Por ejemplo, en [96] se encontraron coeficientes de Kappa entre 0.46 y 0.89 al comparar las decisiones de clasificación de diferentes expertos. Esto provoca que la calidad y fiabilidad de las anotaciones realizadas en las bases de datos usadas puedan limitar en gran medida el comportamiento final de los modelos. No obstante, los modelos obtenidos han

demostrado alto rendimiento en registros de dos bases de datos diferentes, es decir, analizados por personal especializado de diferentes centros.

Desde el punto de vista de la obtención de los modelos, otra limitación importante encontrada a lo largo de este TFG ha sido el coste computacional y la cantidad de tiempo que conlleva entrenar una red convolucional, sobre todo si se trabaja con bases de datos tan extensas como las usadas en este estudio. Esta limitación ha implicado que hubiese que decidirse por variar únicamente el nivel de *dropout* y el número de capas de la CNN, en vez de probar variando más parámetros, lo que igual hubiese llevado a conseguir mejores modelos. En esta línea cabe mencionar también la limitación de que para el modelo conjunto no se ha variado el nivel de *dropout*, y en caso de haberlo hecho igual se podría haber encontrado algún modelo de clasificación mejor.

La última limitación que se va a mencionar está relacionada con las fases del sueño. Este método permite diferenciar hasta 5 fases diferentes, por lo que no serviría para los casos en los que haga falta distinguir las 6 fases del sueño. Esta limitación no se considera importante, ya que en la mayoría de casos las fases 3 y 4 se combinan formando una sola fase de sueño profundo o de ondas lentas, pero se incluye en esta sección porque pueden darse algunas situaciones en las que se deban diferenciar.

Capítulo 7 : CONCLUSIONES Y LÍNEAS FUTURAS

A lo largo de este trabajo se ha indagado en el uso de CNNs como método de detección automática de las fases del sueño, comprobando su rendimiento en la clasificación de registros de diferentes edades. Para finalizar el estudio, una vez conseguidos los principales objetivos del mismo, se van a detallar en este último capítulo las distintas contribuciones realizadas, las conclusiones alcanzadas y las líneas de investigación futuras que se podrían seguir.

7.1. CONTRIBUCIONES

Las contribuciones que este trabajo aporta al campo de investigación sobre métodos de detección automática de las fases del sueño se pueden resumir en tres puntos principales, que se enumeran a continuación.

- En primer lugar la aplicación de técnicas de homogeneización a las señales EEG procedentes de diferentes estudios, con el fin de poder utilizar un mismo modelo en la clasificación de registros de diferentes bases de datos. En concreto se han aplicado a las señales las técnicas de re-muestreo, re-referenciado y estandarizado.
- También se ha indagado en la posibilidad de obtener un modelo de detección automática de las fases del sueño específico para niños de más de dos años, lo que hasta donde alcanza nuestro conocimiento aún no se había mostrado en la literatura científica. El modelo propuesto se basa en el procesado de la señal EEG de registros pediátricos mediante el uso de redes neuronales *Deep learning* (CNNs).
- Por último, se prueba a conseguir un modelo de detección automática de las fases del sueño que presente un alto rendimiento de clasificación tanto para niños como para adultos. Para ello se utiliza también el procesado mediante CNNs de señales EEG procedentes de sujetos de ambas edades. Los resultados obtenidos llevan a pensar que el desarrollo de un modelo conjunto es una opción viable.

7.2. CONCLUSIONES

Las conclusiones obtenidas a lo largo del desarrollo de este TFG se exponen en los siguientes puntos.

- Los métodos de detección automática de las fases del sueño consiguen rendimientos de clasificación cercanos a los obtenidos manualmente por expertos del sueño, presentando una alternativa que permite simplificar el proceso, ganando objetividad y ahorrando tiempo y dinero. Dentro de los distintos métodos disponibles para

desarrollar la tarea de clasificación, las técnicas de *Deep learning* destacan por las altas precisiones alcanzadas. En concreto, en este estudio se ha comprobado que las CNN son un algoritmo capaz de llevar a cabo esta tarea, que no necesita que se implementen etapas previas de extracción y selección de características.

