



UNIVERSIDAD DE VALLADOLID



Escuela Técnica Superior de Ingenieros de Telecomunicación
Universidad de Valladolid

TRABAJO FIN DE MÁSTER

MÁSTER EN INGENIERÍA DE TELECOMUNICACIÓN

CONTRIBUCIONES A LA DETECCIÓN AUTOMÁTICA DE
TDAH EN LA INFANCIA MEDIANTE SOLUCIONES BASADAS
EN DEEP LEARNING

AUTOR: PATRICIA AMADO CABALLERO
TUTOR: CARLOS ALBEROLA LÓPEZ
J.P. CASASECA DE LA HIGUERA

22 de septiembre de 2021

TÍTULO: CONTRIBUCIONES A LA DETECCIÓN AUTOMÁTICA
DE TDAH EN LA INFANCIA MEDIANTE SOLUCIONES
BASADAS EN DEEP LEARNING
AUTOR: PATRICIA AMADO CABALLERO
TUTOR: CARLOS ALBEROLA LÓPEZ
J.P. CASASECA DE LA HIGUERA
DEPARTAMENTO: TSCIT

Miembros del Tribunal

PRESIDENTE: MÁRCOS MARTÍN FERNÁNDEZ
SECRETARIO: IGNACIO DE MIGUEL JIMÉNEZ
VOCAL: LUIS MIGUEL SAN JOSÉ REVUELTA
P. SUPLENTE: MANUEL RODRÍGUEZ CAYETANO
S. SUPLENTE: JAVIER AGUIAR PÉREZ
VOCAL SUPLENTE: RAMÓN DE LA ROSA STEINZ
CALIFICACIÓN:

Resumen del proyecto

El Trastorno por Déficit de Atención e Hiperactividad es considerado uno de los trastornos psiquiátricos infantiles más frecuentes, con una prevalencia en torno al 5% y con un número significativamente mayor de varones diagnosticados con respecto al de mujeres, debido a la dificultad de encontrar marcadores claros que ayuden a determinar la presencia de esta patología en las mujeres. Este hecho, unido a la atenuación de los síntomas con la edad, complica el diagnóstico con el crecimiento. En el presente trabajo, se plantea aunar la utilización de registros de actigrafía con técnicas de aprendizaje profundo para la obtención de patrones que nos ayuden a encontrar diferencias en función de la edad y el sexo de los pacientes diagnosticados con TDAH. Para realizar este análisis, se han creado grupos de pacientes divididos por edad y sexo y se han clasificado mediante una red neuronal convolucional adaptada a cada grupo. Una vez realizada esta clasificación, se han empleado mapas de oclusión como técnica de visualización para ver las características de las señales útiles para la clasificación e interpretar el resultado obtenido en relación con la patología.

Palabras clave

TDAH, Actigrafía, CNN, Mapas de oclusión, Aprendizaje Profundo, Clasificación

Keywords

ADHD, Actigraphy, CNN, Occlusion Sensitivity, Deep Learning, Classification

Abstract

Attention Deficit Hyperactivity Disorder is considered one of the most frequent disorder in

childhood, with a prevalence around 5% and mainly diagnosed in males against females, due to the difficulty of finding markers due to the difficulty of finding clear markers that help to determine the presence of this pathology in females. This fact together with the attenuation of symptoms with age, complicates diagnosis with growth. In this work we propose to combine actigraphy records with Deep learning techniques to obtain patterns that help us to find differences according to the age and sex of patients diagnosed with ADHD. To perform this analysis, groups of patients divided by sex and age have been created and classified using a convolutional neural network adapted to each group. Once this classification was performed, occlusion sensitivity is used as a visualization technique to see the characteristics of the signals useful for the classification and to interpret the result obtained in relation to the pathology.

AGRADECIMIENTOS

En primer lugar, me gustaría agradecerle a mis tutores, Carlos Alberola y Pablo Casaseca por su confianza y su ayuda para que este trabajo saliera adelante.

También me gustaría agradecer al Laboratorio de Procesado de Imagen el haberme proporcionado los medios necesarios para el desarrollo de este trabajo

Me gustaría agradecer a mi familia todo el apoyo que me han dado durante este tiempo, ya que sin ellos, sin su paciencia y su cariño, el camino habría sido mucho más difícil. Gracias por ser mi apoyo incondicional.

ÍNDICE GENERAL

1. Introducción	1
1.1. Motivaciones	1
1.2. Objetivos	2
1.3. Fases y Métodos	2
1.4. Medios necesarios empleados para el desarrollo	3
1.5. Estructura del documento	4
2. Antecedentes	5
2.1. Estado del Arte	5
2.1.1. Resumen de los métodos de diagnóstico objetivo del TDAH	5
2.1.2. Diferencias por sexo y edad en el diagnóstico del TDAH según los métodos tradicionales	10
2.1.3. Utilización de señales biomédicas para el estudio del TDAH	10
2.1.4. El uso de la MRI en el diagnóstico del TDAH	10
2.1.5. Estado del Arte: La interpretación del Deep Learning	11
2.1.5.1. Interpretación de redes neuronales en el diagnóstico de enfermedades	11
2.1.5.2. Interpretabilidad del Deep Learning para el diagnóstico del TDAH	12
2.1.6. Análisis y comparativa	12
2.2. Técnicas y métricas para evaluar el rendimiento de un sistema de reconocimiento de patrones	12
2.2.1. Técnicas para la evaluación del rendimiento de un sistema experto	13
2.2.1.1. Validación cruzada	14
2.2.1.2. Métricas para evaluar la validez de los modelos de clasificación	14
2.2.1.3. Sensibilidad y especificidad	15
2.2.1.4. Valores Predictivos y exactitud	16
2.2.1.5. Razones de verosimilitud	17
3. Métodos	18
3.1. Clasificación sectorizada mediante una CNN	18
3.1.1. Sectorización de pacientes	18
3.1.2. Características de la CNN utilizada para la clasificación de pacientes	19
3.1.3. Características de los entrenamientos clasificación sectorizada	21
3.1.4. Clasificación final de los pacientes	22
3.2. Análisis de las activaciones de la CNN	25
3.2.1. Mapas de activaciones neuronales	25
3.2.1.1. Procesado previo al cálculo de las activaciones de la CNN	25
3.2.1.2. Mapas de activaciones a partir de la excitación de las neuronas	26
3.2.1.3. Eigenfaces: Cálculo de autovectores para el reconocimiento de patrones	30
3.2.2. DeepDream, el algoritmo de Google	30
3.2.3. Algoritmos para la visualización de zonas de activación	31

3.2.3.1.	<i>Mapas de activación Grad-CAM</i>	32
3.2.3.2.	<i>Mapas de oclusión</i>	33
3.2.3.3.	<i>Mapas de oclusión</i>	33
4.	<i>Resultados y discusión</i>	36
4.1.	<i>Prestaciones de la red convolucional para la clasificación sectorizada</i>	36
4.1.1.	<i>Clasificación global de pacientes</i>	37
4.1.1.1.	<i>Scatter plot para los grupos de estudio</i>	40
4.2.	<i>Prestaciones y resultados obtenidos para los diferentes métodos de interpretabilidad de la CNN</i>	44
4.2.1.	<i>Mapas de activaciones</i>	44
4.2.1.1.	<i>Eigenfaces sobre los mapas de activación</i>	44
4.2.1.2.	<i>Aplicación del algoritmo Deep Dream para los cuatro subgrupos de estudio</i>	45
4.2.2.	<i>Visualización de las zonas más relevantes para la clasificación</i>	47
4.2.2.1.	<i>Mapas Grad-CAM</i>	49
4.2.2.2.	<i>Mapas de Oclusión</i>	49
4.3.	<i>Discusión de los resultados cualitativos obtenidos para los mapas de oclusión</i>	62
4.3.1.	<i>Resumen de los Mapas de Oclusión obtenidos en el estudio</i>	62
4.3.2.	<i>Discusión</i>	64
5.	<i>Conclusiones y líneas futuras</i>	65
5.1.	<i>Conclusiones</i>	65
5.2.	<i>Limitaciones y Líneas Futuras</i>	66

ÍNDICE DE FIGURAS

2.1. Ejemplo de funcionamiento de Validación con <i>k</i> -folds para 3 iteraciones	14
2.2. Ejemplo de funcionamiento de Validación Leave One Out para un conjunto de $N=18$ muestras	15
3.1. Esquema que resume el tratamiento de los datos para la generación de las imágenes que nos servirán de entrada para la CNN.	
3.2. Estructura de la CNN utilizada para los entrenamientos	20
3.3. Proceso de entrenamiento del sistema de diagnóstico.	24
3.4. Proceso de entrenamiento para la clasificación por pacientes	24
3.5. Mapa de Activación obtenido para un caso	27
3.6. Mapa de Activación obtenido para un caso	28
3.7. Esquema del procesado de los diferentes mapas de activación	29
3.8. Eigenfaces obtenidas para un caso	31
3.9. Algoritmo para la obtención de los Mapas de Activación de clases	32
3.10. Mapa Grad-CAM obtenido para un caso	34
3.11. Esquema de funcionamiento de la técnica de oclusión	34
3.12. Mapa de Oclusión obtenido para un caso	35
4.1. Scatter Plot para las ventanas de 1800s de Chicas (izquierda) y Chicos (derecha).	
4.2. Scatter Plot para las ventanas de 1800s de secundaria(izquierda) y primaria(derecha).	41
4.3. Scatter Plot para las ventanas de 300s de chicas (izquierda) y chicos (derecha).	42
4.4. Scatter Plot para las ventanas de 300s de secundaria(izquierda) y primaria(derecha).	42
4.5. Scatter Plot para las ventanas de 60s de de chicas (izquierda) y chicos (derecha).	43
4.6. Scatter Plot para las ventanas de 60s de secundaria(izquierda) y primaria(derecha).	43
4.7. Mapas de activación para dos chicas de secundaria en el tramo de actividad diurna.	45
4.8. Mapas de activación para dos chicas de primaria en el tramo de actividad nocturna.	45
4.9. Mapas de Eigenfaces obtenidos para controles(izquierda) y casos(derecha) en la red de secundaria.	46
4.10. Imágenes de activación máxima para un caso (izquierda) y un control(derecha) generadas por el algoritmo DeepDream en la red de secundaria para la actividad nocturna.	46
4.11. Imágenes de activación máxima para un caso (izquierda) y un control(derecha) generadas por el algoritmo DeepDream en la red de chicas para la actividad nocturna.	47
4.12. Imágenes de activación máxima para un caso (izquierda) y un control(derecha) generadas por el algoritmo DeepDream en la red de primaria para la actividad nocturna.	47
4.13. Imágenes de activación máxima para un caso (izquierda) y un control(derecha) generadas por el algoritmo DeepDream en la red de chicos para la actividad nocturna.	48
4.14. Imágenes generadas por el algoritmo DeepDream para nuestras imágenes de entrada para la red de chicas (izquierda) y la red de secundaria, mostrando el resultado para un caso y un control en ambas ejecuciones.	48
4.15. Mapas de Grad-CAM obtenidos para las señales de actividad diurna.	49
4.16. Mapas de Grad-CAM obtenidos para las señales de actividad nocturna.	50

