



Facultade de Informática

UNIVERSIDADE DA CORUÑA

TRABALLO FIN DE GRAO
GRAO EN ENXEÑARÍA INFORMÁTICA
MENCIÓN EN COMPUTACIÓN

Transferencia de aprendizaje en la clasificación de estados de sueño

Estudiante: Diego Pérez Ramos

Dirección: Elena María Hernández Pereira

A Coruña, xuño de 2020.

Agradecimientos

Me gustaría agradecer a mi familia por apoyarme y animarme en todo momento. También me gustaría agradecer a Elena María Hernández Pereira por su ayuda a lo largo de la realización de este trabajo.

Resumen

La clasificación de estados de sueño es una rama de la medicina del sueño que se encarga de detectar anomalías en el sueño de un individuo, utilizando para ello información de señales como EEG, EOG y EMG. En los últimos años han sido muchos los trabajos publicados que tratan de automatizar esta tarea de clasificación. En ellos se han utilizado distintas técnicas, entre las que destaca el Aprendizaje Máquina. Concretamente a través del Aprendizaje Profundo se han conseguido buenos resultados. En este trabajo, se aplica transferencia de aprendizaje para clasificar fases de sueño. Para ello, partiendo de un modelo de Aprendizaje Profundo basado en redes convolucionales, se realiza un ajuste detallado del mismo para llevar a cabo dicha clasificación sobre distintos conjuntos de datos.

Abstract

The classification of sleep stages is a branch of sleep medicine that aims to detect anomalies in someone's sleep, using information from signals such as EEG, EOG and EMG. In recent years there have been many published works that attempt to automate this classification task. In this works, different techniques have been used, among which Machine Learning stands out. Specifically, through Deep Learning, good results have been achieved. In this work, transfer learning is applied to classify sleep stages. For this, from a Deep Learning model based on convolutional networks, a detailed adjustment is made to perform this classification on different data sets.

Palabras clave:

- Aprendizaje Máquina
- Aprendizaje Profundo
- Transferencia de aprendizaje
- Fases del sueño

Keywords:

- Machine learning
- Deep learning
- Transfer learning
- Sleep stages

Índice general

1	Introducción	1
2	Descripción del dominio	3
2.1	El polisomnograma	3
2.2	Fases del sueño	5
3	Estado del arte	7
4	Metodología y planificación	11
4.1	Metodología utilizada	11
4.2	Planificación	11
4.2.1	Estimación de costes del proyecto	12
4.3	Métricas utilizadas	13
4.4	Entorno y herramientas	15
4.5	Conjuntos de datos	15
4.5.1	PhysioNet Sleep-EDF Database	15
4.5.2	ISRUC-Sleep Dataset	16
4.5.3	SHHS PSG Database	16
5	Desarrollo	19
5.1	Modelo de referencia	19
5.2	Preprocesado de datos	21
5.3	Escenarios de transferencia de aprendizaje	23
5.3.1	Escenario 1: Modelos preentrenados	23
5.3.2	Escenario 2: Ajuste de la capa softmax	23
5.3.3	Escenario 3: Ajuste del último bloque convolucional	23
5.3.4	Escenario 4: Ajuste de toda la red	24

6	Resultados y evaluación	25
6.1	Resultados de la transferencia de aprendizaje	25
6.1.1	Physionet-minority	25
6.1.2	ISRUC-minority	26
6.1.3	Physionet-notmajority	27
6.1.4	ISRUC-notmajority	28
6.1.5	Conclusiones de la transferencia de aprendizaje	29
6.2	Visualización de los resultados	30
6.2.1	Mapas de prominencia para el conjunto Physionet	30
6.2.2	Mapas de prominencia para el conjunto ISRUC	32
6.3	Comparación con otros trabajos	34
7	Conclusiones y trabajo futuro	37
7.1	Conclusiones	37
7.2	Trabajo futuro	38
	Lista de acrónimos	41
	Glosario	43
	Bibliografía	45

Índice de figuras

2.1	Ejemplo de un registro polisomnográfico	4
2.2	Ejemplo de un hipnograma siguiendo el estándar R&K	6
4.1	Diagrama de Gantt	12
5.1	Matriz de confusión resultante de la evaluación del conjunto de prueba	21
5.2	Ajustes de los distintos escenarios para un modelo de ejemplo	24
6.1	Ejemplo de mapa de prominencia para la fase W	30
6.2	Ejemplo de mapa de prominencia para la fase N1	31
6.3	Ejemplo de mapa de prominencia para la fase N2	31
6.4	Ejemplo de mapa de prominencia para la fase N3	31
6.5	Ejemplo de mapa de prominencia para la fase REM	32
6.6	Ejemplo de mapa de prominencia para la fase W	32
6.7	Ejemplo de mapa de prominencia para la fase N1	33
6.8	Ejemplo de mapa de prominencia para la fase N2	33
6.9	Ejemplo de mapa de prominencia para la fase N3	33
6.10	Ejemplo de mapa de prominencia para la fase REM	34

Índice de tablas

2.1	Rangos de frecuencia y amplitud de las ondas cerebrales	4
4.1	Coste estimado del ingeniero	13
4.2	Coste estimado del material	13
4.3	Resumen de los conjuntos de datos	16
4.4	Distribución de las clases del conjunto de datos de Physionet	17
4.5	Distribución de las clases del conjunto de datos de ISRUC-Sleep	17
4.6	Distribución de las clases del conjunto de datos de SHHS	17
5.1	Distribuciones para los distintos hiperparámetros	20
5.2	Hiperparámetros de los cinco mejores modelos	20
5.3	Resultados de la evaluación del conjuntos de prueba	21
5.4	Distribución de las clases de ISRUC-Sleep balanceado con la estrategia 'minority'	22
5.5	Distribución de las clases de Physionet balanceado con la estrategia 'minority'	22
5.6	Distribución de las clases de ISRUC-Sleep balanceado con la estrategia 'not majority'	23
5.7	Distribución de las clases de Physionet balanceado con la estrategia 'not majority'	23
6.1	Medidas de rendimiento para el conjunto Physionet-minority	26
6.2	Medidas de rendimiento del escenario 4 para el conjunto Physionet-minority	26
6.3	Medidas de rendimiento para el conjunto ISRUC-minority	27
6.4	Medidas de rendimiento del escenario 4 para el conjunto ISRUC-minority	27
6.5	Medidas de rendimiento para el conjunto Physionet-notmajority	28
6.6	Medidas de rendimiento del escenario 4 para el conjunto Physionet-notmajority	28
6.7	Medidas de rendimiento para el ISRUC-notmajority	29
6.8	Medidas de rendimiento del escenario 4 para el conjunto ISRUC-notmajority	29
6.9	Comparación para el conjunto de datos Physionet	35

6.10 Comparación para el conjunto de datos ISRUC 35

Introducción

LA palabra sueño se puede referir tanto al estado de reposo en que se encuentra un individuo mientras duerme como a la actividad cerebral durante este período. El sueño es esencial para la vida humana, permite descansar tanto física como mentalmente, además de cumplir con ciertas funciones fisiológicas como el aumento de la segregación de la hormona del crecimiento. Además, durante este período se producen en nuestro organismo cambios de temperatura, en el pulso o en la respiración y variaciones importantes de actividad cerebral.

El sueño ha sido siempre una de las condiciones humanas de mayor interés, lo que explica su estudio desde hace ya bastante tiempo. En un principio se relacionaban los sueños con presentimientos o presagios y aspectos mágicos o religiosos, pero con la llegada del encefalograma y la neurología se empezó a investigar el sueño más seriamente. Tanto la curiosidad humana como la necesidad de comprender este fenómeno y su importante relación con nuestra actividad durante el día, han sido motivos suficientes para realizar estudios sobre el sueño. Muchas de las cuestiones planteadas por la humanidad acerca del sueño ya tienen sus respuestas. Pese a esto y los grandes avances realizados, hoy en día siguen surgiendo nuevas preguntas.

Los trastornos del sueño son una patología muy frecuente en la población humana y su estudio es muy reciente. Hasta la fecha se han descubierto más de noventa enfermedades del sueño. Aunque no son graves en su mayoría, pueden tener implicaciones en nuestras vidas como cansancio, bajo rendimiento, problemas cardiovasculares, alteraciones en el organismo o derivar en enfermedades como el insomnio, el síndrome de la apnea del sueño o las hipersomnias [1].

El diagnóstico de los trastornos del sueño está basado en la realización y el análisis de un registro denominado polisomnograma. A partir de este, el médico puede confirmar la existencia de alguna patología que afecte al ciclo normal de sueño de una persona aparentemente sana.

La polisomnografía es una técnica que estudia el sueño mediante el registro de varias se-

ñales fisiológicas. Este registro se realiza mientras el paciente duerme y tiene una duración aproximada de ocho horas. El electroencefalograma, el electromiograma y el electrooculograma son las señales básicas de cualquier polisomnografía. Los expertos analizan manualmente las señales obtenidas para determinar las fases por las que transcurre el sueño del paciente, así como los eventos que hayan surgido durante el mismo. Esta tarea es tediosa y consume mucho tiempo por lo que, para tratar de aliviar el trabajo del experto, se han ido desarrollando, a lo largo de las últimas décadas, sistemas automáticos de clasificación de fases del sueño.

La utilización de técnicas de inteligencia artificial y más concretamente, de técnicas de aprendizaje máquina, ha sido crucial en el desarrollo de sistemas automáticos de clasificación de fases del sueño. Existen numerosas aproximaciones que consiguen resultados comparables con los del experto. En los últimos años y gracias a los avances computacionales, se ha incorporado el aprendizaje profundo a esta tarea, consiguiendo una notable mejora en los sistemas de clasificación de fases del sueño. Los sistemas basados en aprendizaje profundo requieren grandes cantidades de datos así como una elevada potencia de cálculo.

En el campo del aprendizaje profundo la transferencia de aprendizaje es muy popular, consiste en la reutilización de un modelo, desarrollado y entrenado con un gran conjunto de datos, para una nueva tarea relacionada. La aplicación de transferencia de aprendizaje a la clasificación de fases de sueño puede evitar la necesidad de tal volumen de datos y el elevado coste computacional del aprendizaje profundo.

El objetivo de este trabajo es analizar los resultados obtenidos por la transferencia de aprendizaje sobre un modelo de clasificación de fases de sueño. La estructura que seguirá la memoria es la siguiente. En el segundo capítulo se realiza una descripción del contexto del trabajo. En el tercer capítulo se realiza un análisis de antecedentes con vistas a establecer un marco de comparación con el trabajo desarrollado, en la medida de lo posible. En el cuarto capítulo se describen los fundamentos tecnológicos y la metodología utilizada. En el quinto capítulo se realiza una descripción del modelo y de los escenarios de transferencia de aprendizaje desarrollados. El sexto capítulo analiza de forma global los resultados obtenidos y, para finalizar, un capítulo de conclusiones en el que se analiza el trabajo desde un punto de vista crítico aportando ideas para mejorarlo o ampliarlo en un futuro.