- Se ha demostrado que la señal de EEG contiene información suficiente para poder realizar predicciones precisas de las fases del sueño. Además, se ha visto que con un solo canal EEG es suficiente, concretamente se ha probado con el C₄, por lo que en la extracción de las señales no haría falta situar más de un par de electrodos, lo que facilita el proceso tanto para el paciente como para el equipo médico.
- A la hora de intentar aplicar un mismo modelo a diferentes bases de datos, se ha notado la importancia de homogeneizar primero las señales de dichas bases. Los pasos principales al realizar esta tarea son re-muestrear a una misma frecuencia, re-referenciar a la media de los canales utilizados en cada caso, y estandarizar las señales para confinarlas en los mismos límites de amplitud. En caso de no realizar este paso previo, se ha visto que las predicciones que daban los modelos se alejaban mucho de las fases reales, pasando de conseguir precisiones de 0.276 y coeficiente kappa de 0.036 antes de homogeneizar, a precisión de 0.586 y kappa de 0.443 con las bases ya homogeneizadas.
- Para trabajar con modelos de *Deep learning*, se ha visto también la importancia de dividir los datos en grupos de entrenamiento, validación y test, de manera que aseguren que el modelo generaliza correctamente y eviten el sobreentrenamiento de la red. Además, es conveniente que esta división se realice a nivel de sujetos, y no a nivel de épocas, para mantener así los grupos independientes entre sí. Se ha visto que el sobreentrenamiento se puede evitar también variando el parámetro de regularización (*dropout*).
- Se ha comprobado que mediante CNNs se pueden conseguir buenos modelos de clasificación para adultos o niños por separado, llegando a precisión de 0.868 y kappa de 0.816 en el caso de modelo específico de adultos y a precisión de 0.84 y kappa de 0.767 en el modelo específico de niños.
- Los resultados obtenidos al evaluar en niños el modelo entrenado con los registros de adultos (kappa de 0.443) indican que este modelo no es generalizable a pacientes de distintas edades. Lo mismo sucede al evaluar en adultos el modelo entrenado únicamente con registros pediátricos (kappa de 0.25). Estos resultados sugieren la necesidad de incluir en el conjunto de entrenamiento a sujetos de diferentes edades para conseguir modelos generalizables.
- El modelo conjunto propuesto permite una detección de las fases del sueño precisa al evaluarse en un grupo de test con registros de ambas edades, consiguiendo una precisión de 0.814 y un kappa de 0.751. Además, se demuestra que dicho modelo

realiza también buenas predicciones al clasificar las bases de datos por separado, con kappa de 0.7 y 0.79 al evaluarse solo con adultos o niños, respectivamente.

- En cuanto a la arquitectura de la CNN utilizada, se ha visto que el número de capas convolucionales necesarias para que el modelo clasifique las señales EEG correctamente ha de ser lo suficientemente elevado como para permitir a la red aprender la función a modelar, en este caso diferenciar las distintas fases del sueño. El número óptimo sale de 16 capas en el modelo específico para niños y de 18 capas para el modelo conjunto.

Finalmente, se puede concluir que a lo largo de este TFG se ha comprobado que el uso de CNNs permite detectar automáticamente las fases de sueño tanto en niños como en adultos, usando para ello únicamente un canal EEG. Sin embargo, debido a las diferencias que existen en la actividad cerebral entre niños y adultos, los modelos entrenados con registros de una edad solo han permitido clasificar correctamente pacientes de esa misma edad. Combinando registros de varias edades en un mismo grupo de entrenamiento sí que se ha conseguido obtener un modelo de clasificación conjunta, que detecta con precisión 5 fases del sueño tanto en registros adultos como en registros pediátricos.

7.3. LÍNEAS FUTURAS

Existen varios aspectos dentro de este TFG que podrían estudiarse con más detalle en futuras investigaciones, la mayoría de ellos se relacionan con las limitaciones expuestas en el capítulo anterior.