4.17. Ejemplo de espectrograma utilizado para alimentar la CNN.	51
4.18. Mapas de oclusión de actividad diurna para las chicas de secundaria generados para la red entrenada con chicas.	52
4.19. Mapas de oclusión de actividad nocturna para las chicas de secundaria generados para la red entrenada con chicas.	53
4.20. Mapas de oclusión de actividad diurna para las chicas de secundaria generados para la red entrenada con registros de secundaria.	53
4.21. Mapas de oclusión de actividad nocturna para las chicas de secundaria generados para la red entrenada con registros de secundaria.	54
4.22. Mapas de oclusión de actividad diurna para las chicas de primaria para la red entrenada con chicas (CONTROLES).	54
4.23. Mapas de oclusión de actividad diurna de las chicas en primaria para la red entrenada con chicas (CASOS).	55
4.24. Mapas de oclusión de actividad nocturna de las chicas de primaria para la red entrenada con chicas (CONTROLES).	55
4.25. Mapas de oclusión de actividad nocturna de las chicas de primaria para la red entrenada con chicas (CASOS).	56
4.26. Mapas de oclusión de actividad diurna para los chicos de secundaria generados para la red entrenada con registros de secundaria.(CONTROLES)	57
4.27. Mapas de oclusión de actividad diurna para los chicos de secundaria generados para la red entrenada con registros de secundaria.(CASOS)	57
4.28. Mapas de oclusión de actividad nocturna para los chicos de secundaria generados para la red entrenada con registros de secundaria.(CONTROLES)	58
4.29. Mapas de oclusión de actividad nocturna para los chicos de secundaria generados para la red entrenada con registros de secundaria.(CASOS)	58
4.30. Mapas de oclusión de actividad diurna para los chicos de secundaria (controles) en la red de chicos.	59
4.31. Mapas de oclusión de actividad diurna para los chicos de secundaria (casos) en la red de chicos.	59
4.32. Mapas de oclusión de actividad diurna para algunos de los casos (derecha) y controles (izquierda) de los chicos de primaria en la red entrenada con chicos.	60
4.33. Mapas de oclusión de actividad nocturna para algunos de los casos (derecha) y controles (izquierda) en los chicos de primaria en la red de chicos.	61
4.34. Mapas de oclusión de actividad nocturna para algunos de los casos (derecha) y controles (izquierda) en los chicos de primaria en la red de primaria.	61

ÍNDICE DE TABLAS

2.1. Resumen de estudios más relevantes para el TDAH que no están basados en actigrafía (1/2)	7
2.2. Resumen de estudios más relevantes para el TDAH que no están basados en actigrafía (2/2)	8
2.3. Resumen de estudios de actigrafía.	9
2.4. Impacto de las razones de verosimilitud para la utilidad clínica.	17
3.1. Sujetos que forman parte de cada grupo de estudio	19
3.2. Parámetros utilizados en los entrenamientos	21
3.3. Imágenes disponibles para los entrenamientos en cada uno de los grupos	22
3.4. Porcentaje medio de imágenes empleadas en el entrenamiento	23
3.5. Sujetos que forman parte de los conjuntos de train y test en la clasificación de pacientes	23
3.6. Imágenes que maximizan la activación disponibles para los entrenamientos en cada uno de los grupos	26
4.1. Exactitud obtenida del entrenamiento de la CNN para el conjunto de test.	37
4.2. Resultados obtenidos para el conjunto de secundaria.	38
4.3. Resultados obtenidos para el conjunto de primaria.	38
4.4. Resultados obtenidos para el conjunto de chicas.	38
4.5. Resultados obtenidos para el conjunto de chicos.	38
4.6. Exactitud para la clasificación Leave One Out de secundaria.	39
4.7. Exactitud para la clasificación Leave One Out de chicas.	39
4.8. Media de fallos que se producen en la clasificación.	40
4.9. Resumen de los Mapas de oclusión obtenidos para las CNN de chicas.	62
4.10. Resumen de los Mapas de oclusión obtenidos para la CNN de chicos.	62
4.11. Resumen de los Mapas de oclusión para la CNN de primaria.	63
4.12. Resumen de los Mapas de oclusión obtenidos para la CNN de secundaria.	63

Capítulo 1

INTRODUCCIÓN

La utilización de técnicas de Deep Learning ha demostrado ser una herramienta útil para el desarrollo de un método objetivo de diagnóstico del trastorno por déficit de atención e hiperactividad (TDAH). No obstante, debido a la complejidad de dicha patología existen diferencias sustanciales en función de la edad y el sexo de los pacientes diagnosticados. En el presente capítulo, se va a contextualizar este Trabajo de Fin de Máster, donde se ha profundizado en el estudio de dichas técnicas con el objetivo de encontrar patrones en las señales que nos permitan encontrar diferencias en los diferentes grupos de análisis. Indicaremos, asimismo, la estructura del trabajo y hablaremos brevemente de los medios y métodos empleados para su desarrollo.

1.1 MOTIVACIONES

La prevalencia del TDAH actualmente se encuentra en torno al 5% para niños en edad escolar y tiene tendencia a seguir aumentando [1]. Esta elevada prevalencia, hace que sea importante aprender a distinguir sus síntomas a una edad temprana, para garantizar una mejor educación y calidad de vida.

En términos de diagnóstico, existe un mayor número de varones diagnosticados, con una ratio de 3:1 frente a las mujeres [2]. Numerosos estudios han hecho referencia a que esta ratio de diagnóstico se debe a la dificultad de encontrar marcadores claros en las mujeres, lo que hace que la patología pueda pasar desapercibida. Es por ello sumamente importante encontrar un método que nos permita profundizar en estas diferencias y realizar una mejor detección en este sector de población.

Con la edad, los síntomas se atenúan, y a menudo síntomas característicos como la hiperactividad cambian o desaparecen, haciendo difícil detectar la patología si no se diagnosticó previamente. [3]. Es por eso que resulta necesario encontrar un diagnóstico fiable y saber identificar los signos de la patología en esta etapa.

1.2 OBJETIVOS

Las diferencias en la sintomatología de las mujeres respecto a los varones, hace que el diagnóstico del TDAH sea más complicado en las primeras. Además, la influencia de la edad y los cambios que se producen en la manifestación del TDAH con el crecimiento, dificultan aún más dicho diagnóstico. En [4] se presentó un sistema de diagnóstico rápido, fiable y no intrusivo que utilizaba redes neuronales convolucionales (CNN) para realizar una clasificación entre sujetos sanos y pacientes diagnosticados con TDAH. Los resultados que se consiguieron, presentaban porcentajes de sensibilidad, especificidad y exactitud superiores al 90 %.

En este trabajo, se pretende ir un paso más allá, tomando como base el sistema de diagnóstico presentado en [4] para poder encontrar patrones en el comportamiento de la CNN diseñada que podamos asociar con patrones de movimiento en el TDAH. Siguiendo esta premisa, en este trabajo perseguimos:

- Analizar el comportamiento de la CNN diseñada con anterioridad para clasificar grupos reducidos, según edad y sexo.
- Encontrar una metodología que nos permita extraer información útil del comportamiento de la CNN.

1.3 FASES Y MÉTODOS

En esta sección se presenta la mecánica de trabajo que se ha realizado para cumplir los objetivos planteados:

- Sectorización del conjunto completo de pacientes, agrupando por edad y por sexo, con el objetivo de hacer un análisis sobre cuatro grupos de estudio.
- Aplicación de la metodología descrita en [5] para realizar la clasificación de los nuevos grupos empleando el sistema de diagnóstico.
- Desarrollo de una metodología que nos permita extraer mapas de activaciones de la CNN, para caracterizar los patrones de movimiento de los sujetos de estudio mediante dichos mapas.
- Analizar los resultados de la clasificación para grupos.
- Comparar los mapas de activación obtenidos para extraer conclusiones relacionadas con el TDAH.
- Discusión sobre los resultados obtenidos.

1.4 MEDIOS NECESARIOS EMPLEADOS PARA EL DESARROLLO

Para la correcta realización del proyecto planteado, será necesaria la utilización de diferentes herramientas *hardware* y *software*:

Hardware

- Laptop con Windows 10 Home y las siguientes características:
 - Procesador 6xIntel®Core™i7-8750 2.20 GHz.
 - 12 GB de memoria RAM.
 - Disco duro HDD de 1 TB de capacidad.
 - GPU NVIDIA GeForce GTX 1050 con 4 GB de memoria RAM

Software

- MATLAB R2019b [6]: Software de lenguaje de alto nivel que permite el desarrollo de algoritmos, el análisis de datos y computación numérica.
- L^AT_EX : Software utilizado para la composición de textos.

1.5 ESTRUCTURA DEL DOCUMENTO

El presente documento consta de seis capítulos, incluyendo el capítulo de introducción. El contenido de los posteriores capítulos es el que sigue:

Capítulo 2- Antecedentes En este capítulo se presentan los estudios realizados hasta el momento, en términos de la diferenciación en el diagnóstico de grupos sectorizados y se presentan los avances en interpretación de redes neuronales. Además, se explican las diferentes técnicas de reconocimiento de patrones empleadas para la evaluación del sistema experto planteado en el presente trabajo y se analizan diferentes métricas para la evaluación de modelos de clasificación

Capítulo 4- Métodos A lo largo de este capítulo se detalla cómo se ha llevado a cabo la sectorización de pacientes y los procedimientos que se han seguido para realizar la clasificación sectorizada. También se presentan en este capítulo los diferentes métodos para visualizar las activaciones de la CNN.

Capítulo 5- Resultados y discusión Este capítulo está dedicado al estudio de los resultados obtenidos para la clasificación sectorizada y de las diferentes técnicas que se han utilizado para la interpretabilidad de la CNN

Capítulo 6- Conclusiones y líneas futuras En el último capítulo se realiza una visión global del Trabajo fin de Máster, exponiendo aquí las conclusiones extraídas del mismo. Además, se hace mención a las dificultades encontradas a lo largo de su realización y se plantean las líneas de trabajo futuras que pueden surgir a raíz de los resultados obtenidos.