Descripción del dominio

LA medicina del sueño es una rama de la medicina que lleva a cabo el diagnóstico y tratamiento de los trastornos del sueño. Para llevar a cabo estas tareas es necesario realizar un estudio exhaustivo del comportamiento del paciente durante el sueño. La técnica más utilizada en este caso es la polisomnografía (PSG), que consiste en el registro simultáneo de un conjunto de señales fisiológicas que van a permitir analizar la estructura del sueño del paciente. Esta técnica proporciona una gran cantidad de información, que permite un análisis completo de la evolución del sueño del paciente y favorece la elaboración de un diagnóstico.

2.1 El polisomnograma

El polisomnograma es una prueba que se realiza durante el período habitual de sueño del paciente en un laboratorio del sueño u, ocasionalmente, en el domicilio del mismo. Durante esta prueba se registran un conjunto de señales neurofisiológicas, señales respiratorias e información de contexto. La figura 2.1 muestra un fragmento de un registro polisomnográfico. Para la identificación de las distintas fases de sueño, las señales utilizadas principalmente son las señales neurofisiológicas.

Señales neurofisiológicas

Estas señales representan las funciones neurofisiológicas del individuo, siendo las más utilizadas el electroencefalograma (EEG), el electromiograma (EMG) y el electrooculograma (EOG).

- Electroencefalograma. Es la señal neurofisiológica más compleja. Normalmente se analiza el espectro de frecuencias dominantes de la señal dividiéndolo en cuatro categorías de ondas (alpha, beta, theta y delta), que nos dan información de la fase del sueño. Habitualmente las amplitudes de esta señal se encuentran entre 5 y 200 μV , con frecuencias

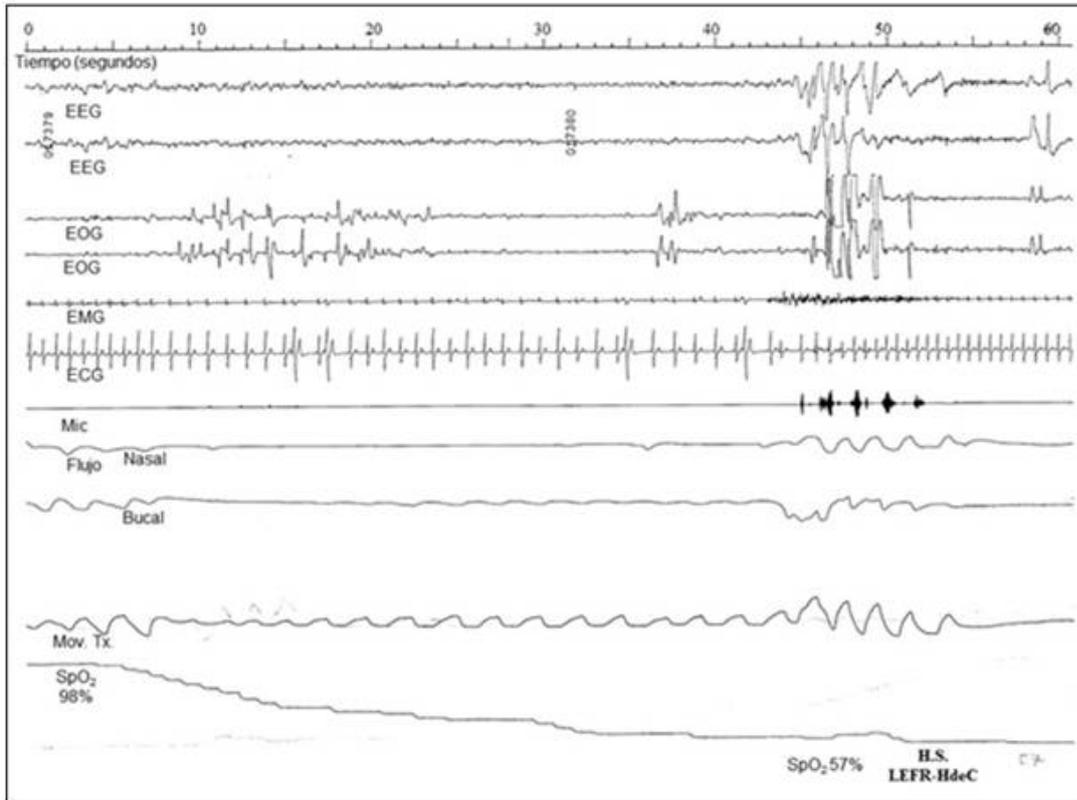


Figura 2.1: Ejemplo de un registro polisomnográfico

entre 1 y 30 Hz. En la tabla 2.1 vemos una división de las ondas según sus rangos de frecuencia y amplitud.

Onda	Rango de frecuencia	Rango de amplitud
Beta (β)	>16 Hz	5-50 μV
Alpha (α)	8-13 Hz	5-50 μV
Theta (θ)	4-8 Hz	50-100 μV
Delta (δ)	<4 Hz	100-200 μV

Tabla 2.1: Rangos de frecuencia y amplitud de las ondas cerebrales

- Electromiograma. Es la señal que registra la actividad neuromuscular durante el sueño, a través de electrodos situados en la superficie de los músculos del área submental y de las piernas. Habitualmente las amplitudes de esta señal se encuentran entre -100 y 100 μV , con frecuencias entre 100 y 500 Hz.

- Electrooculograma. Esta señal representa la diferencia de potencial entre la retina y la cornea del ojo humano y nos proporciona información sobre el movimiento ocular durante las distintas fases del sueño. Habitualmente las amplitudes de esta señal oscilan entre 20 y 250 μV , con frecuencias entre 100 y 500 Hz.

2.2 Fases del sueño

La estructura del sueño se divide en ciclos de aproximadamente 90 minutos que se repiten durante el tiempo que este dure. Existen dos estándares de clasificación de fases del sueño. Según el estándar Rechtschaffen y Kales (R&K) [2] los ciclos de sueño se dividen en seis fases. Por otro lado, la Academia Americana de Medicina del Sueño (American Academy of Sleep Medicine, AASM) [3] establece que cada uno de los ciclos de sueño está dividido en cinco fases. Este segundo estándar se considera una evolución del estándar R&K y es el que se utiliza en la actualidad como guía para la clasificación de estados de sueño que a continuación se describen¹:

- Fase W (fase W). Es una fase de transición en la que el individuo pasa de un estado de vigilia a estar adormecido. La señal de EEG revela la presencia de ritmos alpha de baja amplitud, entre 5 y 50 μV , con frecuencias de 8 a 13 Hz, que se atenúan con los ojos abiertos. El EOG muestra parpadeos rápidos en los ojos con frecuencias de 0.5 a 2 Hz. El EMG muestra amplitudes variables pero mayores que las que aparecen en las fases siguientes.
- Fase N1 (fase 1). También conocida como fase de sueño ligero. Durante esta, tanto la respiración como el ritmo cardíaco se van ralentizando y el sujeto es capaz de percibir la mayor parte de los estímulos externos. En esta fase la señal de EEG presenta una disminución de los ritmos alpha, con amplitudes bajas y frecuencias de entre 4 y 8 Hz. La señal de EOG revela movimientos oculares lentos. La señal de EMG muestra amplitudes menores que en la fase anterior.
- Fase N2 (fase 2). Durante esta fase nos encontramos en un estado de relajación profunda. En el EEG aparecen ondas theta, con amplitudes bajas y frecuencias de entre 11 y 16 Hz. En el EOG no aparecen movimientos oculares y el EMG muestra amplitudes variables pero menores que en las anteriores fases.
- Fase N3 (fases 3 y 4). También llamada fase de sueño profundo o de sueño delta. Determina la calidad del descanso. Durante esta fase el individuo se encuentra en un estado

¹Entre paréntesis se indica la fase correspondiente del estándar R&K.

Estado del arte

DESDE hace tiempo, la clasificación automática de estados de sueño ha sido objeto de estudio y ha estado ligada al campo del aprendizaje máquina. Existen multitud de trabajos que lo demuestran y a continuación se describen algunos de los que tienen relación con el desarrollo de este trabajo fin de grado. En primer lugar se describen varios trabajos de clasificación de etapas de sueño utilizando aprendizaje automático, después se presentan varios trabajos que utilizan aprendizaje profundo, para finalizar con una descripción de trabajos en los que se utiliza transferencia de aprendizaje.

En [4] S. Gudmundsson et al. usan un clasificador basado en máquinas de soporte vectorial (SVM, Support Vector Machines). El conjunto de datos empleado consta de cuatro registros clasificados por expertos siguiendo el estándar R&K. Debido al desbalanceo de clases, los autores unifican las fases 1 y 2 en la fase LS (Light Sleep) y las fases 3 y 4 en la fase SWS (Slow Wave Sleep). Para este trabajo usan un único canal de EEG, muestreado a 256 Hz. A esta señal se le aplica un filtro paso banda (0.5-70 Hz) y un filtro notch (50 Hz). Para la SVM los parámetros son seleccionados usando validación cruzada cinco veces. Utilizan tres métodos de extracción de características distintos, medidas de complejidad de Hjorth, características del espectro de frecuencia y método del histograma. Los autores utilizan cuatro conjuntos de test distintos. Los resultados en términos de precisión para dos de estos conjuntos son de 94% para la fase W, 69,7% para la fase LS, 96,9% para la fase SWS y 83,3% para la fase REM en el primer conjunto de test; 48% para la fase W, 95,6% para la fase LS, 88,1% para la fase SWS y 75,8% para la fase REM en el segundo conjunto de test.

En [5] F. Ebrahimi et al. usan una red de neuronas artificiales (ANN) cuyas entradas se obtienen a partir de los coeficientes Wavelet de la señal de EEG, usando un único canal muestreado a 100 Hz. Los datos utilizados constituyen un subconjunto de PhysioBank [6] [7] de siete registros, anotados por expertos siguiendo el estándar R&K. Debido al desbalanceo, las clases se agrupan de la siguiente forma: fase Despierto (W), fase 1+REM, fase 2 y fase SWS (fases 3 y 4). Los resultados obtenidos en términos de precisión son de 98,5%, 91,4%, 87,5% y

94,6% para cada una de estas fases respectivamente.

En el campo del aprendizaje profundo, O. Tsinalis et al. en [8] utilizan una red convolucional (CNN) para clasificar estados de sueño. El modelo desarrollado está compuesto por dos pares de capas de convolución y pooling, dos capas fully-connected y una capa de salida softmax. El conjunto de datos usado en este trabajo consta de 39 registros polisomnográficos, procedentes de la base de datos de PhysioNet, anotados por expertos siguiendo el estándar R&K. Los autores adaptaron dichas anotaciones al estándar de la AASM. Para este trabajo usan un único canal de EEG, muestreado a 100 Hz. Los resultados obtenidos en términos de precisión son de 61% para la fase W, 34,4% para la fase N1, 89,7% para la fase N2, 79,4% para la fase N3 y 75,4% para la fase REM. En términos de sensibilidad son de 70,5% para la fase W, 60% para la fase N1, 73,2% para la fase N2, 91,2% para la fase N3 y 73,7% para la fase REM.