Con el fin de asegurar que los modelos conseguidos detectan con precisión las fases del sueño en sujetos de todas las edades, se podrían conseguir registros de pacientes entre 9 y 40 años (rango de edades que no se ha tenido en cuenta en el estudio) y comprobar el rendimiento de clasificación de los modelos al evaluar dichos registros. Además, se podría encontrar alguna base de datos con registros EEG de niños menores a 5 años, y ver cómo se comportan los modelos al clasificar dichas señales. Dado que la actividad cerebral cambia en los primeros meses de vida, se espera que la clasificación no sea demasiado buena, pero a partir de los resultados obtenidos se podría estudiar si sería posible desarrollar un método que sí que consiguiese clasificar correctamente los registros de infantes. En general, todo lo que sea probar los modelos con un mayor número de pacientes aportaría información que permitiría adaptar el método, con el fin de que generalice lo mejor posible.

En cuanto al preprocesado de las señales, sería interesante probar a usar otro tipo de referenciado, y ver si este cambio mejora la clasificación. Una opción sería probar a referenciar a un punto en el infinito, pues es una técnica que parece dar buenos resultados en otros estudios [87].

Una línea de investigación novedosa que se podría seguir a partir del trabajo actual es el probar a incluir algo de *dropout* en las capas convolucionales. Esta técnica, hasta donde alcanza nuestro conocimiento, no ha sido muy usada en los estudios publicados hasta la fecha.

Sin embargo, sí que hay algún artículo que sugiere que puede favorecer el comportamiento de la red, siempre que el índice de *dropout* se mantenga en valores bajos. También en relación a la arquitectura de la red, se podría probar a modificar otros parámetros diferentes al número de capas o el dropout, por ejemplo, variar el factor de aprendizaje (*learning rate*) a medida que se avanza en el entrenamiento, con el fin de que la red converja mejor.

En base a los resultados obtenidos, donde se aprecia cómo las predicciones erróneas de los modelos tienden a concentrarse en épocas de las fases 1 y 3&4, podría ser interesante incluir en el método una etapa que arreglase el desbalanceado que existe en el número de muestras de cada fase, y ver si de esta forma las predicciones de la red para las fases menos representadas mejoran. Para ello se podrían usar técnicas como la de SMOTE, previamente mencionada, o métodos específicos para balanceado de clases en CNNs como los que se proponen en [97].

Finalmente, se podría realizar un estudio del coste computacional que requieren los modelos de CNN una vez entrenados, para ver si serían una buena opción para usarse en dispositivos portátiles. En caso de que no lo fuesen, se podría analizar qué cambios habría que realizar en el método para adaptarlo a dispositivos portátiles.