Además se incluye un capítulo en el que se incluyen las referencias bibliográficas que sustentan este proyecto.

Capítulo 2

ANTECEDENTES

En este capítulo, se pretende contextualizar el presente trabajo, haciendo un repaso de los métodos de diagnóstico y presentando las técnicas que hemos utilizado para evaluar el rendimiento de nuestro sistema. Dividiremos el capítulo en dos secciones:

- **Estado del Arte**, donde haremos un repaso de los estudios realizados hasta el momento.
- **Técnicas y métricas**, donde presentaremos las técnicas utilizadas en el presente trabajo.

2.1 ESTADO DEL ARTE

En esta sección se va a hacer un breve repaso de los diferentes estudios relacionados con este trabajo, centrándonos principalmente en cuatro áreas:

- Resumen de los métodos de diagnóstico objetivos hasta el momento
- Repaso de los estudios centrados en las diferencias en sujetos diagnosticados con TDAH por sexo y edad en base a los métodos tradicionales.
- Estudios relacionados con el TDAH y las diferencias relativas al sexo y la edad de los pacientes.
- Estudios que introduzcan técnicas de interpretabilidad de redes neuronales en el campo de las señales biomédicas.

2.1.1 RESÚMEN DE LOS MÉTODOS DE DIAGNÓSTICO OBJETIVO DEL TDAH

El problema del diagnóstico del TDAH ha estado presente en numerosos estudios. EL método de diagnóstico tradicional y que se utiliza actualmente consiste en la aplicación de los criterios marcados en el *Diagnostic And Statistical Manual of Mental Disorders (DSM-V)*. Este método condiciona fuertemente el diagnóstico debido a la subjetividad de las respuestas del entorno del paciente a la evaluación llevada a cabo por los profesionales.

En [7] y [4] también se desarrollaron métodos para poder llevar a cabo un diagnóstico objetivo del TDAH. En estos estudios, así como en [5] se llevo a cabo un análisis exhaustivo de dichos métodos y los resultados que se obtuvieron.

A continuación y a modo de resumen, se incluyen actualizadas las tablas que aparecían en [5] y que aglutinan muy bien la información existente hasta el momento acerca del diagnóstico objetivo.

En las tablas 2.1 y 2.2, se muestran los resultados no basados en actigrafía y en la tabla 2.3 un resumen de los estudios de actigrafía. Los resultados se expresan como lo hacen los autores en los trabajos citados.

F. Información	Autores	Sujetos	Métodos	Resultados
Informes Parentales	Cortese et al. [8]	722 pacientes con TDAH y 683 para el grupo de control	Informes subjetivos del sueño proporcionados por los padres	Los mejores resultados obtienen $p < 0,001$ para la resistencia a la hora de dormir, el inicio del sueño y la dificultad para despertarse por la mañana
Análisis Químico	Tuthill et al. [9]	16 pacientes con TDAH y 261 sujetos sanos	Nivel de plomo en el pelo	$p = 0,006$
	Weiss et al. [10]	277 pacientes con TDAH y 106 sujetos sanos	Nivel de TSH en la sangre	Se encontraron niveles anormales en el 2,2% de los pacientes con TDAH y en el 1% de sujetos sanos
	El-Faddagh et al. [11]	129 mujeres y 126 hombres a lo largo de 11 años	Alelo Exon III gen DRD4	$p = 0,061$ en 4.5 y.o., $p = 0,026$ en 8 y.o., $p = 0,005$ en 11 y.o.
	Van Veen et al. [12]	40 sujetos diagnosticados como TDAH y 40 controles emparejados	Melatonina en saliva	Ritmo circadiano retrasado en pacientes con TDAH ($p < 0,0005$)
Imagen Médica	Kurtek et al. [13]	19 pacientes con TDAH y 15 controles	Análisis de forma de estructuras cerebrales a partir de resonancia magnética	Exactitud = 0,9118
	Ashtari et al. [14]	18 pacientes con TDAH y 15 sujetos sanos	DT-MRI	Los pacientes con TDAH muestran valores más bajos ADHD patients FA (Anisotropía fraccional) en el cerebro medio: $p < 0,01$
	Silk et al. [15]	15 pacientes con TDAH y 15 sujetos sanos	DT-MRI	Los pacientes con TDAH mostraron valores más altos de FA en diferentes regiones del cerebro: región occipito-parietal derecha $p < 0,001$, estriado izquierdo $p < 0,001$, inf. Izquierdo. región temporal $p < 0,005$, derecha inf. región parietal $p < 0,001$ e izquierda inf. región frontal $p < 0,05$
	Zimmer et al. [16]	Encuesta de diferentes estudios	PET/SPECT	El metabolismo de los pacientes con TDAH presenta varias anomalías: glucosa, dopamina (catecolaminas en general) y noradrenalina.
	Riaz et al. [17]	<i>NeuroBureau ADHD-200</i>	Un tipo de resonancia magnética funcional (fMRI) conocida como <i>Resting State fMRI</i> (rsfMRI o R-fMRI)	Exactitud = 0.818, Sensibilidad=1, Especificidad = 0.75
	Sun et al. [18]	83 niños emparejados por edad y sexo con TDAH recién diagnosticado y nunca tratado (40 con TDAH inatento y 43 con TDAH combinado) y 87 sujetos control sanos	Resonancia magnética anatómica y tensor de difusión Imágenes (MR)	No se encontraron diferencias globales entre los niños con TDAH y los sujetos control en el volumen cerebral total o en el volumen total de materia gris y blanca. Exactitud = 0.737 (discriminación de pacientes) y Exactitud = 0.801 (diferentes subtipos)
	Señales Biomédicas	Modarres-Zadeh et al. [19]	5 pacientes con TDAH y 5 controles emparejados antes y después de la medicación	Análisis espectral EEG y test de comportamiento
	Gutierrez et al. [20]	4 pacientes con TDAH no medicados	Relación de la densidad espectral de potencia en las bandas <i>beta</i> y <i>Theta</i> del EEG	La relación se altera en los pacientes con TDAH pero tiende a valores normales después del entrenamiento
	Hyun-Lee et al. [21]	13 niños en tres grupos: TDAH, retraso mental y grupo de control	Análisis de wavelets de EEG y mapas autoorganizados	Exactitud = 0,6
	Dong-Ming et al. [22]	6 jóvenes	Análisis de EEG por entropía multiscales	Se detectaron diferentes estados de atención con $rmExactitud = 0,63158$, lo que revela una alta complejidad en la actividad de EEG incluso cuando no se realiza ninguna actividad

TABLA 2.1: Resumen de estudios más relevantes para el TDAH que no están basados en actigrafía (1/2)

F. Información	Autores	Sujetos	Métodos	Resultados
Señales Biomédicas	Nan et al. [23]	6 personas sanas	Análisis del desplazamiento de información mediante EEG cuando se cambia de atención sostenida a falta de atención	$p < 0,005$ en bandas α y β , $p < 0,05$ en la banda θ
	Hale et al. [24]	35 adultos con TDAH y 104 adultos sanos	Asimetría del EEG en la banda β	Asimetría atípica en el TDAH ($p = 0,00001$)
	Ghasssemi et al. [25]	10 adultos con TDAH y 40 adultos sano	Características morfológicas de las señales ERP	Exactitud = 0,92 en la banda γ
	Castro-Cabrera et al. [26]	23 sujetos y 23 controles	Análisis de características morfológicas, wavelets y características no lineales de señales ERP.	Morphological: Sensibilidad = 0,96, Especificidad = 0,65, Exactitud = 0,81 \pm 0,1100, AUC = 0,81; dinámicas no lineales: Sensibilidad = 0,69, Especificidad = 0,78, Exactitud = 0,7391 \pm 0,1200, AUC = 0,72; wavelets: Sensibilidad = 1, Especificidad = 0,78, Exactitud = 0,8695 \pm 0,0973, AUC = 0,91; mejor combinación de características: Sensibilidad = 0,96, Especificidad = 0,87, Exactitud = 0,9130 \pm 0,0100, AUC = 0,94
	Demanelle et al. [27]	2 pacientes con TDAH y 2 controles emparejados	Sincronía de fase de señales MEG adquiridas en regiones distantes del cerebro.	No reportado
Öztoprak et al. [28]	Entre 37 y 44 niños en el grupo con TDAH y 32 a 38 niños en el grupo de control sano.	Medidas ERP en las señales EEG durante la realización de una tarea de tipo Stroop.	Exactitud = 0.995 e (train) Exactitud = 1 (10 sujetos test)	
PSG	Gruber et al. [29]	32 pacientes con TDAH y 64 sujetos sanos adultos durante 5 noches consecutivas	Informes de sueño	Los informes de sueño pueden ser usados para el diagnóstico del TDAH
	Silvestri et al. [30]	16 sujetos diagnosticados como TDAH inactivo y 39 con otros tipos de TDAH	Informes de sueño	Tanto los aspectos macro como microestructurales del sueño presentaron diferencias significativas. El sueño de los sujetos con TDAH presentó más trastornos. Las mayores diferencias presentaron $p < 0,0001$
	Cortese et al. [8]	Meta-Analysis incluyendo 722 sujetos con TDAH y 683 controles	Informes de sueño	Los sujetos con TDAH presentan más trastornos del sueño La característica más informativa presento $p < 0,001$
Análisis del movimiento mediante infrarrojos	Teicher et al. [31]	17 sujetos diagnosticados con TDAH y 10 sujetos sanos	Evaluación de los movimientos de la cabeza mediante dispositivos infrarrojos	Sensibilidad= 0,94, Especificidad=1 logrado mediante la combinación de la complejidad de los movimientos de la cabeza evaluados mediante el análisis fractal con otras variables significativas (edad, puntuación en una tarea continuada)
	Teicher et al. [32]	18 pacientes con TDAH y 11 de contro	Evaluación de los movimientos de la cabeza mediante dispositivos infrarrojos	Los movimientos en la cabeza para pacientes con TDAH son más frecuentes ($p < 0,002$), 3,4 veces más frecuentes $p > 0,01$, al cubrir áreas grandes ($p > 0,001$) y más lineales y menos complejos ($p > 0,00004$)

TABLA 2.2: Resumen de estudios más relevantes para el TDAH que no están basados en actigrafía (2/2)