En [9] A. Supratak et al. evalúan un modelo de aprendizaje profundo sobre dos conjuntos de datos: Montreal Archive of Sleep Studies (MASS) [10] y Sleep-EDF [6] [7]. Del primer conjunto utilizaron 62 registros, anotados según el estándar AASM, a través de un único canal de EOG muestreado a 256 Hz. Del segundo conjunto, anotado según el estándar AASM, utilizaron 20 registros a través de dos canales de EEG muestreados a 100 Hz. La arquitectura del modelo empleado está dividida en dos partes, el aprendizaje de representación y el aprendizaje residual secuencial. La primera aprende filtros de extracción de características invariables al tiempo y está formada por dos CNNs. Cada CNN consiste en cuatro capas convolucionales y dos capas max-pooling. Cada capa convolucional realiza una convolución 1D, una normalización y aplica una función de activación ReLU. La segunda parte aprende la información temporal, es decir, recuerda las características de la entrada anterior para clasificar la actual. Para ello usan dos capas LSTM bidireccional y una conexión de acceso directo. Esta última contiene una capa fully-connected que transforma las características extraídas en un vector compatible con la salida de las capas LSTM bidireccional. Además, esta capa realiza una normalización y aplica una función de activación ReLU. Para las fases W, N1, N2, N3 y REM consiguen unos resultados de 87,3%, 60,4%, 89,8%, 83,8% y 88,4% en términos de precisión y 87,2%, 59,3%, 90,7%, 79,4% y 90,2% en términos de sensibilidad evaluando el canal de EOG; evaluando uno de los dos canales de EEG consiguen valores de 86%, 43,5%, 90,5%, 77,1% y 80,9% en términos de precisión y 83,4%, 50,1%, 81,7%, 94,2% y 83,9% en términos de sensibilidad.

En [11] H. Phan et al. clasifican una secuencia de varias etapas de sueño. Para ello proponen un modelo en el que la entrada es una secuencia de imágenes que computa una secuencia de salidas del mismo tamaño. Para este trabajo usan 200 registros polisomnográficos de la base de datos MASS, anotados de acuerdo al estándar de la AASM. Los canales utilizados fueron un canal de EEG, la media de los dos canales de EOG y la media de los dos canales de EMG, todos ellos remuestreados a 100 Hz. La arquitectura propuesta se divide en cuatro bloques. El primer bloque son capas secuenciales de bancos de filtros, cuya función es aprender

filtros en el dominio de la frecuencia, para suavizar y de esta forma reducir la dimensión de las imágenes de entrada. Debido a las diferentes características de las señales de EEG, EOG y EMG, usan capas de bancos de filtros separadas para cada una. El segundo bloque consta de redes de neuronas recurrentes (RNN) bidireccionales junto con el mecanismo de atención, para aprender características de las señales. El tercer bloque está formado por RNNs bidireccionales seguidas de capas de atención. En este bloque, a partir de la secuencia de vectores de características computadas anteriormente, se crea un único vector de características para cada etapa de la secuencia de entrada. Finalmente, en el último bloque de la red usan RNN bidireccionales secuenciales seguidas de capas softmax que computan las salidas. Para las fases W, N1, N2, N3 y REM consiguen unos resultados de 90%, 65%, 89,1%, 84% y 90,8% en términos de precisión y 89,4%, 60,9%, 90,7%, 80,3% y 91,9% en términos de sensibilidad.

Finalmente, y dentro del campo de transferencia de aprendizaje en la clasificación de etapas de sueño, en [12] H. Phan et al. trabajan sobre los conjuntos de datos MASS y Sleep-EDF con 200 y 20 registros respectivamente. Todos estos registros han sido anotados siguiendo el estándar de la AASM. El primer conjunto utiliza un canal de EEG y el segundo conjunto, un canal de EEG y uno de EOG, todos ellos muestreados a 100 Hz. El modelo utilizado para hacer transferencia de aprendizaje es SeqSleepNet [11] para un único canal. Este modelo se entrena utilizando el conjunto MASS, ya que dispone de un mayor número de registros; y se realiza la transferencia sobre el conjunto Sleep-EDF. En este trabajo estudian dos escenarios. En el primero, el modelo es entrenado con el canal de EEG, para luego transferir el conocimiento sobre el canal de EEG de Sleep-EDF. En el segundo escenario se entrena el modelo con el canal de EEG de MASS, para a continuación transferir el conocimiento al canal de EOG de Sleep-EDF. Los mejores resultados obtenidos en este trabajo son de 85,5% de precisión media, 79,6% de sensibilidad media y 80% de F1 score medio.

En [13] A. A. Gharbali et al. utilizan la red AlexNet [14] para realizar transferencia de aprendizaje sobre un conjunto de 10 registros procedente de la base de datos de ISRUC-Sleep [15]. Utilizan un único canal de EEG, muestreado a 200 Hz sobre el que aplican un filtro paso banda (0.3 - 25 Hz) y un filtro notch (50 Hz). Los registros están anotados siguiendo el estándar de la AASM. Usando la transformada Wavelet continua (CWT), extraen el espectro de frecuencia de la señal de EEG y lo transforman en imágenes que se clasifican usando AlexNet. AlexNet está formada por 25 capas, ocho de las cuales tienen pesos que se pueden aprender, de las cuales cinco son capas convolucionales y tres son capas fully-connected. En este trabajo, para las fases W, N1, N2, N3 y REM, consiguen obtener una precisión de 96,3%, 67,6%, 84,1%, 89,2% y 82,4% y una sensibilidad de 92%, 64%, 79%, 85% y 78% para cada fase respectivamente.

Metodología y planificación

EN este capítulo describiremos la metodología utilizada a lo largo de este trabajo así como la planificación seguida, el coste estimado del proyecto y las métricas de rendimiento empleadas. Finalizaremos con los capítulos de entorno y herramientas utilizadas y conjuntos de datos.

4.1 Metodología utilizada

Este trabajo ha sido desarrollado utilizando la metodología de desarrollo en espiral [16] [17]. Esta metodología consta de un bucle principal para la ejecución de cada fase del trabajo. Las tareas a realizar en cada iteración del bucle son:

1. **Análisis de objetivos:** consiste en determinar los objetivos que queremos alcanzar, teniendo en cuenta los resultados obtenidos en la iteración anterior del bucle, si esta existe.
2. **Planificación y análisis de riesgos:** consiste en determinar la forma en la que se van a tratar de alcanzar los objetivos propuestos y determinar posibles amenazas que se puedan producir.
3. **Implementación:** consiste en llevar a cabo las tareas determinadas durante la planificación para conseguir los objetivos y, de esta forma, obtener unos resultados.
4. **Evaluación:** consiste en analizar los resultados obtenidos durante la implementación y determinar la necesidad o no de una nueva iteración de la fase actual.

4.2 Planificación

La planificación de este trabajo esta organizada en fases o tareas que son ejecutadas siguiendo la metodología vista anteriormente. Dichas fases deben ser llevadas a cabo en el

siguiente orden:

1. Estudio del dominio de trabajo: la medicina del sueño.
2. Estudio del aprendizaje profundo, transferencia de aprendizaje y sus aplicaciones a la medicina del sueño.
3. Estudio y preprocesado de los conjuntos de datos.
4. Implementación del modelo de referencia para la transferencia de aprendizaje.
 - 4.1. Implementación del modelo base.
 - 4.2. Búsqueda de los mejores hiperparámetros para el modelo.
 - 4.3. Entrenamiento y estudio de los resultados.
5. Realización de la transferencia de aprendizaje.
 - 5.1. Implementación de los distintos escenarios de transferencia de aprendizaje.
 - 5.2. Estudio de los resultados de cada escenario.
6. Evaluación, análisis y comparación de los resultados.

En la figura 4.1 podemos observar el diagrama de Gantt de la planificación detallada por semanas de este trabajo.



Figura 4.1: Diagrama de Gantt

4.2.1 Estimación de costes del proyecto

La estimación de costes de este proyecto se llevará a cabo asumiendo la participación de un único ingeniero con poca experiencia, con un salario de 12,5€ por hora. El salario por hora ha sido calculado partiendo de un sueldo anual de 24.000€. En cuanto al coste del material, se tiene

en cuenta los gastos de internet y el coste del equipo. Tanto los conjuntos de datos utilizados como las herramientas son de libre uso por lo que no influyen en el coste del proyecto. En la tabla 4.1 se observa la distribución de horas y el coste de cada fase para el ingeniero. En la tabla 4.2 se observa el coste del material necesario durante el proyecto. El coste total estimado del proyecto es de 5.450€.

Fase	Horas	Coste (€)
1	15	187,5
2	15	187,5
3	50	625
4.1	55	687,5
4.2	45	562,5
4.3	25	312,5
5.1	50	625
5.2	30	375
6	45	562,5
Total	330	4.125

Tabla 4.1: Coste estimado del ingeniero

Material	Coste (€)
Internet	125
Equipo informático	1.200
Total	1.325

Tabla 4.2: Coste estimado del material

4.3 Métricas utilizadas

A lo largo de este trabajo es muy importante ser capaz, en todo momento, de evaluar unos resultados y determinar la calidad de estos. Para ello se hace uso de las siguientes métricas:

- Exactitud: La exactitud se refiere a la proporción de ejemplos clasificados correctamente entre el número total de ejemplos.

$$exactitud = \frac{VP + VN}{VP + VN + FP + FN}$$

Donde VP se corresponde con los verdaderos positivos, VN con los verdaderos negativos, FP con los falsos positivos y FN con los falsos negativos.

- Precisión: La precisión se refiere a la proporción de ejemplos clasificados correctamente como positivos entre el número total de ejemplos clasificados como positivos.

$$precisión = \frac{VP}{VP + FP}$$

- Sensitividad: La sensitividad, también conocida como tasa de verdaderos positivos, se refiere a la proporción de ejemplos de una clase clasificados correctamente como positivos entre el número total de ejemplos de dicha clase.

$$sensitividad = \frac{VP}{VP + FN}$$

- Medida F1: La métrica de F1 score es la media armónica entre la precisión y la sensitividad, y resulta de gran ayuda cuando nos encontramos ante conjuntos de datos desbalanceados.

$$F1score = 2 * \frac{(sensitividad * precisión)}{(sensitividad + precisión)}$$

- Macro F1: La métrica macro F1 hace referencia a la media de F1 score de todas las clases.

$$macroF1 = \frac{\sum_{c=1}^M F1score_c}{M}$$

Donde M es el número de clases y $F1score_c$ es el valor de F1 para la clase c .

- Índice kappa: el índice kappa es una métrica que mide el acuerdo entre dos clasificadores.

$$kappa = \frac{p_o - p_e}{1 - p_e}$$

Donde p_o se refiere al acuerdo observado entre los clasificadores y p_e se refiere a la probabilidad de acuerdo por azar entre estos.

- Función de pérdida: la función de pérdida es una métrica para evaluar las diferencias entre las predicciones y los resultados reales de un modelo. Durante el entrenamiento de una red neuronal el objetivo es minimizar dicha función. A lo largo de este trabajo se usa la función de pérdida entropía cruzada categórica. Habitual en problemas de clasificación donde un único resultado es correcto.

$$p\acute{e}rdida = - \sum_{c=1}^M y_{o,c} \log(p_{o,c})$$

Donde M es el número de clases, $y_{o,c}$ es 0 ó 1 e indica si la observación o está correctamente clasificada como clase c y $p_{o,c}$ es la probabilidad de pertenencia a la clase c de la observación o .