REFERENCIAS

- [1] J. D. Bronzino, *The Biomedical Engineering Handbook. Medical Devices and Systems*, vol. 53, no. 9. 2013.
- [2] L. Sörnmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. 2005.
- [3] S. Cerutti and C. Marchesi, *Advanced Methods of Biomedical Signal Processing*. New Jersey: John Wiley & Sons, Inc., 2011.
- [4] A. L. C. J. e S. F. Q. Conrad Iber, Sonia Ancoli-Israel, *The AASM Manual for the Scoring of Sleep and Associated Westchester*. 2007.
- [5] F. Rosenblatt, "Perceptron Simulation Experiments," *Proc. IRE*, vol. 48, no. 3, pp. 301–309, 1960, doi: 10.1109/JRPROC.1960.287598.
- [6] L. F. Haas, "Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography.," *J. Neurol. Neurosurg. Psychiatry*, vol. 74, no. 1, p. 9, Jan. 2003, doi: 10.1136/jnnp.74.1.9.
- [7] A. R. Hassan and A. Subasi, "A decision support system for automated identification of sleep stages from single-channel EEG signals," *Knowledge-Based Syst.*, vol. 128, pp. 115–124, 2017, doi: 10.1016/j.knosys.2017.05.005.
- [8] P. J. Uhlhaas and W. Singer, "Abnormal neural oscillations and synchrony in schizophrenia," *Nature Reviews Neuroscience*, vol. 11, no. 2. Nature Publishing Group, pp. 100–113, Feb. 2010, doi: 10.1038/nrn2774.
- [9] L. De Gennaro, M. Ferrara, and M. Bertini, "The boundary between wakefulness and sleep: Quantitative electroencephalographic changes during the sleep onset period," *Neuroscience*, vol. 107, no. 1, pp. 1–11, Nov. 2001, doi: 10.1016/S0306-4522(01)00309-8.
- [10] S. Scholle and G. Zwacka, "Arousals and obstructive sleep apnea syndrome in children," *Clin. Neurophysiol.*, vol. 112, no. 6, pp. 984–991, Jun. 2001, doi: 10.1016/S1388-2457(01)00508-9.
- [11] M. Le Van Quyen *et al.*, "Large-scale microelectrode recordings of high-frequency gamma oscillations in human cortex during sleep," *J. Neurosci.*, vol. 30, no. 23, pp. 7770–7782, Jun. 2010, doi: 10.1523/JNEUROSCI.5049-09.2010.
- [12] J. L. Cantero, M. Atienza, J. R. Madsen, and R. Stickgold, "Gamma EEG dynamics in neocortex and hippocampus during human wakefulness and sleep," *Neuroimage*, vol. 22, no. 3, pp. 1271–1280, Jul. 2004, doi: 10.1016/j.neuroimage.2004.03.014.
- [13] S. Vanhatalo, J. M. Palva, M. D. Holmes, J. W. Miller, J. Voipio, and K. Kaila, "Infraslow oscillations modulate excitability and interictal epileptic activity in the human cortex during sleep," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 14, pp. 5053–5057, Apr. 2004, doi: 10.1073/pnas.0305375101.
- [14] JASPER and H. H., "The ten-twenty electrode system of the international federation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 10, pp. 370–375, 1958.
- [15] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "SeqSleepNet: End-to-End Hierarchical Recurrent Neural Network for Sequence-to-Sequence Automatic Sleep Staging," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 400–410, 2019, doi: 10.1109/TNSRE.2019.2896659.

- [16] R. Boostani, F. Karimzadeh, and M. Nami, "A comparative review on sleep stage classification methods in patients and healthy individuals," *Comput. Methods Programs Biomed.*, vol. 140, pp. 77–91, 2017, doi: 10.1016/j.cmpb.2016.12.004.
- [17] D. Zhao, "Computer Methods and Programs in Biomedicine Comparative analysis of different characteristics of automatic sleep stages," *Comput. Methods Programs Biomed.*, vol. 175, pp. 53–72, 2019, doi: 10.1016/j.cmpb.2019.04.004.
- [18] O. Faust, H. Razaghi, R. Barika, E. J. Ciaccio, and U. R. Acharya, "A review of automated sleep stage scoring based on physiological signals for the new millennia," *Comput. Methods Programs Biomed.*, vol. 176, pp. 81–91, 2019, doi: 10.1016/j.cmpb.2019.04.032.
- [19] A. Kishi, Z. R. Struzik, B. H. Natelson, F. Togo, and Y. Yamamoto, "Dynamics of sleep stage transitions in healthy humans and patients with chronic fatigue syndrome," *Am. J. Physiol. Integr. Comp. Physiol.*, vol. 294, no. 6, pp. R1980–R1987, Jun. 2008, doi: 10.1152/ajpregu.00925.2007.
- [20] M. Eisermann, A. Kaminska, M. L. Moutard, C. Soufflet, and P. Plouin, "Normal EEG in childhood: From neonates to adolescents," *Neurophysiol. Clin.*, vol. 43, no. 1, pp. 35–65, 2013, doi: 10.1016/j.neucli.2012.09.091.
- [21] I. Feinberg and I. G. Campbell, "Sleep EEG changes during adolescence: An index of a fundamental brain reorganization," *Brain Cogn.*, vol. 72, no. 1, pp. 56–65, 2010, doi: 10.1016/j.bandc.2009.09.008.
- [22] I. Feinberg, S. Hibi, and V. R. Carlson, "Changes in EEG Amplitude During Sleep with Age," Springer, Boston, MA, 1977, pp. 85–98.
- [23] L. Tarokh and M. A. Carskadon, "Developmental changes in the human sleep EEG during early adolescence," *Sleep*, vol. 33, no. 6, pp. 801–809, 2010, doi: 10.1093/sleep/33.6.801.
- [24] RECHTSCHAFFEN and A., "A manual of standardized terminology, technique and scoring system for sleep stages of human subjects," *Public Heal. Serv.*, 1968.
- [25] R. B. Berry *et al.*, "Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events," *J. Clin. Sleep Med.*, vol. 8, no. 5, pp. 597–619, 2012, doi: 10.5664/jcsm.2172.
- [26] R. B. Berry *et al.*, "AASM scoring manual updates for 2017 (version 2.4)," *Journal of Clinical Sleep Medicine*, vol. 13, no. 5. American Academy of Sleep Medicine, pp. 665–666, 2017, doi: 10.5664/jcsm.6576.
- [27] J. C. Principe, S. K. Gala, and T. G. Chang, "Sleep Staging Automaton Based on the Theory of Evidence," *IEEE Trans. Biomed. Eng.*, vol. 36, no. 5, pp. 503–509, 1989, doi: 10.1109/10.24251.
- [28] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980, doi: 10.1007/BF00344251.
- [29] G. H. A. K. I. S. R. S. N Srivastava, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.
- [30] H. G. Jo, J. Y. Park, C. K. Lee, S. K. An, and S. K. Yoo, "Genetic fuzzy classifier for sleep stage identification," *Comput. Biol. Med.*, vol. 40, no. 7, pp. 629–634, 2010, doi: 10.1016/j.combiomed.2010.04.007.
- [31] "Sleep-EDF Database Expanded v1.0.0." <https://physionet.org/content/sleep-edfx/1.0.0/>