F. Información	Autores	Sujetos	Métodos	Resultados
Actividad Diurna	Dane et al. [33]	Se evaluaron 42 sujetos TDAH (20 de tipo inatento, 22 de tipo combinado) y 22 controles.	Registros de Actigrafía de 24 horas de duración promedio en 2h de duración	Solo diferencias durante la tarde. Los tipos de TDAH no presentaron diferencias en el análisis actigráfico.
	Teicher et al. [34]	11 pacientes con TDAH y 12 controles	Registros de Actigrafía de 72 horas promediados y muestreados cada 5 minutos	La distribución de las puntuaciones de actividad (porcentaje de tiempo empleado en actividad baja, media o alta) constituye un índice relevante para el diagnóstico de TDAH (la distribución de la actividad es significativamente sesgado a la derecha en pacientes con TDAH)
	Kam et al. [35]	10 pacientes con TDAH y 7 controles	3 horas de registros de actigrafía muestreados a 32 Hz	La distribución promedio obtenida a partir de los histogramas de actividad de 1 minuto presentó mayor nivel de actividad y variabilidad en el grupo de TDAH ($p < 0,05$)
	Martín et al. [7, 36]	31 pacientes con TDAH y 35 controles de 6 años de edad	Registros de Actigrafía durante 24 h muestreados a 1 Hz	Exactitud = 0.8571 , Sensibilidad = 0.9500 y Especificidad = 0.7727
	Muñoz et al. [37]	22 pacientes, 11 diagnosticados como TDAH y 11 sujetos sanos	Dos acelerómetros: uno de muñeca y otro de tobillo para analizar los datos obtenidos durante 6 horas escolares y una red convolucional.	Exactitud = 0.8570 , Sensibilidad = 0.6 y Especificidad = 1 para la muñeca y Exactitud= 0.937, Sensibilidad =0.8 y Especificidad=1 para el tobillo
	Mahony et al. [38]	43 pacientes, 24 con TDAH y 19 controles de 6 a 11 años de edad	Señales de actigrafía obtenidas mediante IMUs o unidades de medición inercial, durante 1h	Exactitud = 0.9512 , Sensibilidad = 0.9444 y Especificidad = 0.9565
	Amado et al. [4]	148 pacientes, 73 con TDAH y 75 controles de 6 a 15 años de edad	Registros de Actigrafía durante 24 h muestreados a 1 Hz utilizando redes neuronales convolucionales para clasificar espectrogramas de ventanas de actividad.	Exactitud = 0.9857 , Sensibilidad = 0.9762 y Especificidad = 0.9952
Informes de Sueño	Cortese et al. [8]	722 sujetos componen el grupo de TDAH y 683 componen el grupo de control	Los informes Actigráficos del sueño	Los niños con TDAH presentaron una menor eficiencia del sueño, un tiempo de sueño real y tiempos promedio para quedarse dormidos. Los resultados más significativos arrojaron $p < 0,001$
	Gruber et al. [29]	38 niños con TDAH y 64 controles durante 5 noches consecutivas	Informes de sueño actigráfico	Se encontró un aumento de la inestabilidad en el inicio del sueño, la eficiencia del sueño y el sueño verdadero en pacientes con TDAH
Ritmo biológico circadiano	Van Veen et al. [12]	40 sujetos diagnosticados como TDAH y 40 controles emparejados	Registros actimétricos de 24 horas	Comienzo y finalización retardados del sueño en sujetos con TDAH. Mayor relevancia lograda para el inicio del sueño: $p = 0,006$

TABLA 2.3: Resumen de estudios de actigrafía.

2.1.2 DIFERENCIAS POR SEXO Y EDAD EN EL DIAGNOSTICO DEL TDAH SEGÚN LOS MÉTODOS TRADICIONALES

Existen diversos estudios recientes, que basados en este método de diagnóstico han resaltado la influencia de la edad y el sexo en la manifestación de la patología. En esta sección, repasaremos algunos de ellos.

En el estudio realizado por Murray *et al.* [39] se encontró que las mujeres tenían más probabilidades de mostrar síntomas en los primeros años de la adolescencia, mientras que los hombres mostraban síntomas desde su niñez. Se encontró también relación entre el diagnóstico tardío de las chicas con patrones de hiperactividad e impulsividad, debido a la aparición tardía de los síntomas.

En un metanálisis realizado en 2021, Loyer *et al.* encontraron diferencias en la manifestación del TDAH en función del sexo en niños y adolescentes. Se encontró que los niños con TDAH eran más hiperactivos que las niñas con TDAH y los niños tenían más dificultades en términos de inhibición de la respuesta motora y flexibilidad cognitiva. [40]

2.1.3 UTILIZACIÓN DE SEÑALES BIOMÉDICAS PARA EL ESTUDIO DEL TDAH

Se entiende por señales biomédicas aquellas señales originadas por el cuerpo humano que se utilizan para el diagnóstico o la investigación médica. Para el diagnóstico del TDAH, tiene especial interés el **electroencefalograma**(EEG). Esta técnica no invasiva permite el estudio del sistema nervioso central a través de señales eléctricas generadas en los impulsos nerviosos que son recogidas a través de electrodos.

En relación al análisis del TDAH en subgrupos de edad y sexo, hay varios estudios que emplean las EEG junto con señales de actividad electrodérmica (EDA); que se refiere a la variación continua en las características eléctricas de la piel, para la caracterización de las diferencias entre sexos en la adolescencia. En [41], Hermens *et al.* sugirieron que los procesos psicofisiológicos podrían ser fundamentales para caracterizar el TDAH en cada sexo, y encontraron evidencias de que los déficits en los sistemas de excitación / atención venían marcados en la expresión conductual del TDAH en cada sexo. Lazzaro *et al.* [42]. encontraron que los adolescentes manifestaban estados de baja actividad no presentes en los niños.

En otro estudio, en el que se emplearon señales EEG para caracterizar las diferencias entre los subtipos de TDAH en función del sexo [43], se encontró que el componente hiperactivo / impulsivo maduran con la edad y que la falta de atención parece ser más generalizado con el aumento de la edad. En cuanto a las diferencias de sexo, se encontraron que las diferencias de grupo eran más pequeñas en mujeres que en hombres.

2.1.4 EL USO DE LA MRI EN EL DIAGNÓSTICO DEL TDAH

La **imagen por resonancia magnética** (MRI) consiste en la utilización de campos magnéticos (en lugar de radiaciones ionizantes) para la obtención de las diferentes imágenes. En relación con nuestro problema, existen estudios en los que se han encontrado diferencias en como se

ve afectado el cerebro de un paciente diagnosticado con TDAH en función del sexo, donde se hayaron diferencias en las redes fronto-subcorticales, estando estas más afectadas en las niñas. [44]

2.1.5 ESTADO DEL ARTE: LA INTERPRETACIÓN DEL DEEP LEARNING

Uno de los objetivos que se persigue con este trabajo es desarrollar un método que nos permita interpretar el comportamiento de la CNN en términos de las señales de actividad. Existen algunos trabajos que han utilizado técnicas para realizar estudios similares en relación con el diagnóstico de enfermedades. En concreto dividiremos los estudios en dos bloques:

- Estudios que realizan interpretación de redes neuronales en el diagnóstico de enfermedades.
- Estudios que aplican interpretabilidad en deep learning para el diagnóstico del TDAH.

2.1.5.1 INTERPRETACIÓN DE REDES NEURONALES EN EL DIAGNÓSTICO DE ENFERMEDADES

Existen varios estudios recientes que aplican metodologías de interpretación de redes, utilizando como señales de entrada señales EEG e imágenes MRI.

En [45], Jonas et al. utilizan señales EEG junto con una red neuronal convolucional para predecir el pronóstico de pacientes en coma tras haber sufrido un paro cardíaco. Una vez realizada la clasificación, utilizan el cálculo de mapas de activación de clases ponderados por gradiente (*Grad-CAM*) para identificar las características de las señales EEG que la red utilizó para la correcta clasificación, obteniendo resultados prometedores. La técnica de (*Grad-CAM*) también ha sido empleada por Uyttenhove et al. al. para interpretar los resultados de una red convolucional que detecta sujetos que padecen epilepsia [46], obteniendo como resultado que dicha red es capaz de aprender características sensibles con conexiones a marcadores de epilepsia conocidos.

Existen estudios que han utilizado imágenes por resonancia magnética como señales de entrada de redes neuronales para su posterior interpretación. En concreto, Cai et al. utilizan una versión mejorada del mapa de activación de clases (CAM) para encontrar texturas en las imágenes MRI que se correlacionen con los subtipos de distrofia muscular. [47]. En un estudio sobre las migrañas, se encontraron patrones de activación diferentes en las MRI del cerebro usando redes convolucionales y CAM. [48]. También se han utilizado los mapas de activación de clases, en este caso Grad-CAM para encontrar patrones que ayuden a mejorar el diagnóstico del cáncer de mama mediante MRI [49].

Otro estudio, realizado en 2018, utilizaba señales de electroencefalograma (EEG), electromiograma (EMG) y electrooculograma (EOG) para entrenar una red convolucional y después obtener los mapas de activación mediante Grad-CAM guiado para identificar características en estas señales [50].

2.1.5.2 INTERPRETABILIDAD DEL DEEP LEARNING PARA EL DIAGNÓSTICO DEL TDAH

En el estudio planteado por Chen et *al.* al. en 2019 emplearon redes convolucionales y mapeo de activación de clases ponderado por gradiente (Grad-CAM) para detectar anomalías de frecuencia espacial personalizadas en EEG de niños con TDAH. Las señales de entrada utilizadas para este estudio fueron señales EEG [51]. El estudio pudo detectar anomalías EEG de niños con TDAH visualizando el proceso de toma de decisiones y encontró diferencias significativas con respecto al grupo de control.

2.1.6 ANÁLISIS Y COMPARATIVA

En este capítulo, se ha realizado un repaso de diversos estudios en las dos líneas de investigación en las que se enmarca este trabajo: el diagnóstico del TDAH para grupos diferenciados de sexo y edad y el empleo del deep learning para obtener información útil en el diagnóstico de enfermedades. Según lo que hemos visto a lo largo de este capítulo, podemos afirmar que:

- Los estudios que hablan de las diferencias de los síntomas por sexo y edad, coinciden en una déficit en el diagnóstico en las niñas.
- Las técnicas de visualización de mapas de activaciones en redes neuronales se han utilizado en el estudio de diversas enfermedades, pero no existen estudios conocidos que utilicen actimetría para ello.
- Aunque si se ha empleado *Deep Learning* para diseñar métodos objetivos de diagnóstico para el TDAH, no hemos encontrado ninguno que aplique metodologías de interpretación.