4.4 Entorno y herramientas

Para la elaboración del trabajo se han usado distintas herramientas y librerías que facilitan el desarrollo del mismo. El proyecto ha sido realizado en el lenguaje de programación Python (versión 3.7) [18] [19] debido a su simplicidad y a la gran cantidad de librerías y herramientas que facilitan el manejo de datos, la creación de modelos de aprendizaje máquina, etc. El entorno de programación empleado es PyCharm, un IDE de Python que facilita la instalación de las distintas librerías. Las librerías utilizadas son Keras [20] y TensorFlow [21] para el manejo de redes neuronales, MNE [22], pandas [23] y Visbrain [24] para la lectura y manejo de los datos, Numpy para el manejo de vectores e imbalanced-learn [25] para el balanceo de las clases. Para la búsqueda de hiperparámetros se usa la librería Hyperopt [26]. Además, para la visualización de los mapas de prominencia (saliency maps) se ha usado la herramienta tf-keras-vis [27] junto con Matplotlib [28].

4.5 Conjuntos de datos

Para poder realizar transferencia de aprendizaje es necesario disponer de distintos conjuntos de datos. A continuación se describen los conjuntos de datos públicos utilizados en el trabajo.

4.5.1 PhysioNet Sleep-EDF Database

PhysioNet Sleep-EDF [6] [7] es una base de datos que contiene 197 registros polisomnográficos. En este trabajo usaremos el conjunto de datos perteneciente al estudio Sleep Cassette que contiene 153 registros con una duración aproximada de 20 horas cada uno. La población del estudio incluye 77 individuos, hombres y mujeres, de entre 25 y 101 años sin medicación relacionada con el sueño. Cada registro contiene dos canales de EEG, un canal de EOG y un canal de EMG. Los canales de EEG y EOG están muestreados a 100 Hz mientras que el canal de EMG está muestreado a 1 Hz. Esta base de datos dispone de ficheros de anotaciones con la clasificación de fases de sueño para cada uno de los registros polisomnográficos. Dicha clasificación ha sido realizada por expertos en medicina del sueño siguiendo el estándar R&K.

4.5.2 ISRUC-Sleep Dataset

ISRUC-Sleep [15] es un conjunto de datos que contiene 100 registros polisomnográficos con duración aproximada de ocho horas cada uno. La población de este estudio incluye hombres y mujeres de entre 20 y 85 años, con evidencias de trastornos del sueño. Cada registro contiene seis canales de EEG, dos canales de EOG y tres canales de EMG, entre otras señales. Todos los canales de EEG, EOG y EMG han sido muestreados a 200 Hz. Las anotaciones han sido realizadas por dos expertos siguiendo el estándar AASM.

4.5.3 SHHS PSG Database

Sleep Heart Health Study (SHHS) PSG [7] [29] es una base de datos que contiene registros polisomnográficos de individuos adultos mayores de 40 años. Del conjunto total de registros se utilizarán 992 de aproximadamente 11 horas de duración. Dichos registros contienen dos canales de EEG, dos canales de EOG y un canal de EMG. Los canales de EEG y EMG están muestreados a 125 Hz mientras que los canales de EOG están muestreados a 50 Hz. Las anotaciones han sido realizadas por expertos siguiendo el estándar de la AASM.

La tabla 4.3 recoge la información resumida de los conjuntos de datos descritos en términos de cantidad de registros, duración de los mismos, señales registradas y estándar seguido por las anotaciones.

Conjuntos de datos	Registros	Horas/Registro	EEG	EOG	EMG	Estándar
Physionet	153	20	100 Hz	100 Hz	1 Hz	R&K
ISRUC-Sleep	100	8	200 Hz	200 Hz	200 Hz	AASM
SHHS Dataset	992	11	125 Hz	50 Hz	125 Hz	AASM

Tabla 4.3: Resumen de los conjuntos de datos

En las tablas 4.4, 4.5 y 4.6 se puede observar la distribución de las distintas clases de cada conjunto de datos. Podemos observar que el conjunto de datos de ISRUC-Sleep tiene una distribución de clases bastante balanceada con respecto a los otros dos conjuntos. En general, las clases N1 y N3 son las que menos presencia tienen, mientras que las clases W y N2 son las más abundantes.

Fase	W	N1	N2	N3	REM	Total
Nº Ejemplos	65.795	21.469	68.633	12.991	25.767	194.655
Proporción	33,8%	11,03%	35,26%	6,67%	13,24%	100%

Tabla 4.4: Distribución de las clases del conjunto de datos de Physionet

Fase	W	N1	N2	N3	REM	Total
Nº Ejemplos	20.979	11.513	28.287	17.480	11.928	90.187
Proporción	23,26%	12,77%	31,36%	19,38%	13,23%	100%

Tabla 4.5: Distribución de las clases del conjunto de datos de ISRUC-Sleep

Fase	W	N1	N2	N3	REM	Total
Nº Ejemplos	452.173	44.602	434.026	112.756	150.838	1.194.395
Proporción	37,86%	3,73%	36,34%	9,44%	12,63%	100%

Tabla 4.6: Distribución de las clases del conjunto de datos de SHHS

El conjunto de datos de SHHS será el conjunto utilizado para entrenar el modelo de referencia y los conjuntos de datos ISRUC-Sleep y Physionet serán los conjuntos sobre los que se realizará transferencia de aprendizaje.

Desarrollo

EN este capítulo describiremos tanto el modelo de referencia basado en Aprendizaje Profundo, el preprocesado realizado sobre los datos, así como los distintos escenarios de transferencia de aprendizaje que se llevarán a cabo en este trabajo.

5.1 Modelo de referencia

Para realizar transferencia de aprendizaje es necesario tener un modelo que haya sido previamente entrenado con un conjunto de datos relativamente grande. Para este trabajo partiremos de un modelo basado en el que se describe en I. Fernández-Varela et al. [30]. Dicho modelo ha sido entrenado con los datos de Sleep Heart Health Study (SHHS) utilizando dos canales de EEG, dos canales de EOG y un canal de EMG, remuestreados a 125 Hz. El modelo está basado en un ensemble de cinco modelos convolucionales, que recibe como entrada una ventana de 30 segundos de cada una de las señales mencionadas. El primer bloque de cada uno de los modelos es el bloque convolucional y está formado por una capa convolucional 1D, una capa de normalización por lotes, una capa de activación y capa una de average pool con factor de reducción de 2. El tamaño del kernel es el mismo para todas las capas pero el número de filtros se va duplicando en cada capa hasta un máximo de 1024. Este bloque convolucional se repite un número de veces en función de cada uno de los modelos del ensemble. La salida de este bloque se utiliza como entrada en una capa densa con activación softmax que devuelve la probabilidad de pertenencia a cada clase.

A diferencia del esquema original [30], donde se usan cinco canales, y tras haber comprobado que los resultados obtenidos son similares, se entrenará un modelo siguiendo dicho esquema con la información de tres canales: uno de EEG, uno de EOG y uno de EMG. De esta manera se reduce la carga computacional y los tiempos de entrenamiento. A las señales de EEG y EMG se aplica un filtro notch (60 Hz) para eliminar interferencias y de la señal de EMG, se eliminan aquellas frecuencias no relacionadas con movimientos musculares, a través de un

filtro paso alto (15 Hz).

Los valores del número de bloques convolucionales, el tamaño del kernel, el número de filtros iniciales y el ratio de aprendizaje se obtienen mediante una búsqueda de hiperparámetros optimizada usando la técnica Tree-structured Parzen Estimator (TPE) [31] [32]. El número máximo de filtros ha sido reducido de 1024 a 256 debido a los recursos computacionales limitados de los que se dispone. Se prueban un total de 50 modelos de los cuales, utilizando el índice kappa, se seleccionan a los cinco mejores que formarán parte del ensemble final. En la tabla 5.1 se presentan las distribuciones usadas para cada uno de los hiperparámetros y en la tabla 5.2 se muestra los valores obtenidos para los cinco mejores modelos.

Hiperparámetro	Distribución
Nº de bloques	Uniforme entre 1 y 10
Tamaño del kernel	Uniforme entre 3 y 50
Filtros iniciales	Elección entre 8, 16, 32 o 64
Ratio de aprendizaje	Log-uniforme entre -10 y -1

Tabla 5.1: Distribuciones para los distintos hiperparámetros

	Filtros iniciales	Tamaño del kernel	Ratio de aprendizaje	Bloques convolucionales
Modelo 1	8	9	$7,44 \times 10^{-5}$	8
Modelo 2	32	6	$5,65 \times 10^{-4}$	7
Modelo 3	8	8	$3,81 \times 10^{-3}$	5
Modelo 4	16	6	$8,27 \times 10^{-3}$	9
Modelo 5	16	8	$3,97 \times 10^{-3}$	8

Tabla 5.2: Hiperparámetros de los cinco mejores modelos

A partir de los cinco mejores modelos se crea un ensemble que es entrenado con los registros pertenecientes al conjunto de datos SHHS. El entrenamiento se lleva a cabo durante 50 epochs y con un tamaño de batch de 64. Se dividen los datos en un 80% para entrenamiento, un 10% para validación y un 10% para prueba. Como medida para evitar el sobreajuste se utiliza early stopping, monitorizando las pérdidas de validación con una paciencia de 10 epochs. La tabla 5.3 presenta los resultados obtenidos y la figura 5.1 muestra la matriz de confusión resultante de la evaluación del conjunto de prueba. Podemos destacar que la fase más difícil de clasificar es la fase N1 que se confunde en la mayoría de las veces con las fases W y N2. La

principal causa de esta confusión es el desbalanceo entre las clases.

Fase	Precisión	Sensitividad	F1 score
W	0,929	0,953	0,941
N1	0,411	0,184	0,254
N2	0,872	0,91	0,891
N3	0,878	0,808	0,842
REM	0,854	0,886	0,869
Media	0,789	0,748	0,759

Tabla 5.3: Resultados de la evaluación del conjuntos de prueba

		PREDICCIÓN				
		W	N1	N2	N3	REM
REAL	W	37589	511	913	21	399
	N1	1364	934	2161	3	608
	N2	874	591	42646	1199	1553
	N3	3	0	2109	8928	10
	REM	615	238	1074	14	15069

Figura 5.1: Matriz de confusión resultante de la evaluación del conjunto de prueba

5.2 Preprocesado de datos

La transferencia de aprendizaje requiere un análisis previo de los conjuntos de datos sobre los que se va a aplicar. Estos conjuntos, PhysioNet Sleep-EDF y ISRUC-Sleep, contienen una cantidad de datos mucho menor que el conjunto con el que se entrena el modelo de referencia. Al igual que el conjunto SHHS, las señales se remuestran a 125 Hz para adaptarlas al modelo de referencia.

El conjunto de datos de PhysioNet está etiquetado según el estándar R&K [2] y para adaptarlo al estándar AASM [3] se unifican las etiquetas de 'Sleep stage 3' y 'Sleep stage 4' en una única etiqueta correspondiente a la fase N3, además de eliminar aquellos registros correspondientes a epochs con etiquetas no válidas. Uno de los mayores inconvenientes de este conjunto es la gran cantidad de ejemplos de la fase W en comparación con el resto de clases.

Esto es debido al tiempo de duración de los registros, muy superior a las 8 horas de un registro típico. Para paliar este problema se eliminarán aquellos ejemplos de la fase W pertenecientes a las horas durante las que el sujeto está despierto.