- (accessed Jun. 26, 2020).
- [32] "About Mass." <https://massdb.herokuapp.com/en/> (accessed Jun. 26, 2020).
- [33] "Data | ISRUC-SLEEP Dataset." https://sleeptight.isr.uc.pt/?page_id=48 (accessed Jun. 26, 2020).
- [34] "Sleep Heart Health Study - Sleep Data - National Sleep Research Resource - NSRR." <https://sleepdata.org/datasets/shhs> (accessed May 30, 2020).
- [35] A. R. Hassan and M. I. H. Bhuiyan, "An automated method for sleep staging from EEG signals using normal inverse Gaussian parameters and adaptive boosting," *Neurocomputing*, vol. 219, pp. 76–87, 2017, doi: 10.1016/j.neucom.2016.09.011.
- [36] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jan. 2002, doi: 10.1613/jair.953.
- [37] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proceedings of the International Joint Conference on Neural Networks*, 2008, pp. 1322–1328, doi: 10.1109/IJCNN.2008.4633969.
- [38] K. C. Chua, V. Chandran, U. R. Acharya, and C. M. Lim, "Application of higher order statistics/spectra in biomedical signals-A review," *Medical Engineering and Physics*, vol. 32, no. 7. Elsevier, pp. 679–689, Sep. 01, 2010, doi: 10.1016/j.medengphy.2010.04.009.
- [39] "Machine Learning | Coursera." <https://www.coursera.org/learn/machine-learning> (accessed May 30, 2020).
- [40] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier Inc., 2016.
- [41] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: 10.1016/j.patrec.2009.09.011.
- [42] M. Diyxh, Y. Li, and P. Wen, "EEG sleep stages classification based on time domain features and structural graph similarity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 11, pp. 1159–1168, 2016, doi: 10.1109/TNSRE.2016.2552539.
- [43] M. Diyxh and Y. Li, "Complex networks approach for EEG signal sleep stages classification," *Expert Syst. Appl.*, vol. 63, pp. 241–248, 2016, doi: 10.1016/j.eswa.2016.07.004.
- [44] P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis, "Robust Data Mining," *Robust Data Min.*, pp. 27–33, 2013, doi: 10.1007/978-1-4419-9878-1.
- [45] S. F. Liang, C. E. Kuo, Y. H. Hu, Y. H. Pan, and Y. H. Wang, "Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 6, pp. 1649–1657, 2012, doi: 10.1109/TIM.2012.2187242.
- [46] S. I. Dimitriadis, C. Salis, and D. Linden, "A novel, fast and efficient single-sensor automatic sleep-stage classification based on complementary cross-frequency coupling estimates," *Clin. Neurophysiol.*, vol. 129, no. 4, pp. 815–828, 2018, doi: 10.1016/j.clinph.2017.12.039.
- [47] G. Zhu, Y. Li, and P. P. Wen, "Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal," *IEEE J. Biomed. Heal. Informatics*, vol. 18, no. 6, pp. 1813–1821, 2014, doi: 10.1109/JBHI.2014.2303991.
- [48] S. Seifpour, H. Niknazar, M. Mikaeili, and A. M. Nasrabadi, "A new automatic sleep staging system based on statistical behavior of local extrema using single channel EEG signal," *Expert*