Con todo lo planteado, podemos justificar nuestra propuesta de desarrollo de un sistema de diagnóstico experto utilizando la sectorización de pacientes y técnicas de *Deep Learning*.

2.2 TÉCNICAS Y MÉTRICAS PARA EVALUAR EL RENDIMIENTO DE UN SISTEMA DE RECONOCIMIENTO DE PATRONES

El **Reconocimiento de Patrones** es la disciplina científica que estudia los procesos que permiten extraer información de un conjuntos de datos, con el objetivo de llevar a cabo acciones posteriores como clasificación de los datos en unas determinadas categorías o labores de regresión para predecir un determinado valor no observado. En lo que sigue, y dadas las características de este trabajo, nos centraremos en la tarea de clasificación. Dicho esto, en un sistema de reconocimiento de patrones se realizan las siguientes etapas:

- **Adquisición de datos de entrada.**
- **Extracción de características** a partir de los datos de entrada. Se espera que estas características resuman la información contenida en los datos y que mantengan poder discriminante suficiente para que sean útiles en la etapa de clasificación.

- **Clasificación** del conjunto de datos de estudio en grupos o clases, basándose en las características extraídas.

En esta sección se presentarán las técnicas y métricas que se han utilizado en la evaluación del sistema experto planteado en este trabajo.

2.2.1 TÉCNICAS PARA LA EVALUACIÓN DEL RENDIMIENTO DE UN SISTEMA EXPERTO

Existen diversas técnicas que permiten evaluar el rendimiento en los sistemas de reconocimiento de patrones. En este trabajo nos centramos en el aprendizaje supervisado, de forma que el sistema que diseñaremos parte de la disposición de una base de datos *suficientemente* grande de ejemplos etiquetados como para que el sistema experto pueda aprender a partir de ellos. Entendemos por *etiquetados*, en nuestro dominio de aplicación, señales de actigrafía que vengan acompañadas de un diagnóstico (como sano –control– o enfermo –caso–), así como de información demográfica adicional como edad o sexo.

Pues bien, el aprendizaje consiste en ajustar los parámetros libres del sistema experto a partir de un subconjunto de datos de la base de datos que acabamos de mencionar. Al subconjunto empleado para entrenar se le denomina *conjunto de entrenamiento*. Normalmente se acompaña el conjunto de entrenamiento de un conjunto de *validación*, el cual permite seleccionar los *metaparámetros* de nuestro problema, por ejemplo, la arquitectura de clasificador a seleccionar –si se han entrenado varias arquitecturas– o cualquier otro parámetro que no se fije mediante el entrenamiento. Finalmente, las prestaciones del sistema de reconocimiento de patrones se miden a partir de parámetros de rendimiento cuyos valores se obtienen haciendo funcionar el sistema experto finalmente elegido sobre un *conjunto de test*. Los tres subconjuntos que acabamos de mencionar (entrenamiento, validación y test) deben ser disjuntos, con el objetivo de que las medidas sean fiables. De no serlo, es decir, de coincidir datos, pongamos de entrenamiento y test, las medidas obtenidas serán, con toda seguridad, excesivamente optimistas.

En el caso en el que las bases de datos sean muy grandes, emplear este procedimiento para una determinada selección de datos en los tres conjuntos podría ser suficiente. Ahora bien, cuando estas bases de datos son pequeñas, hecho que ocurre con frecuencia en aplicaciones clínicas donde no es sencillo reclutar a elevadas cantidades de pacientes, el hecho de que los datos se seleccionen de una u otra manera para construir estos tres conjuntos puede tener relevancia, de tal manera que una ordenación podría traer consigo resultados pesimistas, pero otra, por meras razones de tamaño muestral, podría traer consigo unos resultados optimistas. Por este motivo, se recurre habitualmente a un procedimiento consistente en la validación cruzada. Esta técnica consiste en llevar a cabo múltiples asignaciones de datos a los diferentes conjuntos de entrenamiento, validación y test y, para cada una de esas ordenaciones, medir las prestaciones del sistema. Con ello conseguimos no una (única) medida de las prestaciones, sino un muestro de la función de densidad del parámetro de medida de prestaciones que estemos empleando, lo cual nos permite proporcionar una medida en términos estadísticos y así aproximarnos mejor al verdadero valor del parámetro que estamos buscando.

Dentro de la validación cruzada podemos distinguir, asimismo, varias técnicas. Dos de ellas son la validación cruzada con *k-folds* y, un caso particular y muy singular del anterior, la validación

cruzada *Leave One Out*, que por su popularidad merece un análisis separado.

2.2.1.1 VALIDACIÓN CRUZADA

Como ya hemos apuntado, la validación cruzada se caracteriza por la utilización del análisis estadístico para la evaluación de los resultados, garantizando así la independencia entre las particiones de entrenamiento y test. Su principio de funcionamiento consiste en realizar un análisis variando los conjuntos de entrenamiento y test, manteniendo la misma proporción, para después calcular, por ejemplo, la media aritmética de los resultados. En función de cómo se distribuyen las particiones, tendremos diferentes tipos de validación cruzada. Para el desarrollo de nuestro sistema experto hemos empleado:

- Validación cruzada con k-folds:** Consiste en dividir los datos de muestra en k subconjuntos, empleando uno de ellos para los datos de test y el resto ($k-1$) para los datos de entrenamiento. Este proceso se repite durante k iteraciones, hasta que se han evaluado todos los conjuntos de test disponibles. Mediante esta técnica, tenemos la ventaja de que se evalúan todos los datos, pero la carga computacional es alta. Podemos ver en la figura 2.1 un ejemplo de funcionamiento de este tipo de validación.

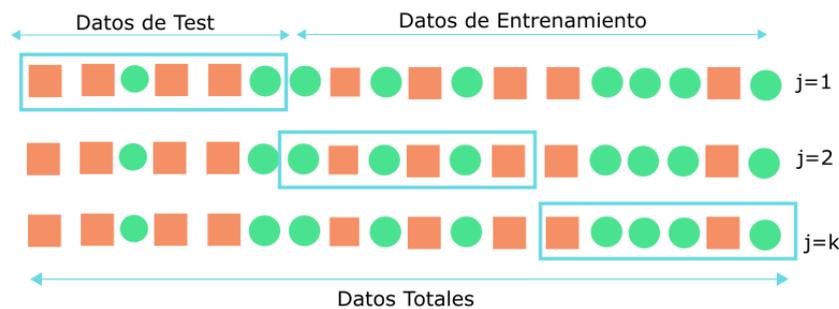


FIGURA 2.1: Ejemplo de funcionamiento de Validación con k-folds para 3 iteraciones

- Validación *Leave One Out*:** En este caso, los datos se separan de forma que cada iteración consta de una única muestra en el conjunto de test. Esta técnica nos permite ajustar muy bien el error, pero la carga computacional es elevada debido a que el número de iteraciones es directamente proporcional al número de muestras del que disponemos. Su diagrama de funcionamiento puede verse en la figura 2.2, donde se ha patente que este método es el caso particular del anterior para $k=N$, con N el tamaño del registro de datos.

2.2.1.2 MÉTRICAS PARA EVALUAR LA VALIDEZ DE LOS MODELOS DE CLASIFICACIÓN

A la hora de evaluar las propiedades de un sistema de diagnóstico, existen diversas métricas que nos proporcionan una idea del comportamiento del mismo. En función de la interpretación que podemos hacer de los resultados obtenidos para estos indicadores, podemos hablar de dos grupos diferenciados:

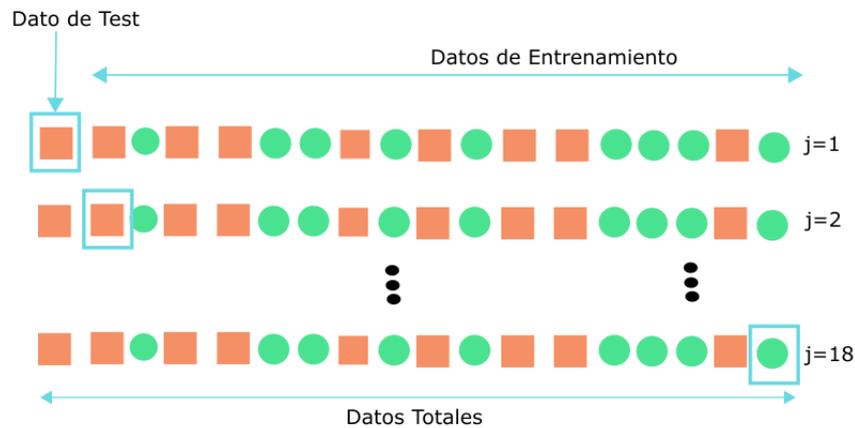


FIGURA 2.2: Ejemplo de funcionamiento de Validación Leave One Out para un conjunto de $N=18$ muestras

- **Intrínsecos del desempeño de la prueba** Permiten evaluar una prueba diagnóstica sin tener en cuenta la prevalencia de la patología en la población estudiada. En el presente estudio, se han analizado la sensibilidad y la especificidad.
- **Indicadores de la eficacia real de la prueba** : En este grupo podemos incluir los valores predictivos, que si bien responden a la probabilidad de que la persona efectivamente esté enferma ante un test positivo, variarán enormemente con la población estudiada. También se puede incluir la exactitud, que mide la probabilidad de detectar la presencia o ausencia de enfermedad.

Para comprender el significado correcto de estas métricas, debemos distinguir entre cuatro grupos de pacientes:

- **Verdaderos Positivos (VP)**: Pacientes cuyo resultado en una prueba diagnóstica es positivo y realmente están afectados por la enfermedad estudiada.
- **Falsos Negativos (FN)**: Aquellos pacientes afectados por la enfermedad que no son detectados como positivos por la prueba diagnóstica.
- **Falsos Positivos (FP)** Se refiere a los pacientes para los que la prueba diagnóstica da un resultado positivo y realmente son sujetos sanos.
- **Verdaderos Negativos (VN)**: Sujetos sanos para los que la prueba diagnóstica es negativa.

2.2.1.3 SENSIBILIDAD Y ESPECIFICIDAD

Son independientes de la prevalencia de la enfermedad en la población estudiada y pueden definirse como:

- **Sensibilidad:** Se entiende como la capacidad de detección de la enfermedad, es decir, nos proporciona la probabilidad de clasificar correctamente a un enfermo.