Por otra parte, teniendo en cuenta el desbalanceo entre clases existente en los conjuntos mencionados (tablas 4.4 y 4.5), se realiza un sobremuestreo de las clases minoritarias mediante la técnica SMOTE (Synthetic Minority Over-sampling Technique) [33]. Se descarta el submuestreo de las clases mayoritarias ya que los conjuntos de datos son pequeños y no es recomendable reducir la cantidad de ejemplos disponibles.

En una primera aproximación se hace uso de los conjuntos de datos balanceados mediante la estrategia 'minority' de SMOTE. Esta estrategia consiste en aumentar, en cada registro, el número de ejemplos de la clase minoritaria. La distribución de clases de los conjuntos de datos tras aplicar dicha técnica se observa en las tablas 5.4 y 5.5. En general podemos observar que el número de ejemplos de las clases minoritarias aumenta en comparación a los conjuntos de datos no balanceados, consiguiendo proporciones de ejemplos más balanceadas.

Fase	W	N1	N2	N3	REM	Total
Nº Ejemplos	26.432	21.952	28.287	19.920	19.860	116.451
Proporción	27,70%	18,85%	24,29%	17,11%	17,05%	100%

Tabla 5.4: Distribución de las clases de ISRUC-Sleep balanceado con la estrategia 'minority'

Fase	W	N1	N2	N3	REM	Total
Nº Ejemplos	65.975	45.676	68.633	48.262	33.517	261.883
Proporción	25,12%	17,44%	26,21%	18,43%	12,80%	100%

Tabla 5.5: Distribución de las clases de Physionet balanceado con la estrategia 'minority'

En una segunda aproximación se hace uso de la estrategia 'not majority' de SMOTE. Esta estrategia consiste en aumentar, en cada registro, el número de ejemplos de todas las clases excepto la clase mayoritaria. Con esto se pretende igualar más la proporción de ejemplos de todas las clases. La distribución de clases de los conjuntos de datos tras aplicar dicha técnica se observa en las tablas 5.6 y 5.7. En ellas podemos observar que tanto en la fase REM para el conjunto de ISRUC, como en la fase N3 para el conjunto de Physionet, el número de ejemplos es menor que en el resto de clases. Esto se debe a que hay registros sin ejemplos de estas clases en dichos conjuntos.

Fase	W	N1	N2	N3	REM	Total
Nº Ejemplos	33.945	33.945	33.945	33.945	32.916	168.696
Proporción	20,12%	20,12%	20,12%	20,12%	19,51%	100%

Tabla 5.6: Distribución de las clases de ISRUC-Sleep balanceado con la estrategia 'not majority'

Fase	W	N1	N2	N3	REM	Total
Nº Ejemplos	82.183	82.183	82.183	66.491	82.183	395.223
Proporción	20,79%	20,79%	20,79%	16,82%	20,79%	100%

Tabla 5.7: Distribución de las clases de Physionet balanceado con la estrategia 'not majority'

5.3 Escenarios de transferencia de aprendizaje

Sobre el modelo de referencia se utilizarán distintos escenarios de transferencia basándose en modelos preentrenados, el ajuste de la capa softmax y el ajuste del último bloque convolucional [12] [34] [35].

5.3.1 Escenario 1: Modelos preentrenados

En este escenario, también conocido como transferencia directa, se utiliza el modelo preentrenado como un clasificador sobre el nuevo conjunto de datos. No se realiza ningún tipo de ajuste de pesos de la red.

5.3.2 Escenario 2: Ajuste de la capa softmax

En este escenario se usa el modelo de referencia como un extractor de características fijo, es decir, los pesos de todas las capas permanecen fijos, a excepción de la capa densa con activación softmax. Este escenario es usado habitualmente cuando el nuevo conjunto de datos es similar al conjunto de datos con el que se entrenó el modelo original. En nuestro caso, los conjuntos PhysioNet Sleep-EDF e ISRUC-Sleep cumplen con este requisito, a pesar de existir diferencias en la frecuencia de muestreo de las señales.

5.3.3 Escenario 3: Ajuste del último bloque convolucional

En este escenario, además de ajustar la capa softmax, se ajusta el último bloque convolucional, es decir, aquel que se encuentra en lo alto de la jerarquía de la red. El resto de bloques

convolucionales permanecen fijos. Las primeras capas de la red convolucional contienen características más genéricas y, a medida que se avanza en la jerarquía, las capas contienen características más específicas del conjunto de datos original. De esta forma, ajustando solamente el último bloque convolucional, estamos asumiendo que las características genéricas de ambos conjuntos son similares y solo nos interesa que el modelo aprenda las características más específicas del conjunto de datos.

5.3.4 Escenario 4: Ajuste de toda la red

En este escenario se ajustan todos los bloques convolucionales. Lo que se pretende al ajustar el modelo entero es conseguir que este aprenda tanto las características más genéricas como las más específicas del nuevo conjunto de datos.

En la figura 5.2 se observa un esquema de un modelo de ejemplo en el que se indica gráficamente los bloques que se ajustarían en cada escenario. Para el escenario 1 no se ajustaría ningún bloque por lo que no aparece representado en dicha figura.

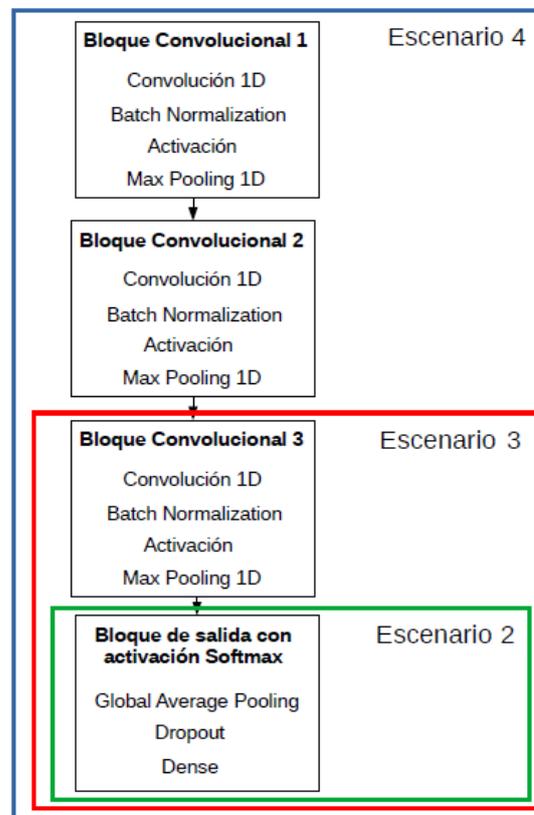


Figura 5.2: Ajustes de los distintos escenarios para un modelo de ejemplo

Resultados y evaluación

EN este capítulo describiremos los resultados obtenidos en cada uno de los escenarios de transferencia de aprendizaje, compararemos los resultados de cada uno de ellos sobre los distintos conjuntos de datos y presentaremos de forma gráfica algunos de los casos más destacados. Para finalizar haremos una comparación con los resultados de otros trabajos.

6.1 Resultados de la transferencia de aprendizaje

Este apartado se divide en cuatro subapartados, uno por cada conjunto de datos y estrategia de balanceo, que serán denominados: Physionet-minority, Physionet-notmajority, ISRUC-minority e ISRUC-notmajority. En cada subapartado se describen los resultados obtenidos en cada escenario de transferencia de aprendizaje. Haremos referencia a cada escenario según su orden de descripción en el capítulo anterior, es decir, escenario 1 (escenario de modelo preentrenado), escenario 2 (ajuste de la capa softmax), escenario 3 (ajuste del último bloque convolucional) y escenario 4 (ajuste de toda la red).

6.1.1 Physionet-minority

Los resultados, en términos de exactitud y macro F1 obtenidos al aplicar transferencia de aprendizaje sobre el conjunto de datos de Physionet balanceado con la estrategia 'minority', bajo los cuatro escenarios, se pueden observar en la tabla 6.1. Se observa que el modelo original (escenario 1) tiene dificultades para realizar la clasificación de las fases de sueño. De hecho, esta clasificación mejora cuanto más se ajusta el modelo preentrenado, consiguiendo valores cercanos al 77% para la medida macro F1 en el escenario 4.

En la tabla 6.2 se presentan los resultados en términos de precisión, sensibilidad y F1 score del escenario que presenta los mejores resultados (escenario 4). En este caso los mejores resultados son los obtenidos en la clasificación de la fase W con un F1 score de 89,3%. La fase N1 obtiene un F1 score de 59,7%, siendo este un resultado bastante bueno teniendo en cuenta

que es la fase más difícil de clasificar y que presenta un menor número de ejemplos respecto al resto.

Escenario	Exactitud	Macro F1
Escenario 1	0,523	0,489
Escenario 2	0,575	0,559
Escenario 3	0,625	0,605
Escenario 4	0,772	0,769

Tabla 6.1: Medidas de rendimiento para el conjunto Physionet-minority

Fase	Precisión	Sensitividad	F1 score
W	0,869	0,919	0,893
N1	0,747	0,497	0,597
N2	0,667	0,859	0,751
N3	0,801	0,738	0,799
REM	0,808	0,804	0,806

Tabla 6.2: Medidas de rendimiento del escenario 4 para el conjunto Physionet-minority

6.1.2 ISRUC-minority

Los resultados obtenidos al aplicar transferencia de aprendizaje sobre el conjunto de datos de ISRUC-Sleep balanceado con la estrategia 'minority', bajo los cuatro escenarios, se pueden observar en la tabla 6.3. Podemos observar que para este conjunto de datos los resultados mejoran ligeramente con respecto a los del caso anterior (tabla 6.1), en los tres primeros escenarios. En cuanto al escenario 4, aún siendo el que mejores resultados proporciona, consigue valores inferiores a los del apartado anterior. Al igual que pasaba en este caso anterior, el modelo tiene dificultades para realizar la clasificación que mejora cuanto más se ajusta el modelo preentrenado.

En la tabla 6.4 se presentan los resultados del mejor escenario, el escenario 4, en términos de precisión, sensitividad y F1 score. En este caso los mejores resultados son los obtenidos en la clasificación de la fase N3 sin conseguir aumentar el rendimiento del conjunto Physionet-minority. El resto de las fases no consiguen valores de rendimiento destacables, siendo la fase N1 la peor clasificada.

Escenario	Exactitud	Macro F1
Escenario 1	0,538	0,514
Escenario 2	0,652	0,636
Escenario 3	0,663	0,655
Escenario 4	0,690	0,691

Tabla 6.3: Medidas de rendimiento para el conjunto ISRUC-minority

Fase	Precisión	Sensitividad	F1 score
W	0,655	0,864	0,745
N1	0,571	0,519	0,544
N2	0,715	0,725	0,720
N3	0,945	0,679	0,79
REM	0,676	0,638	0,656

Tabla 6.4: Medidas de rendimiento del escenario 4 para el conjunto ISRUC-minority

6.1.3 Physionet-notmajority

Los resultados obtenidos al aplicar transferencia de aprendizaje sobre el conjunto de datos de Physionet balanceado con la estrategia 'not minority', bajo los cuatro escenarios, se pueden observar en la tabla 6.5. En general, los valores obtenidos no mejoran con respecto al caso Physionet-minority. Se observa igualmente una tendencia de mejora cuanto más ajuste se realiza sobre el modelo preentrenado. Podríamos destacar los resultados obtenidos por el escenario 3 ya que son superiores a los dos casos previos (tablas 6.1 y 6.3) en términos de ambas medidas de rendimiento.