- Syst. Appl.*, vol. 104, pp. 277–293, 2018, doi: 10.1016/j.eswa.2018.03.020.
- [49] C. C. Chang and C. J. Lin, “LIBSVM: A Library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011, doi: 10.1145/1961189.1961199.
- [50] C. S. Huang, C. L. Lin, W. Y. Yang, L. W. Ko, S. Y. Liu, and C. T. Lin, “Applying the fuzzy C-means based dimension reduction to improve the sleep classification system,” *IEEE Int. Conf. Fuzzy Syst.*, vol. 1, pp. 1–5, 2013, doi: 10.1109/FUZZ-IEEE.2013.6622495.
- [51] P. Ghasemzadeh, H. Kalbkhani, S. Sartipi, and M. G. Shayesteh, “Classification of sleep stages based on LSTAR model,” *Appl. Soft Comput. J.*, vol. 75, pp. 523–536, 2019, doi: 10.1016/j.asoc.2018.11.007.
- [52] M. Peker, “An efficient sleep scoring system based on EEG signal using complex-valued machine learning algorithms,” *Neurocomputing*, vol. 207, pp. 165–177, 2016, doi: 10.1016/j.neucom.2016.04.049.
- [53] P. Memar and F. Faradji, “A Novel Multi-Class EEG-Based Sleep Stage Classification System,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 84–95, 2018, doi: 10.1109/TNSRE.2017.2776149.
- [54] X. Li, L. Cui, S. Tao, J. Chen, X. Zhang, and G. Q. Zhang, “HyCLASSS: A Hybrid Classifier for Automatic Sleep Stage Scoring,” *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 2, pp. 375–385, 2018, doi: 10.1109/JBHI.2017.2668993.
- [55] D. JIANG, Y. nan LU, Y. MA, and Y. WANG, “Robust sleep stage classification with single-channel EEG signals using multimodal decomposition and HMM-based refinement,” *Expert Syst. Appl.*, vol. 121, pp. 188–203, 2019, doi: 10.1016/j.eswa.2018.12.023.
- [56] H. Ghimatgar, K. Kazemi, M. S. Helfroush, and A. Aarabi, “An automatic single-channel EEG-based sleep stage scoring method based on hidden Markov Model,” *J. Neurosci. Methods*, vol. 324, no. May, p. 108320, 2019, doi: 10.1016/j.jneumeth.2019.108320.
- [57] A. R. Hassan and M. I. H. Bhuiyan, “Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting,” *Comput. Methods Programs Biomed.*, vol. 140, pp. 201–210, 2017, doi: 10.1016/j.cmpb.2016.12.015.
- [58] N. Michielli, U. R. Acharya, and F. Molinari, “Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals,” *Comput. Biol. Med.*, vol. 106, no. January, pp. 71–81, 2019, doi: 10.1016/j.compbimed.2019.01.013.
- [59] C. Sun, J. Fan, C. Chen, W. Li, and W. Chen, “A Two-Stage Neural Network for Sleep Stage Classification Based on Feature Learning, Sequence Learning, and Data Augmentation,” *IEEE Access*, vol. 7, pp. 109386–109397, 2019, doi: 10.1109/access.2019.2933814.
- [60] Z. Mousavi, T. Yousefi Rezaii, S. Sheykhivand, A. Farzamia, and S. N. Razavi, “Deep convolutional neural network for classification of sleep stages from single-channel EEG signals,” *J. Neurosci. Methods*, vol. 324, no. April, 2019, doi: 10.1016/j.jneumeth.2019.108312.
- [61] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J. F. Payen, “A convolutional neural network for sleep stage scoring from raw single-channel EEG,” *Biomed. Signal Process. Control*, vol. 42, pp. 107–114, 2018, doi: 10.1016/j.bspc.2017.12.001.
- [62] J. Zhang, R. Yao, W. Ge, and J. Gao, “Orthogonal convolutional neural networks for automatic sleep stage classification based on single-channel EEG,” *Comput. Methods Programs Biomed.*, vol. 183, p. 105089, 2020, doi: 10.1016/j.cmpb.2019.105089.