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (2.1)$$

- **Especificidad:** Se entiende como la probabilidad de obtener un resultado negativo para un sujeto sano, es decir, la probabilidad de clasificar correctamente a los sujetos sanos.

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (2.2)$$

Para considerar buena una prueba diagnóstica, los resultados de sensibilidad y especificidad deben ser superiores al 80 %.

2.2.1.4 VALORES PREDICTIVOS Y EXACTITUD

Los valores predictivos miden la eficacia real de una prueba diagnóstica. Dependen de la prevalencia de la enfermedad y proporcionan información sobre la probabilidad de padecer o no una enfermedad cuando se conoce el resultado de la prueba diagnóstica.

- **Exactitud:** Mide la probabilidad de que el sistema diagnóstico prediga correctamente la presencia o ausencia de la enfermedad.

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.3)$$

- **Valor Predictivo Positivo (VPP):** Se refiere a la probabilidad de padecer la enfermedad si el resultado de la prueba diagnóstica es positivo.

$$\text{VPP} = \frac{VP}{FP + VP} \quad (2.4)$$

- **Valor Predictivo Negativo (VPN):** Indica la probabilidad de que un sujeto esté sano ante un resultado negativo de la prueba diagnóstica.

$$\text{VPN} = \frac{VN}{VN + FN} \quad (2.5)$$

Los valores predictivos permiten evaluar el comportamiento de la prueba diagnóstica en una determinada población, por lo que son útiles para medir la relevancia de la sensibilidad y la especificidad en dicha población. No obstante, al estar los resultados altamente relacionados con la prevalencia de la enfermedad, no aportan utilidad clínica, por lo que resulta necesario encontrar métricas robustas para este fin, como las razones de verosimilitud.

LR+	LR-	Utilidad
>10	<0.1	Altamente relevante
5-10	0.1-0.2	Buena
2-5	0.5-0.2	Regular
<2	>0.5	Mala

TABLA 2.4: Impacto de las razones de verosimilitud para la utilidad clínica.

2.2.1.5 RAZONES DE VEROSIMILITUD

La razón de verosimilitud o *Likelihood Ratio* (LR) se entiende como la razón entre la posibilidad de observar un resultado en los pacientes con la enfermedad frente a la posibilidad de obtener el mismo resultado en pacientes sin la patología.[52]

Este criterio tiene una gran utilidad clínica para la toma de decisiones, ya que son valores independientes respecto a la prevalencia de la enfermedad. En términos de sensibilidad y especificidad, las razones de verosimilitud se pueden calcular aplicando las expresiones 2.6 y 2.7

$$LR(+)=\frac{Sensibilidad}{1-Especificidad} \quad (2.6)$$

$$LR(-)=\frac{1-Sensibilidad}{Especificidad} \quad (2.7)$$

En función del valor obtenido para cada parámetro, podremos valorar la utilidad de la prueba diagnóstica. En la tabla 2.4 se resumen los intervalos, y la valoración subjetiva asociada, para estos parámetros:

Capítulo 3

MÉTODOS

El objetivo de este trabajo es diseñar un método de diagnóstico que nos permita encontrar una explicación a las diferencias entre pacientes sanos y pacientes diagnosticados con TDAH a partir del funcionamiento interno de las CNN. Para ello, realizaremos primero un estudio de las prestaciones de la red convolucional que se diseñó en [5], pero centrándonos en grupos de estudio reducidos, y con los resultados obtenidos se aplicarán diferentes metodologías sobre la red para extraer características de la misma. En este capítulo, nos centraremos en detallar el procedimiento que se ha empleado para el desarrollo del sistema, concretamente:

- Presentación de los grupos de estudio y los criterios establecidos para crear los conjuntos de entrenamiento y test de la CNN.
- Adaptación de los resultados obtenidos de las CNNs para poder aplicar las diferentes metodologías de interpretación.
- Análisis de las activaciones de las capas de la red neuronal mediante diferentes técnicas de interpretación de redes neuronales.

3.1 CLASIFICACIÓN SECTORIZADA MEDIANTE UNA CNN

En estudios anteriores [5] se presentó un método de diagnóstico basado en la utilización de señales actigráficas para entrenar una CNN. Dichas señales han sido obtenidas mediante un actígrafo de muñeca (*ActiGraph GT3x*). Este actígrafo se encarga de medir la aceleración en los tres ejes cartesianos, registrando una muestra cada segundo (*frecuencia de funcionamiento de 1 Hz*). Vamos a utilizar este sistema para realizar una clasificación por grupos, haciendo una sectorización por sexo y edad.

3.1.1 SECTORIZACIÓN DE PACIENTES

Disponemos de información actigráfica de un total de 148 pacientes, de los cuales 73 han sido clasificados como afectados por TDAH según el *Diagnostic And Statistical Manual Of Mental Disorders (DSM-V)* [1] y a los que denominaremos *Casos*. Los 75 restantes se conocerán ahora

como *Controles*. Para el estudio finalmente se han utilizado 70 casos y 69 controles, debido a que no se disponía de información suficiente para la realización de este nuevo estudio en todos los pacientes.

El TDAH es más frecuente en hombres que en mujeres, lo que ha llevado a diversos autores [53, 54, 55] a estudiar las diferencias existentes en cómo se manifiesta esta patología en ambos grupos, por lo que resulta de interés que procedamos nosotros de una manera similar, de cara a encontrar diferencias en la actividad de casos y controles, en su caso, en función del sexo. Como terminología, nos referiremos como *chicos* a los varones y como *chicas* a las mujeres. Lo mismo ocurre con las diferencias en función de la edad [3], distinguiendo entre dos etapas: la equivalente a los estudios de educación primaria (6-11 años) y la adolescencia. En nuestro caso, contamos con información de sujetos entre los 6 y los 15 años por lo que haremos la división tomando por un lado los sujetos entre 6 y 11 años, grupo al que llamaremos *primaria*, y los sujetos entre 12 y 15 años, los cuales formarán parte del grupo de *secundaria*. En la tabla 3.1 se presenta la división entre casos y controles para cada uno de los grupos.

GRUPO DE ESTUDIO	CASO	CONTROL
<i>Chicas</i>	14	13
<i>Chicos</i>	56	56
<i>Secundaria</i>	11	8
<i>Primaria</i>	59	61

TABLA 3.1: Sujetos que forman parte de cada grupo de estudio

3.1.2 CARACTERÍSTICAS DE LA CNN UTILIZADA PARA LA CLASIFICACIÓN DE PACIENTES

En [5] se utilizaron las señales de actigrafía para obtener mediante el espectrograma una representación tiempo-frecuencia de la señal y crear a partir de ella imágenes 2D para entrenar la CNN. Para la generación de dichas imágenes, se siguió el siguiente proceso:

- Primero se crearon dos registros para diferenciar entre la actividad nocturna y la diurna.
- Se fragmentaron las señales en ventanas de 1800, 300 y 60 segundos, correspondiente a actividades de larga, media y corta duración.
- Por último, se crearon espectrogramas a partir de estas ventanas.

En el esquema de la figura 3.1 se resume este proceso. Una vez establecidos los diferentes conjuntos de entrenamiento, se realizaron seis entrenamientos diferentes, uno por cada grupo de ventanas en el día y en la noche, sobre la CNN representada en la figura 3.2

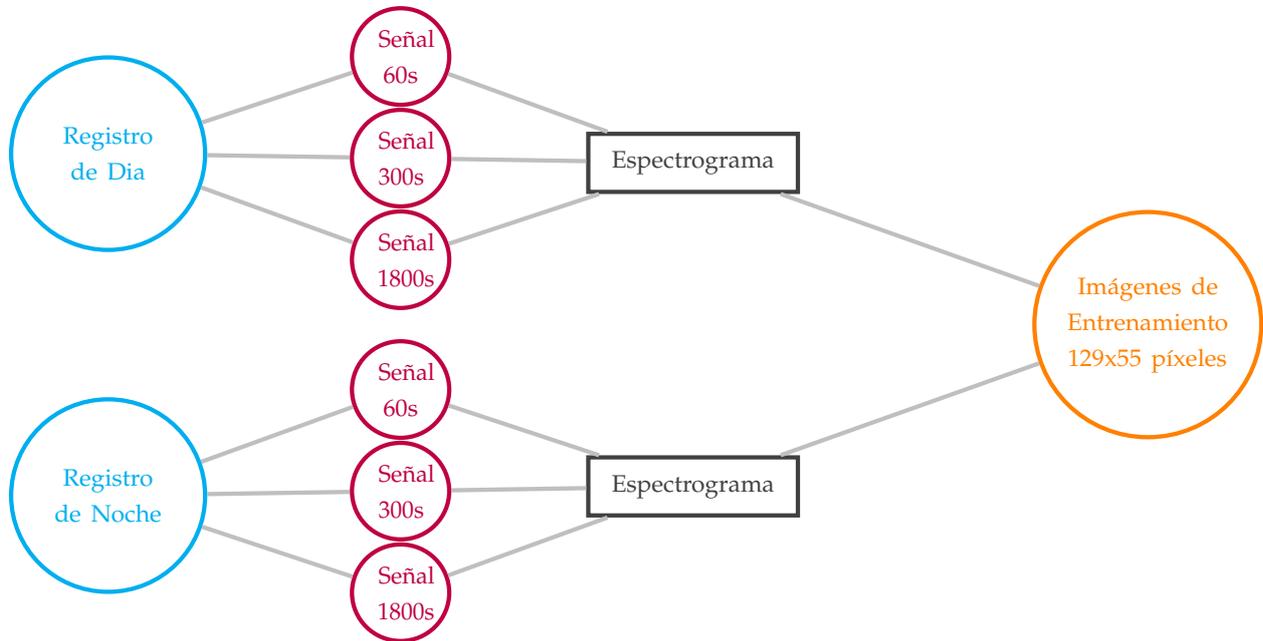


FIGURA 3.1: Esquema que resume el tratamiento de los datos para la generación de las imágenes que nos servirán de entrada para la CNN.