En la tabla 6.6 se presentan los resultados del mejor escenario, el escenario 4, en términos de precisión, sensitividad y F1 score. En este caso y aún a pesar de ser la fase peor clasificada en este escenario, la fase N1 consigue valores de rendimiento superiores a los casos antes estudiados. Ocurre lo mismo con la fase N3, que consigue un aumento en las medidas de rendimiento del 3%.

Escenario	Exactitud	Macro F1
Escenario 1	0,429	0,420
Escenario 2	0,491	0,470
Escenario 3	0,704	0,691
Escenario 4	0,736	0,743

Tabla 6.5: Medidas de rendimiento para el conjunto Physionet-notmajority

Fase	Precisión	Sensitividad	F1 score
W	0,876	0,889	0,882
N1	0,569	0,662	0,611
N2	0,635	0,696	0,664
N3	0,889	0,767	0,824
REM	0,806	0,674	0,734

Tabla 6.6: Medidas de rendimiento del escenario 4 para el conjunto Physionet-notmajority

6.1.4 ISRUC-notmajority

Los resultados obtenidos al aplicar transferencia de aprendizaje sobre el conjunto de datos de ISRUC-Sleep balanceado con la estrategia 'not majority', bajo los cuatro escenarios, se pueden observar en la tabla 6.7. En este caso se observa la misma tendencia que en los apartados anteriores: los resultados mejoran cuanto más se ajusta el modelo preentrenado. En este caso, la estrategia de balanceo permite una mejora importante en el rendimiento del último escenario, consiguiendo mejores resultados que los conseguidos sobre el conjunto Physionet con esta misma estrategia de balanceo.

En la tabla 6.8 se presentan los resultados del mejor escenario, el escenario 4, en términos de precisión, sensibilidad y F1 score. En ella se puede observar que la estrategia de balanceo not majority resulta más apropiada que la estrategia minority, consiguiendo mejoras en las medidas de rendimiento de casi todas las fases. Destacan los valores obtenidos para la fase N2, que sale perjudicada tanto en este conjunto de datos como en Physionet, por la estrategia de balanceo.

Escenario	Exactitud	Macro F1
Escenario 1	0,527	0,511
Escenario 2	0,667	0,658
Escenario 3	0,647	0,637
Escenario 4	0,761	0,761

Tabla 6.7: Medidas de rendimiento para el ISRUC-notmajority

Fase	Precisión	Sensitividad	F1 score
W	0,803	0,877	0,838
N1	0,569	0,681	0,619
N2	0,779	0,596	0,675
N3	0,863	0,867	0,865
REM	0,832	0,783	0,807

Tabla 6.8: Medidas de rendimiento del escenario 4 para el conjunto ISRUC-notmajority

6.1.5 Conclusiones de la transferencia de aprendizaje

El análisis de los resultados obtenidos nos lleva a concluir que la transferencia de aprendizaje no ha resultado según lo esperado. Se puede comprobar como los resultados mejoran cuanto más se ajusta el modelo preentrenado, lo que es síntoma de que dicho modelo necesita de un ajuste a los datos que intenta clasificar.

El comportamiento de ambos conjuntos con respecto a las técnicas de balanceo no es comparable ya que se podría decir que los resultados están cruzados. Los mejores resultados de Physionet se consiguen con la técnica minority mientras que para el conjunto ISRUC ocurre lo contrario. Sin embargo si se puede comprobar que a menor número de ejemplos, peores resultados. Esto es lo que ocurre en el caso del conjunto ISRUC con la estrategia de balanceo minority (tabla 6.4).

Finalmente, cabe destacar la ligera mejora de F1 score para la fase N1 al pasar de los conjuntos balanceados con la estrategia 'minority' a los balanceados con la estrategia 'not majority'. Como se ha dicho desde un principio, a pesar de que los resultados de clasificación de la fase N1 son los más bajos, son unos resultados más que aceptables.

6.2 Visualización de los resultados

En este apartado representaremos de forma gráfica las muestras dentro de un ejemplo en las que el modelo pone más atención a la hora de clasificar dicho ejemplo. Para esta tarea usamos la herramienta `tf-keras-vis` [27] que nos proporciona los mapas de prominencia (saliency maps) correspondientes a cada entrada. El objetivo de los mapas de prominencia es facilitar el análisis del funcionamiento de una red neuronal a través de una imagen [36] [37].

Presentaremos estas gráficas para cada uno de los conjuntos de datos y para cada una de las fases de sueño. Dentro de cada conjunto de datos, las gráficas se corresponden al escenario y método de balanceo que han conseguido el mejor resultado. Estos ejemplos están compuestos por una figura que representa las tres señales de entrada a la red y otra figura que representa el mapa de prominencia de dicha entrada. En esta última figura el eje y representa la importancia que tiene cada muestra para la clasificación de la señal, en un rango de 0 a 1.

6.2.1 Mapas de prominencia para el conjunto Physionet

En las figuras 6.1, 6.2, 6.3, 6.4 y 6.5 se observa un ejemplo de cada fase de sueño para Physionet-minority en el escenario 4. De forma general se puede observar que los valores más altos en los mapas de prominencia corresponden a zonas de la señal donde los rangos de amplitud son menores o a zonas que limitan con estos picos de amplitud. Estos valores también nos permiten identificar en qué momentos del epoch se identifican características propias de la fase de sueño.

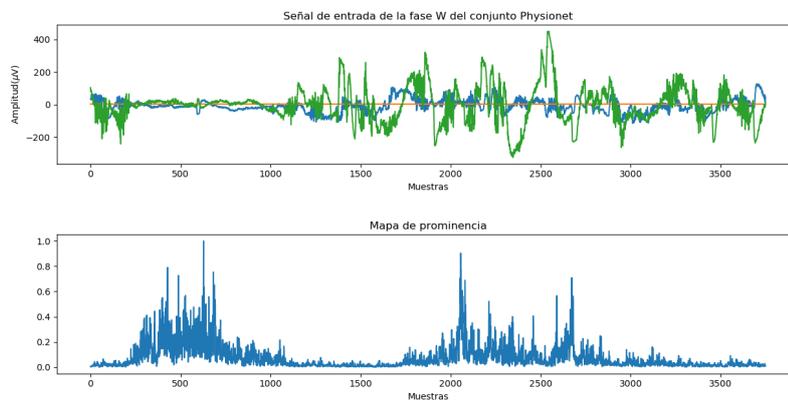


Figura 6.1: Ejemplo de mapa de prominencia para la fase W

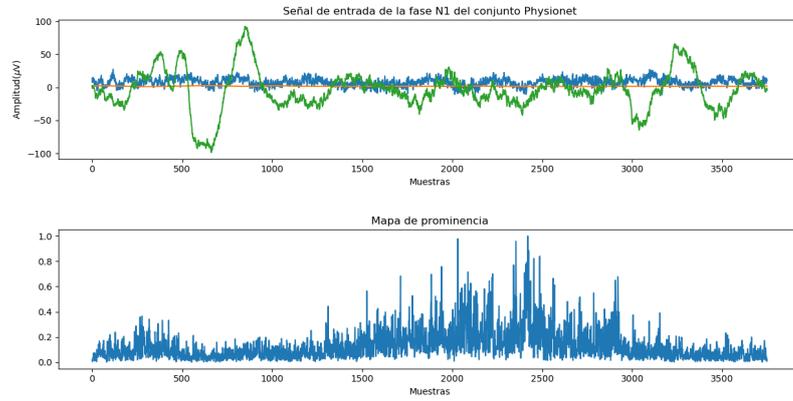


Figura 6.2: Ejemplo de mapa de prominencia para la fase N1

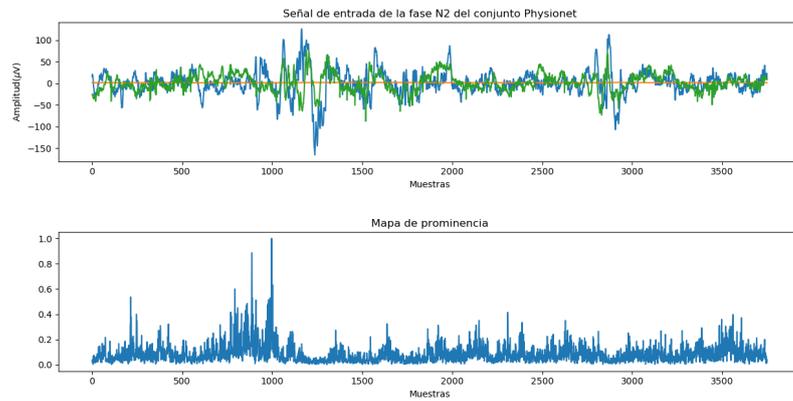


Figura 6.3: Ejemplo de mapa de prominencia para la fase N2

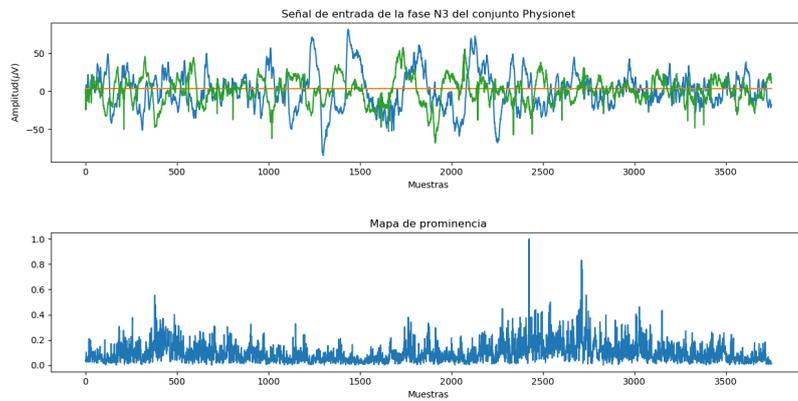


Figura 6.4: Ejemplo de mapa de prominencia para la fase N3

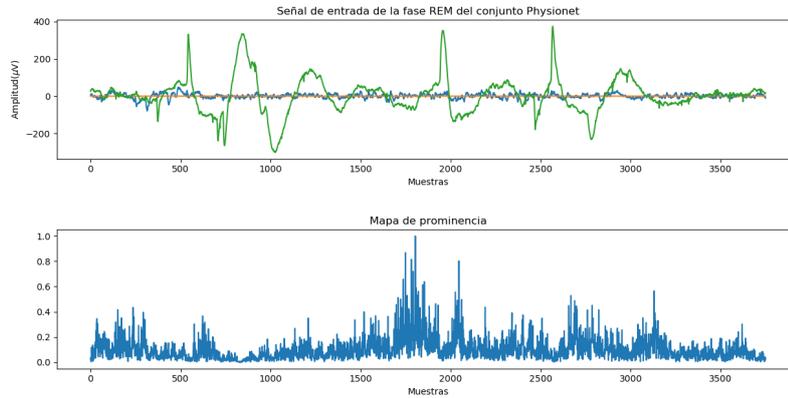


Figura 6.5: Ejemplo de mapa de prominencia para la fase REM

6.2.2 Mapas de prominencia para el conjunto ISRUC

En las figuras 6.6, 6.7, 6.8, 6.9 y 6.10 se observa un ejemplo de cada fase de sueño para ISRUC-notmajority en el escenario 4. Al igual que para el caso anterior las muestras del ejemplo a las que la red le da una mayor importancia son aquellas donde el rango de amplitud de la señal es más bajo. En este caso sin embargo, solamente en las fases W y N1 los mapas muestran claramente en que momentos del epoch se centra el modelo para realizar dicha clasificación.