- [63] S. Mousavi, F. Afghah, and U. Rajendra Acharya, "Sleeppegnet: Automated sleep stage scoring with sequence to sequence deep learning approach," *PLoS One*, vol. 14, no. 5, May 2019, doi: 10.1371/JOURNAL.PONE.0216456.
- [64] E. Alickovic and A. Subasi, "Ensemble SVM method for automatic sleep stage classification," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 6, pp. 1258–1265, 2018, doi: 10.1109/TIM.2018.2799059.
- [65] S. F. Quan *et al.*, "The Sleep Heart Health Study: Design, Rationale, and Methods," *Sleep*, vol. 20, no. 12, pp. 1077–1085, 1997, doi: 10.1093/sleep/20.12.1077.
- [66] S. Redline *et al.*, "The Childhood Adenotonsillectomy Trial (CHAT): Rationale, Design, and Challenges of a Randomized Controlled Trial Evaluating a Standard Surgical Procedure in a Pediatric Population," *Sleep*, vol. 34, no. 11, pp. 1509–1517, Nov. 2011, doi: 10.5665/sleep.1388.
- [67] M. Baumert *et al.*, "The effect of adenotonsillectomy for childhood sleep apnoea on cardiorespiratory control," *ERJ Open Res.*, vol. 2, no. 2, pp. 1395–1404, 2016, doi: 10.1183/23120541.00003-2016.
- [68] "scipy.signal.resample — SciPy v1.4.1 Reference Guide." <https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.resample.html> (accessed Jun. 06, 2020).
- [69] J. Schmidhuber, "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61. Elsevier Ltd, pp. 85–117, Jan. 01, 2015, doi: 10.1016/j.neunet.2014.09.003.
- [70] D. C. Cireşan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," *IJCAI Int. Jt. Conf. Artif. Intell.*, pp. 1237–1242, 2011, doi: 10.5591/978-1-57735-516-8/IJCAI11-210.
- [71] Y. Bengio, I. J. Goodfellow, and A. Courville, "CHAPTER 9. Convolutional Networks," *Mit Press*, pp. 334–376, 2015, doi: 10.1007/978-3-319-94463-0_8.
- [72] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," 2012. Accessed: May 30, 2020. [Online]. Available: <http://code.google.com/p/cuda-convnet/>.
- [73] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.
- [74] R. Collobert and J. Weston, "A unified architecture for natural language processing," in *Proceedings of the 25th international conference on Machine learning - ICML '08*, 2008, pp. 160–167, doi: 10.1145/1390156.1390177.
- [75] H. Cecotti and A. Gräser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, 2011, doi: 10.1109/TPAMI.2010.125.
- [76] A. Page, C. Shea, and T. Mohsenin, "Wearable seizure detection using convolutional neural networks with transfer learning," in *Proceedings - IEEE International Symposium on Circuits and Systems*, Jul. 2016, vol. 2016-July, pp. 1086–1089, doi: 10.1109/ISCAS.2016.7527433.
- [77] A. Drouin-Picaro and T. H. Falk, "Using deep neural networks for natural saccade classification from electroencephalograms," in *2016 IEEE EMBS International Student Conference: Expanding the Boundaries of Biomedical Engineering and Healthcare, ISC 2016 - Proceedings*, Jul. 2016, doi: 10.1109/EMBSISC.2016.7508606.
- [78] "Deep Learning | Coursera." <https://www.coursera.org/specializations/deep-learning> (accessed