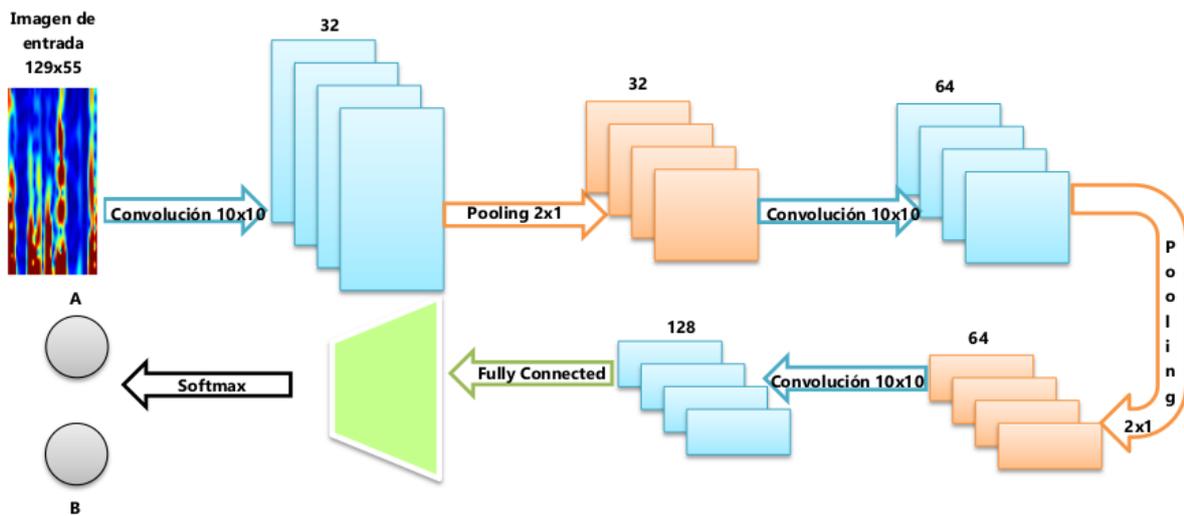


FIGURA 3.2: Estructura de la CNN utilizada para los entrenamientos

Para los entrenamientos, se empleó el **Stochastic Gradient Descent with Momentum (SGDM)**, método iterativo que consiste en minimizar la función de error buscando en la dirección contraria al gradiente de dicha función [56, 57]:

$$\theta_{l+1} = \theta_l - \alpha \nabla E(\theta_l) + \gamma(\theta_l - \theta_{l-1}), \quad (3.1)$$

donde l es el número de iteración, α es la tasa de aprendizaje, γ es el momento, θ el vector de parámetros y E la función de error.

Los parámetros configurables elegidos para los entrenamientos se resumen en la tabla 3.2.

Learning Rate	Momentum Coefficient	Minibatch Size	Epoc Number
0.01	0.9	32	75

TABLA 3.2: Parámetros utilizados en los entrenamientos

3.1.3 CARACTERÍSTICAS DE LOS ENTRENAMIENTOS CLASIFICACIÓN SECTORIZADA

En los entrenamientos por grupos sectorizados, se ha utilizado la misma metodología de entrenamiento que se planteaba en [5] y que hemos resumido en la sección 3.1.2, repitiendo el proceso para cada grupo. Podemos agrupar los entrenamientos en tres tipos:

- **Periodos de actividad**, realizaremos entrenamientos independientes para las actividades nocturna y la diurna.
- **Ventanas de actividad**, se realizarán entrenamientos independientes para las ventanas de corta, media y larga duración.
- **Grupos sectorizados**, entrenaremos por separado con los cuatro grupos definidos en la sección 3.1.1.

Realizaremos un total de 24 entrenamientos, con 10 folds de prueba en cada uno. La cantidad de imágenes de las que dispondremos en cada entrenamiento podemos verla en la tabla 3.3.

Uno de los mayores problemas en el entrenamiento de redes convolucionales es el sobreentrenamiento para alguna de las clases, por lo que lo ideal es que el número de imágenes con las que entrenamos esté balanceado para ambas. En general, podemos ver en la tabla 3.4 que nuestras clases están bastante balanceadas, con alguna excepción como se ve en los entrenamientos de actividad nocturna para las ventanas de 60s en el grupo de los chicos y de primaria.

Para realizar los entrenamientos se ha utilizado validación cruzada (2.2.1.1), por lo que hemos realizado una división por pacientes de cada grupo en dos conjuntos, de forma que el 70 % de los pacientes se ha utilizado como conjunto de entrenamiento y ha servido para realizar el aprendizaje. El 30 % restante se ha utilizado para definir el conjunto de test.

En la figura 3.3 se puede ver el esquema completo del entrenamiento.

Ventanas de 1800				
GRUPO DE ESTUDIO	DÍA		NOCHE	
	<i>Caso</i>	<i>Control</i>	<i>Caso</i>	<i>Control</i>
<i>Chicas</i>	1311	1316	967	934
<i>Chicos</i>	5252	5791	3748	4199
<i>Secundaria</i>	996	779	691	566
<i>Primaria</i>	5567	6328	4024	4567
Ventanas de 300				
GRUPO DE ESTUDIO	DÍA		NOCHE	
	<i>Caso</i>	<i>Control</i>	<i>Caso</i>	<i>Control</i>
<i>Chicas</i>	8093	7957	4392	4370
<i>Chicos</i>	32489	35478	17256	19213
<i>Secundaria</i>	6208	4843	3157	2406
<i>Primaria</i>	34374	38592	18491	21177
Ventanas de 60				
GRUPO DE ESTUDIO	DÍA		NOCHE	
	<i>Caso</i>	<i>Control</i>	<i>Caso</i>	<i>Control</i>
<i>Chicas</i>	40322	39641	8077	10121
<i>Chicos</i>	162053	178112	33773	46308
<i>Secundaria</i>	30864	24347	6781	5534
<i>Primaria</i>	171511	193406	35069	50895

TABLA 3.3: Imágenes disponibles para los entrenamientos en cada uno de los grupos

3.1.4 CLASIFICACIÓN FINAL DE LOS PACIENTES

La CNN nos permite establecer una clasificación de las ventanas de actividad, por lo que será necesario implementar algoritmos que nos permitan realizar la clasificación a nivel de pacientes. Para ello, hemos seguido la metodología descrita en [5]; la figura 3.4 muestra un esquema de cómo se realiza este procesado. De forma breve, introducimos a un clasificador dos variables, a saber el porcentaje de ventanas de actividad diurna que han resultado ser caso, y ese porcentaje para ventanas de actividad nocturna.

Con los modelos de los clasificadores entrenados se hacen predicciones sobre el mismo conjunto de test sobre el que se evaluaba la CNN. El número total de pacientes pertenecientes a los conjuntos de test y train de pacientes puede verse en la tabla 3.5.

CHICAS				
Ventana	Casos Dia	Controles Dia	Casos Noche	Controles Noche
60 s	50.43 %	49.57 %	44.38 %	55.62 %
300 s	50,42 %	49.58 %	50.13 %	49.87 %
1800 s	49.90 %	50.10 %	50.87 %	49.13 %
CHICOS				
Ventana	Casos Dia	Controles Dia	Casos Noche	Controles Noche
60 s	47.64 %	52.36 %	42.17 %	57.83 %
300 s	47.80 %	52.20 %	47.32 %	52.68 %
1800 s	47.56 %	52.44 %	47.16 %	52.84 %
SECUNDARIA				
Ventana	Casos Dia	Controles Dia	Casos Noche	Controles Noche
60 s	50,43 %	44.10 %	55.06 %	44.94 %
300 s	50,43 %	43.82 %	56.75 %	43.25 %
1800 s	50,43 %	43.89 %	54.97 %	45.03 %
PRIMARIA				
Ventana	Casos Dia	Controles Dia	Casos Noche	Controles Noche
60 s	47 %	53 %	40.79 %	59.21 %
300 s	47.11 %	52.89 %	46.61 %	53.39 %
1800 s	46.80 %	53.20 %	46.84 %	53.12 %

TABLA 3.4: Porcentaje medio de imágenes empleadas en el entrenamiento

GRUPO DE ESTUDIO	TRAIN	TEST
<i>Chicas</i>	19	8
<i>Chicos</i>	80	32
<i>Secundaria</i>	14	5
<i>Primaria</i>	85	35

TABLA 3.5: Sujetos que forman parte de los conjuntos de train y test en la clasificación de pacientes

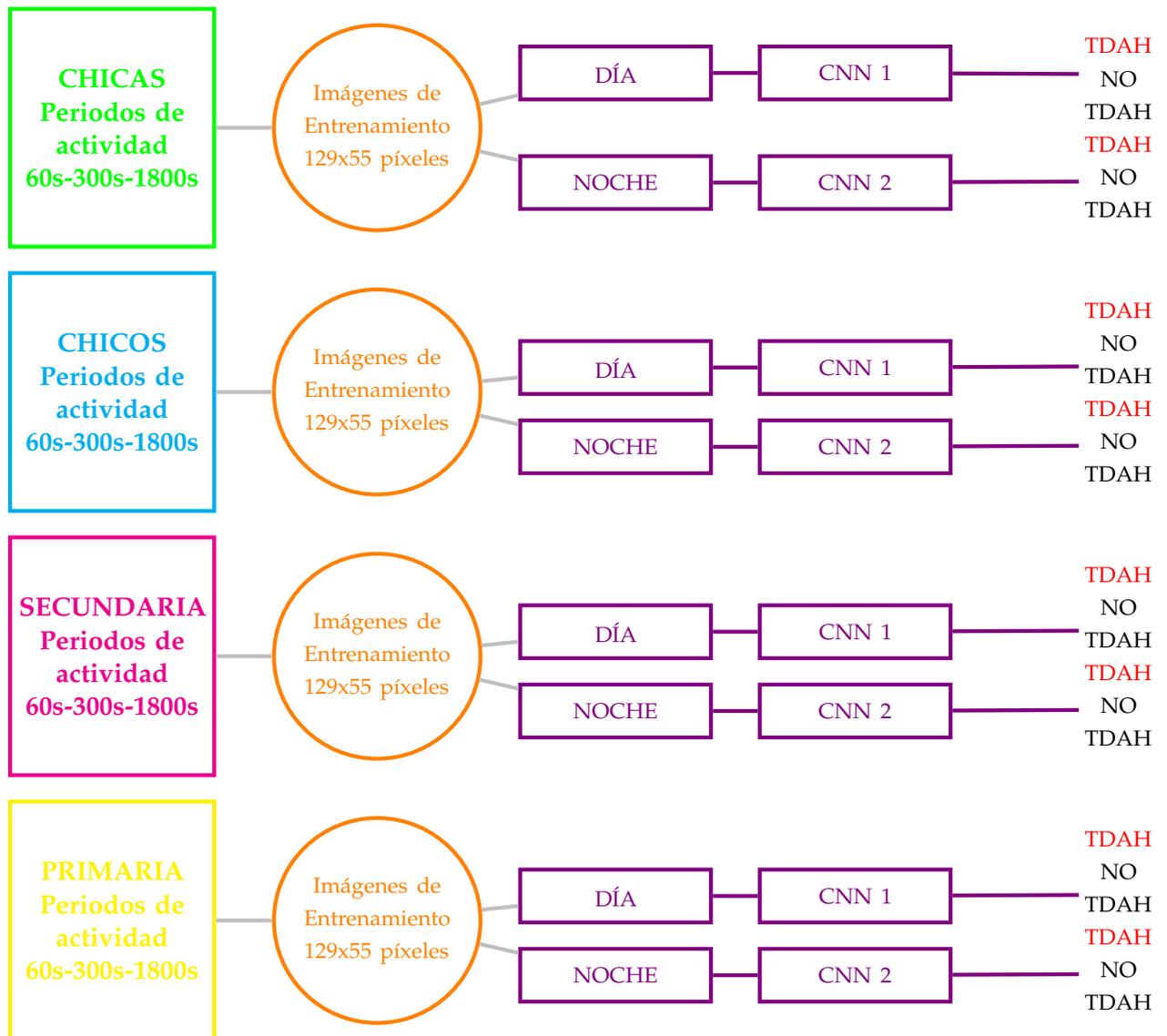


FIGURA 3.3: Proceso de entrenamiento del sistema de diagnóstico.