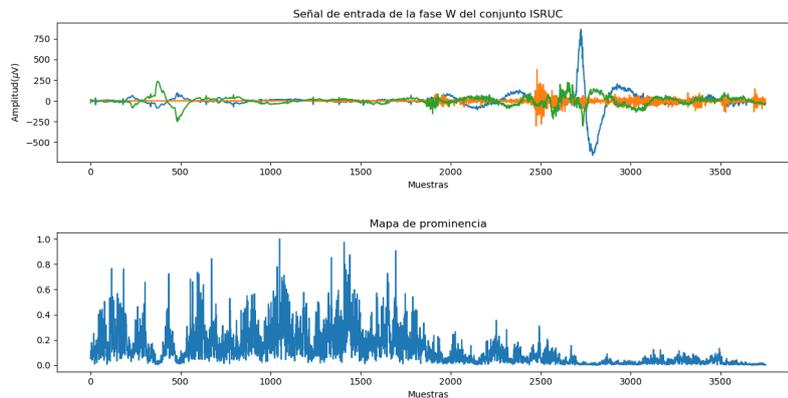


Figura 6.6: Ejemplo de mapa de prominencia para la fase W

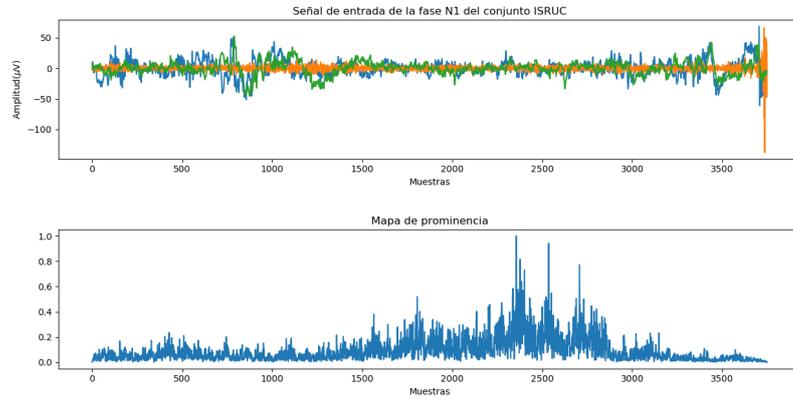


Figura 6.7: Ejemplo de mapa de prominencia para la fase N1

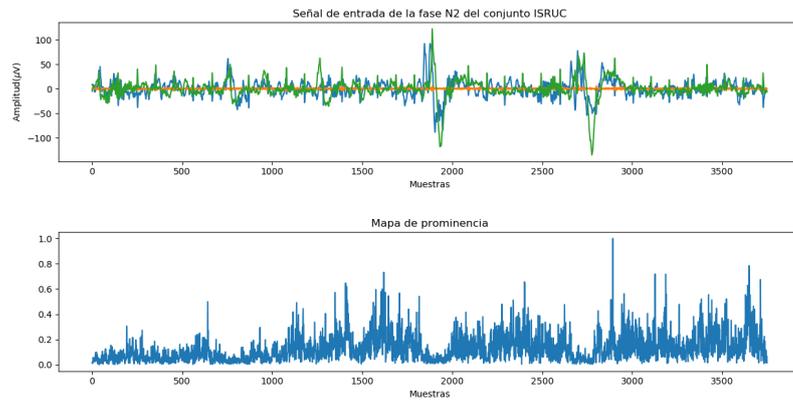


Figura 6.8: Ejemplo de mapa de prominencia para la fase N2

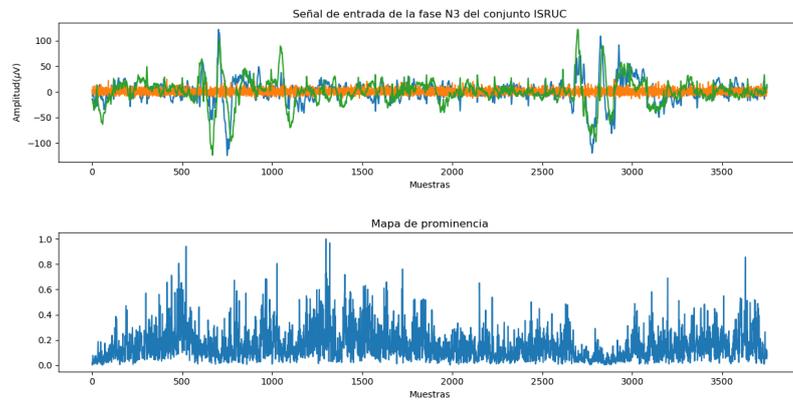


Figura 6.9: Ejemplo de mapa de prominencia para la fase N3

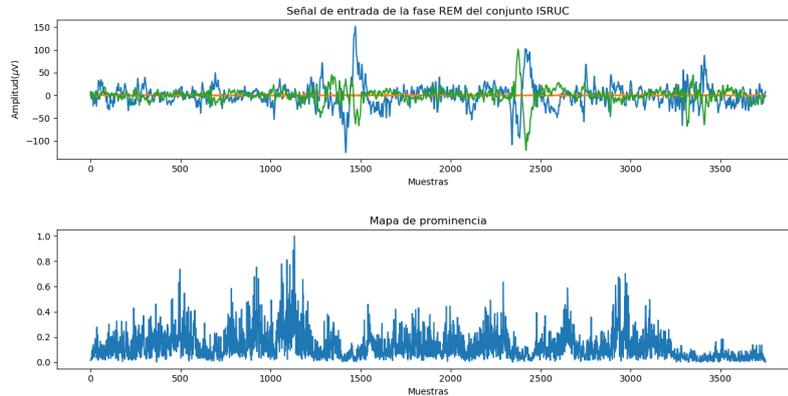


Figura 6.10: Ejemplo de mapa de prominencia para la fase REM

6.3 Comparación con otros trabajos

En este apartado compararemos los resultados obtenidos en este trabajo con los resultados de otros trabajos similares descritos en el capítulo 3. Para realizar esta comparación se usará la métrica F1 score para cada fase de sueño, ya que proporciona una idea global de los resultados sobre el mejor escenario y método de balanceo, para cada conjunto de datos. Las tablas 6.9 y 6.10 presentan los datos sobre los que se realiza dicha comparación. A pesar de que la comparación no es justa ya que los conjuntos de datos no contienen los mismos registros y las señales utilizadas no siempre coinciden, los resultados obtenidos pueden competir con los de otros trabajos.

Para el conjunto de datos Physionet observamos que los resultados para las fases N2 y N3 son algo más bajos que para los trabajos de O.Tsinalis et al. [8] y A.Supratak et al. [9]. En cambio las fases W y N1 presentan un rendimiento bastante superior a la vez que aceptable. En cuanto a la fase REM se alcanza un valor de F1 muy similar al del resto de trabajos, tan solo un 7% por debajo del trabajo de A.Supratak et al. [9].

Para el conjunto de datos ISRUC observamos unos resultados para las fases N1, N3 y REM muy similares al los del trabajo de A.A.Gharbali et al. [13]. Sin embargo, el rendimiento para el resto de las fases se degrada ligeramente para el caso W y en mayor medida para la fase N2.

Trabajo	W	N1	N2	N3	REM
Escenario 4 + Physionet-minority	0,88	0,65	0,74	0,78	0,75
O. Tsinalis et al. [8]	0,65	0,44	0,81	0,85	0,75
A. Supratak et al. (EEG) [9]	0,85	0,45	0,86	0,85	0,82

Tabla 6.9: Comparación para el conjunto de datos Physionet

Trabajo	W	N1	N2	N3	REM
Escenario 4 + ISRUC-notmajority	0,84	0,62	0,68	0,87	0,81
A. A. Gharbali et al. [13]	0,94	0,66	0,82	0,87	0,80

Tabla 6.10: Comparación para el conjunto de datos ISRUC

Conclusiones y trabajo futuro

EN este capítulo se comentan las conclusiones generales del trabajo así como posibles trabajos futuros.

7.1 Conclusiones

Como hemos visto, a lo largo de este trabajo se han cumplido los objetivos planteados, hemos desarrollado un modelo sobre el que aplicar transferencia de aprendizaje y hemos analizado los resultados obtenidos por el mismo. Los resultados generales han sido satisfactorios y similares a otros trabajos relacionados. A lo largo del desarrollo del trabajo han surgido complicaciones que han frenado el avance del mismo.

En cuanto a los conjuntos de datos la principal dificultad encontrada ha sido los diferentes formatos tanto en las señales como en las anotaciones, esto implica el uso de diferentes herramientas de lectura para cada conjunto de datos.

La escasez de ejemplos de uso para redes convolucionales que trabajen con señales ha sido una dificultad a tener en cuenta, sobre todo a la hora de usar la herramienta de visualización de mapas de prominencia, principalmente desarrollada para trabajar con imágenes.

Una de las principales dificultades a la hora de realizar las tareas de implementación ha sido la incompatibilidad de ciertas versiones de las herramientas utilizadas.

Uno de los conocimientos a mi parecer más importantes, adquiridos durante el desarrollo de este trabajo, es la importancia del estudio de los conjuntos de datos antes de realizar la transferencia de aprendizaje. Esto nos permite detectar desbalanceo entre clases, que condiciona los resultados de cualquier tarea de clasificación.

A lo largo del estudio del grado apenas se hace referencia al Aprendizaje Profundo, tan solo se dan unos conceptos básicos acerca de este tema. Igual de ausente se encuentra en el ámbito práctico. Esto conlleva que para la realización de trabajos de Aprendizaje Profundo, como este, se parta de una base demasiado pobre.

En general el conocimiento adquirido a lo largo del desarrollo de este trabajo ha sido muy amplio, desde conceptos más específicos como conocimiento teórico sobre la medicina del sueño, desarrollo de modelos de Aprendizaje Profundo o aplicación de transferencia de aprendizaje sobre nuevos conjuntos de datos, hasta conocimientos más básicos como la estructuración y desarrollo de una memoria o la planificación de un trabajo de este estilo.

7.2 Trabajo futuro

Como trabajo futuro y con la finalidad de mejorar los resultados obtenidos en este, se propone lo siguiente:

- Usar un mayor número de señales como entradas a los modelos, como el ECG o señales de movimiento de extremidades.
- Probar nuevos escenarios de transferencia de aprendizaje que nos permitan mejorar los resultados sin la necesidad de ajustar toda la red.
- Probar distintas estrategias para balancear los datos que nos permitan minimizar las clasificaciones incorrectas.