- May 30, 2020).
- [79] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical Evaluation of Rectified Activations in Convolutional Network," 2015, [Online]. Available: <http://arxiv.org/abs/1505.00853>.
- [80] "Convolutional Neural Network - Long - Medium." <https://medium.com/@Aj.Cheng/convolutional-neural-network-d9f69e473feb> (accessed Jun. 26, 2020).
- [81] B. Gao and L. Pavel, "On the Properties of the Softmax Function with Application in Game Theory and Reinforcement Learning," pp. 1–10, 2017, [Online]. Available: <http://arxiv.org/abs/1704.00805>.
- [82] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.
- [83] "Iterator Objects — Python 3.8.3 documentation." <https://docs.python.org/3/c-api/iterator.html> (accessed May 30, 2020).
- [84] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960, doi: 10.1177/001316446002000104.
- [85] M. Junghöfer, T. Elbert, D. M. Tucker, and C. Braun, "The polar average reference effect: A bias in estimating the head surface integral in EEG recording," *Clin. Neurophysiol.*, vol. 110, no. 6, pp. 1149–1155, Jun. 1999, doi: 10.1016/S1388-2457(99)00044-9.
- [86] J. Kayser and C. E. Tenke, "Issues and considerations for using the scalp surface Laplacian in EEG/ERP research: A tutorial review," *International Journal of Psychophysiology*, vol. 97, no. 3. Elsevier B.V., pp. 189–209, Sep. 01, 2015, doi: 10.1016/j.ijpsycho.2015.04.012.
- [87] D. Yao, "A method to standardize a reference of scalp EEG," *Physiol. Meas.*, vol. 22, pp. 693–711, 2001.
- [88] "Puntuaciones estandarizadas - MATLAB zscore - MathWorks España." <https://es.mathworks.com/help/stats/zscore.html> (accessed Jun. 02, 2020).
- [89] "tf.nn.dropout | TensorFlow Core v2.2.0." https://www.tensorflow.org/api_docs/python/tf/nn/dropout (accessed Jun. 03, 2020).
- [90] H. Wu and X. Gu, "Towards dropout training for convolutional neural networks," *Neural Networks*, vol. 71, pp. 1–10, Nov. 2015, doi: 10.1016/j.neunet.2015.07.007.
- [91] "42: A blog on A.I.!" https://nchlis.github.io/2017_08_10/page.html (accessed Jun. 03, 2020).
- [92] J. E. Heiss, C. M. Held, P. A. Estévez, C. A. Perez, C. A. Holzmänn, and J. P. Pérez, "Classification of sleep stages in infants: A neuro fuzzy approach," *IEEE Eng. Med. Biol. Mag.*, vol. 21, no. 5, pp. 147–151, 2002, doi: 10.1109/MEMB.2002.1044185.
- [93] V. Gerla, K. Paul, L. Lhotska, and V. Krajca, "Multivariate analysis of full-term neonatal polysomnographic data," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 1, pp. 104–110, 2009, doi: 10.1109/TITB.2008.2007193.
- [94] A. Piryatinska, G. Terdik, W. A. Woyczynski, K. A. Loparo, M. S. Scher, and A. Zlotnik, "Automated detection of neonate EEG sleep stages," *Comput. Methods Programs Biomed.*, vol. 95, no. 1, pp. 31–46, 2009, doi: 10.1016/j.cmpb.2009.01.006.
- [95] P. Chriskos, C. A. Frantidis, P. T. Gkivogkli, P. D. Bamidis, and C. Kourtidou-Papadeli, "Automatic

-
- Sleep Staging Employing Convolutional Neural Networks and Cortical Connectivity Images," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 31, no. 1, pp. 113–123, 2020, doi: 10.1109/TNNLS.2019.2899781.
- [96] C. Stepnowsky, D. Levendowski, D. Popovic, I. Ayappa, and D. M. Rapoport, "Scoring accuracy of automated sleep staging from a bipolar electroocular recording compared to manual scoring by multiple raters," *Sleep Med.*, vol. 14, no. 11, pp. 1199–1207, Nov. 2013, doi: 10.1016/j.sleep.2013.04.022.
- [97] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning Deep Representation for Imbalanced Classification," 2016.