Apéndices

Lista de acrónimos

PSG *Polisomnografía*

EEG *Electroencefalograma*

EMG *Electromiograma*

EOG *Electrooculograma*

R&K *Rechtschaffen y Kales*

AASM *Academia Americana de Medicina del Sueño*

REM *Rapid Eye Movement*

MOR *Movimiento ocular rápido*

SVM *Support Vector Machine*

LS *Light Sleep*

SWS *Slow Wave Sleep*

ANN *Artificial Neural Network*

CNN *Convolutional Neural Network*

MASS *Montreal Archive of Sleep Studies*

LSTM *Long Short-Term Memory*

ReLU *Rectifier Linear Unit*

RNN *Recurrent Neural Network*

EDF *European Data Format*

CWT *Continuous Wavelet Transform*

IDE *Integrated Development Environment*

SHHS *Sleep Heart Health Study*

TPE *Tree-structured Parzen Estimator*

SMOTE *Synthetic Minority Over-sampling Technique*

ECG *Electrocardiogram*

Glosario

Insomnio Trastorno que causa falta de sueño y dificultad para conciliarlo.

Síndrome de la apnea del sueño Trastorno que causa interrupciones en la respiración durante el sueño.

Hipersomnias Trastorno del sueño que causa episodios de somnolencia diurna excesiva.

Capa pooling En español capa de agrupamiento. Capa usada para reducir la dimensión espacial en una red neuronal convolucional.

Capa fully-connected En español capa totalmente conectada. Capa que conecta todas las neuronas de una capa con todas las neuronas de otra capa.

Softmax Función que proporciona probabilidades para cada posible clase de salida.

ReLU Función de activación que permite al modelo resolver problemas no lineales.

Maxpooling Tipo de capa de agrupamiento que reduce la dimensión espacial siguiendo la política de seleccionar al candidato máximo.

Bibliografía

- [1] J. G. Pérez-Larraya, J. Toledo, E. Urrestarazu, and J. Iriarte, *Clasificación de los trastornos del sueño*, 2007. [En línea]. Disponible en: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272007000200003
- [2] A. Kales and A. Rechtschaffen, *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*, 1968. [En línea]. Disponible en: <https://archive.org/details/RKManual/page/n5>
- [3] C. Iber, S. Ancoli-Israel, A. Chesson, and S. F. Quan, *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, 2007. [En línea]. Disponible en: https://www.researchgate.net/publication/285881487_The_AASM_Manual_for_the_Scoring_of_Sleep_and_Associated_Events_Rules_Terminology_and_Technical_Specifications
- [4] S. Gudmundsson, T. P. Runarsson, and S. Sigurdsson, *Automatic Sleep Staging using Support Vector Machines with Posterior Probability Estimates*, 2005. [En línea]. Disponible en: https://www.researchgate.net/publication/4242568_Automatic_Sleep_Staging_using_Support_Vector_Machines_with_Posterior_Probability_Estimates
- [5] F. Ebrahimi, M. Mikaeili, E. Estrada, and H. Nazeran, *Automatic Sleep Stage Classification Based on EEG Signals by Using Neural Networks and Wavelet Packet Coefficients*, 2008. [En línea]. Disponible en: https://www.researchgate.net/publication/23932036_Automatic_Sleep_Stage_Classification_Based_on_EEG_Signals_by_Using_Neural_Networks_and_Wavelet_Packet_Coefficients
- [6] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. C. Kamphuisen, and J. J. L. Oberyé, *Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG*, 2000. [En línea]. Disponible en: https://www.researchgate.net/publication/3035248_Analysis_of_a_sleep-dependent_neuronal_feedback_loop_The_slow-wave_microcontinuity_of_the_EEG

-
- [7] A. Goldberger, L. Glass, P. C. Ivanov, J. E. Mietus, C.-K. Peng, L. A. N. Amaral, J. M. Hausdorff, R. G. Mark, G. B. Moody, and H. E. Stanley, *PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals*, 2000. [En línea]. Disponible en: https://www.researchgate.net/publication/12467774_PhysioBank_PhysioToolkit_and_PhysioNet_Components_of_a_New_Research_Resource_for_Complex_Physiologic_Signals
- [8] O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, *Automatic Sleep Stage Scoring with Single-Channel EEG Using Convolutional Neural Networks*, 2016. [En línea]. Disponible en: <https://arxiv.org/abs/1610.01683>
- [9] A. Supratak, H. Dong, C. Wu, and Y. Guo, *DeepSleepNet: a Model for Automatic Sleep Stage Scoring based on Raw Single-Channel EEG*, 2017. [En línea]. Disponible en: <https://arxiv.org/abs/1703.04046>
- [10] C. O'Reilly, N. Gosselin, J. Carrier, and T. Nielsen, *Montreal Archive of Sleep Studies: an open-access resource for instrument benchmarking and exploratory research*, 2014. [En línea]. Disponible en: <https://onlinelibrary.wiley.com/doi/full/10.1111/jsr.12169>
- [11] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. D. Vos, *SeqSleepNet: End-to-End Hierarchical Recurrent Neural Network for Sequence-to-Sequence Automatic Sleep Staging*, 2018. [En línea]. Disponible en: <https://arxiv.org/abs/1809.10932>
- [12] H. Phan, O. Y. Chén, P. Koch, A. Mertins, and M. D. Vos, *Deep Transfer Learning for Single-Channel Automatic Sleep Staging with Channel Mismatch*, 2019. [En línea]. Disponible en: <https://arxiv.org/abs/1904.05945>
- [13] A. A. Gharbali, S. Najdi, and J. M. Fonseca, *Transfer Learning of Spectrogram Image for Automatic Sleep Stage Classification*, 2018. [En línea]. Disponible en: https://www.researchgate.net/publication/325573102_Transfer_Learning_of_Spectrogram_Image_for_Automatic_Sleep_Stage_Classification
- [14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, 2012. [En línea]. Disponible en: https://www.researchgate.net/publication/267960550_ImageNet_Classification_with_Deep_Convolutional_Neural_Networks
- [15] K. Sirvan, T. Sousa, J. M. Santos, and U. Nunes, *ISRUC-Sleep: A comprehensive public dataset for sleep researchers*, 2015. [En línea]. Disponible en: https://www.researchgate.net/publication/283734463_ISRUC-Sleep_A_comprehensive_public_dataset_for_sleep_researchers

- [16] “What is spiral model- advantages, disadvantages and when to use it?” Último acceso 20-06-2020. [En línea]. Disponible en: <http://tryqa.com/what-is-spiral-model-advantages-disadvantages-and-when-to-use-it/>
- [17] “Modelo en espiral,” Último acceso 20-06-2020. [En línea]. Disponible en: https://es.ryte.com/wiki/Modelo_en_Espiral
- [18] S. Raschka, *Python Machine Learning*, 1st ed. Packt Publishing, 2015.
- [19] I. Vasilev, D. Slater, G. Spacagna, P. Roelants, and V. Zocca, *Python Deep Learning: Exploring deep learning techniques and neural network architectures with PyTorch, Keras, and TensorFlow*, 2nd ed. Packt Publishing, 2019.
- [20] Chollet, François *et al.*, “Keras,” 2015, Último acceso 20-06-2020. [En línea]. Disponible en: <https://keras.io>
- [21] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, Último acceso 20-06-2020. [En línea]. Disponible en: <https://www.tensorflow.org/>
- [22] A. Gramfort, M. Luessi, E. Larson, D. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. Hämäläinen, *MNE software for processing MEG and EEG data*, 2014. [En línea]. Disponible en: <https://www.ncbi.nlm.nih.gov/pubmed/24161808>
- [23] W. McKinney, *pandas: a Foundational Python Library for DataAnalysis and Statistics*, 2011. [En línea]. Disponible en: <https://es.scribd.com/document/71048089/pandas-a-Foundational-Python-Library-for-Data-Analysis-and-Statistics>
- [24] Combrisson, Etienne, Vallat, Raphael, O’Reilly, Christian, Jas, Mainak, Pascarella, Annalisa, Saive, Anne-lise, Thiery, Thomas, Meunier, David, Altukhov, Dmitrii, Lajnef, Tarek, Ruby, Perrine, Guillot, Aymeric, Jerbi, and Karim, *Visbrain: A Multi-Purpose GPU-Accelerated Open-Source Suite for Multimodal Brain Data Visualization*, 2019. [En línea]. Disponible en: <https://www.frontiersin.org/article/10.3389/fninf.2019.00014>

-
- [25] G. Lemaitre, F. Nogueira, and C. K. Aridas, *Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning*, 2017. [En línea]. Disponible en: <http://jmlr.org/papers/v18/16-365.html>
- [26] J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, and D. D. Cox, *Hyperopt: a Python library for model selection and hyperparameter optimization*, 2015. [En línea]. Disponible en: <https://iopscience.iop.org/article/10.1088/1749-4699/8/1/014008/meta>
- [27] Y. Kubota, “tf-keras-vis,” 2019, Último acceso 20-06-2020. [En línea]. Disponible en: <https://github.com/keisen/tf-keras-vis>
- [28] J. D. Hunter, “Matplotlib: A 2d graphics environment,” 2007, Último acceso 20-06-2020. [En línea]. Disponible en: <https://matplotlib.org/>
- [29] S. Redline, M.H.Sanders, B. Lind, S. Quan, C. Iber, D. Gottlieb, W. Bonekat, D. Rapoport, P. Smith, and J. Kiley, *Methods for obtaining and analyzing unattended polysomnography data for a multicenter study*. Sleep Heart Health Research Group, 1998. [En línea]. Disponible en: <https://www.ncbi.nlm.nih.gov/pubmed/11300121>
- [30] I. Fernández-Varela, E. Hernández-Pereira, D. Alvarez-Estevez, and V. Moret-Bonillo, *Una red convolucional para la clasificación de las fases de sueño*, 2018.
- [31] J. Bergstra, R. Bardenet, B. Kégl, and Y. Bengio, *Algorithms for Hyper-Parameter Optimization*, 2011. [En línea]. Disponible en: https://www.researchgate.net/publication/216816964_Algorithms_for_Hyper-Parameter_Optimization
- [32] J. Bergstra, D. Yamins, and D. Cox, *Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms*, 2013. [En línea]. Disponible en: https://www.researchgate.net/publication/339440322_Hyperopt_A_Python_Library_for_Optimizing_the_Hyperparameters_of_Machine_Learning_Algorithms
- [33] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, *SMOTE: Synthetic Minority Over-sampling Technique*, 2002. [En línea]. Disponible en: <https://arxiv.org/abs/1106.1813>
- [34] “Cs231n: Convolutional neural networks for visual recognition,” 2015, Último acceso 20-06-2020. [En línea]. Disponible en: <https://cs231n.github.io/transfer-learning/>
- [35] S. Kostadinov, “What is deep transfer learning and why is it becoming so popular?” 2019, Último acceso 20-06-2020. [En línea]. Disponible en: <https://towardsdatascience.com/what-is-deep-transfer-learning-and-why-is-it-becoming-so-popular-91acdcc2717a>

BIBLIOGRAFÍA

- [36] A. Sharma, “What are saliency maps in deep learning?” 2018, Último acceso 20-06-2020. [En línea]. Disponible en: <https://analyticsindiamag.com/what-are-saliency-maps-in-deep-learning/>
- [37] A. Schreiber, “Saliency maps for deep learning part 1: Vanilla gradient,” 2019, Último acceso 20-06-2020. [En línea]. Disponible en: <https://medium.com/@thelastalias/saliency-maps-for-deep-learning-part-1-vanilla-gradient-1d0665de3284>