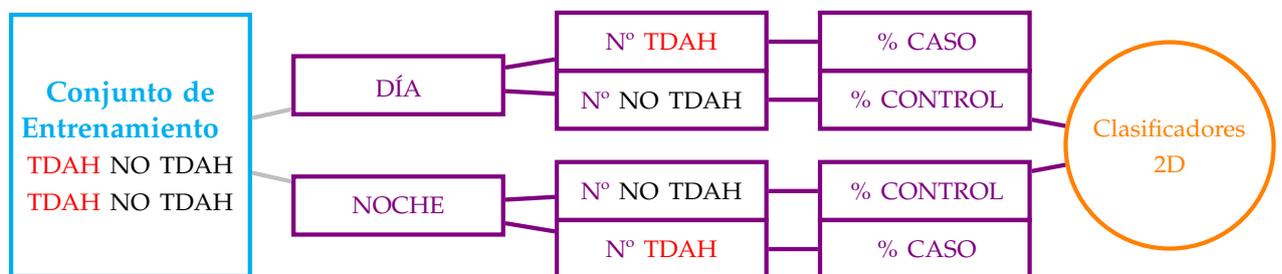


FIGURA 3.4: Proceso de entrenamiento para la clasificación por pacientes

3.2 ANÁLISIS DE LAS ACTIVACIONES DE LA CNN

Podemos dividir los métodos utilizados para la interpretación de redes convolucionales en dos tipos, en función de su principio de funcionamiento:

- Técnicas de análisis del funcionamiento de cada neurona de la CNN, como la visualización de los filtros, pesos o activaciones de las neuronas de la red [58].
- Técnicas basadas en la obtención de mapas que reflejen a partir de una imagen de entrada la importancia que tienen los píxeles en la predicción de la CNN [59].

En el presente trabajo se han estudiado varias de estas técnicas con el objetivo de diseñar mapas asociados a cada uno de los sujetos e intentar encontrar la relación entre el comportamiento de la CNN y las señales de entrada.

3.2.1 MAPAS DE ACTIVACIONES NEURONALES

En [60], Molnar plantea la interpretación de los resultados obtenidos mediante técnicas de aprendizaje automático como un problema matemático de optimización. Las redes convolucionales aprenden conceptos abstractos de las imágenes con las que alimentamos la red, especialmente en las capas ocultas. En estas redes, la imagen se va transformando a través de las capas convolucionales, aprendiendo características más complejas con cada transformación. Típicamente, las primeras capas convolucionales suelen centrarse en aprender características relacionadas con los bordes y texturas más básicas y las últimas capas suelen centrarse en la detección de partes de objetos de la imagen. Para visualizar las características, se suele emplear una imagen de entrada que maximice la activación de la red y se ven las zonas de activación para cada neurona. Debido a que la cantidad de neuronas que posee una red convolucional es muy grande, se suelen visualizar las características mediante los canales de cada capa, obteniendo un mapa de activación por cada una de las capas.

Nuestro objetivo será crear mapas propios de activación que nos permitan extraer características propias asociadas a cada uno de los grupos estudiados. Para ello se han realizado los siguientes pasos:

- Extracción de las imágenes que maximicen la activación de la red.
- Obtención de la activación de las neuronas de la CNN.
- Creación de mapas asociados a cada sujeto de caso o control y aplicación de técnicas *PCA* para extraer conclusiones de dichos mapas.

3.2.1.1 PROCESADO PREVIO AL CÁLCULO DE LAS ACTIVACIONES DE LA CNN

Para poder extraer las activaciones de las neuronas, es importante que la imagen de entrada maximice la activación de la red para una de las clases sobre las que se esté evaluando. En nuestro caso, estas imágenes se corresponderán con las imágenes que tras el entrenamiento de la red neuronal son clasificadas correctamente cuando se evalúa esta imagen en el conjunto de test.

En nuestro caso concreto, hemos seleccionado las imágenes cuya tasa de acierto se corresponde con más de un 90 %. En la tabla 3.6 podemos ver cuántas imágenes de entrada maximizan la red para cada conjunto.

Ventanas de 1800				
GRUPO DE ESTUDIO	DÍA		NOCHE	
	<i>Caso</i>	<i>Control</i>	<i>Caso</i>	<i>Control</i>
<i>Chicas</i>	298	456	267	435
<i>Chicos</i>	1546	2399	1513	2315
<i>Secundaria</i>	395	254	219	132
<i>Primaria</i>	1326	2398	1421	2553
Ventanas de 300				
GRUPO DE ESTUDIO	DÍA		NOCHE	
	<i>Caso</i>	<i>Control</i>	<i>Caso</i>	<i>Control</i>
<i>Chicas</i>	1893	2895	956	1104
<i>Chicos</i>	8325	12838	4579	7664
<i>Secundaria</i>	2238	1740	882	207
<i>Primaria</i>	9089	14698	5666	8258
Ventanas de 60				
GRUPO DE ESTUDIO	DÍA		NOCHE	
	<i>Caso</i>	<i>Control</i>	<i>Caso</i>	<i>Control</i>
<i>Chicas</i>	1778	3577	687	2213
<i>Chicos</i>	107682	125191	6187	18450
<i>Secundaria</i>	5764	3112	1618	402
<i>Primaria</i>	114558	135427	8179	22687

TABLA 3.6: Imágenes que maximizan la activación disponibles para los entrenamientos en cada uno de los grupos

Con estos conjuntos de imágenes, calcularemos las activaciones de la red para la creación de los mapas de activación.

3.2.1.2 MAPAS DE ACTIVACIONES A PARTIR DE LA EXCITACIÓN DE LAS NEURONAS

En una red convolucional, cada una de las capas está formada por neuronas que, en función de la imagen con la que alimentemos la entrada, reaccionarán y permitirán la clasificación final de dicha imagen en uno de los grupos sobre los que se realice el estudio. Si se encuentra que determinadas neuronas reaccionan siempre para un grupo de estudio concreto (Caso o Control en nuestro problema), nos será posible establecer un patrón de activación que supondrá el primer paso a la hora de elaborar un sistema que nos permita dar sentido al comportamiento de la red neuronal.

Utilizaremos los conjuntos de imágenes definidos en 3.2.1.1 para visualizar las activaciones de las neuronas de la red.

En función de la capa en la que centremos nuestro análisis la red detectará patrones diferentes, desde texturas y bordes en las primeras capas hasta partes de objetos en las más profundas [60].

En nuestro caso, las imágenes con las que alimentamos la red no son imágenes al uso, sino una representación de las señales de actimetría en tiempo-frecuencia, por lo que nos interesa que los patrones que obtengamos nos muestren activaciones en las bajas, medias o altas frecuencias. Es por ello que, vamos a analizar las activaciones de la última capa convolucional de nuestra red, antes de la capa Relu y las de clasificación, puesto que será más probable encontrar patrones más complejos en esta capa que en las anteriores.

Con las activaciones de la capa seleccionada, elaboraremos un mapa por cada uno de los sujetos de estudio, mediante el siguiente procesado:

- Se calculan las activaciones que generan cada una de las imágenes de máxima activación asociadas a cada sujeto. Obtenemos entonces para cada imagen un mapa de activaciones; donde cada uno de los rectángulos representa una neurona y su color la intensidad con la que se activa. Cada mapa estará compuesto por las activaciones de las 128 neuronas de la capa convolucional, como se ilustra en la figura 3.5. Por lo tanto, los mapas serán matrices de 256x224.
- Se obtienen para cada sujeto un mapa por cada una de las imágenes de máxima activación.
- Los mapas obtenidos para cada imagen se promedian, obteniendo finalmente un mapa de activación por cada sujeto, donde cada rectángulo representará la activación media de esa neurona para el sujeto de estudio.
- Esto se realiza para la activación nocturna y diurna.

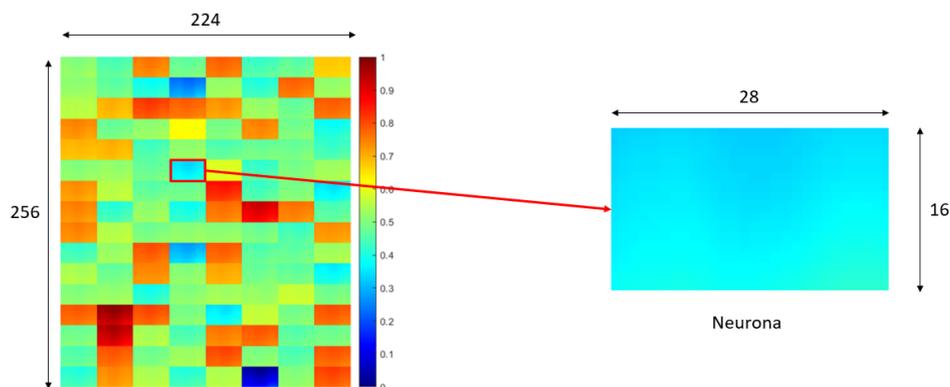


FIGURA 3.5: Mapa de Activación obtenido para un caso

Este proceso, que está esquematizado en la figura 3.7, se repite para todos los grupos de entrenamiento presentados en 3.1.3.

Una vez obtenemos estos mapas de activación, podemos empezar a realizar el procesado que nos permita dar sentido a las diferencias en los mapas. Este procesado consiste en la obtención de los autovectores de las imágenes obtenidas de los mapas mediante la técnica de

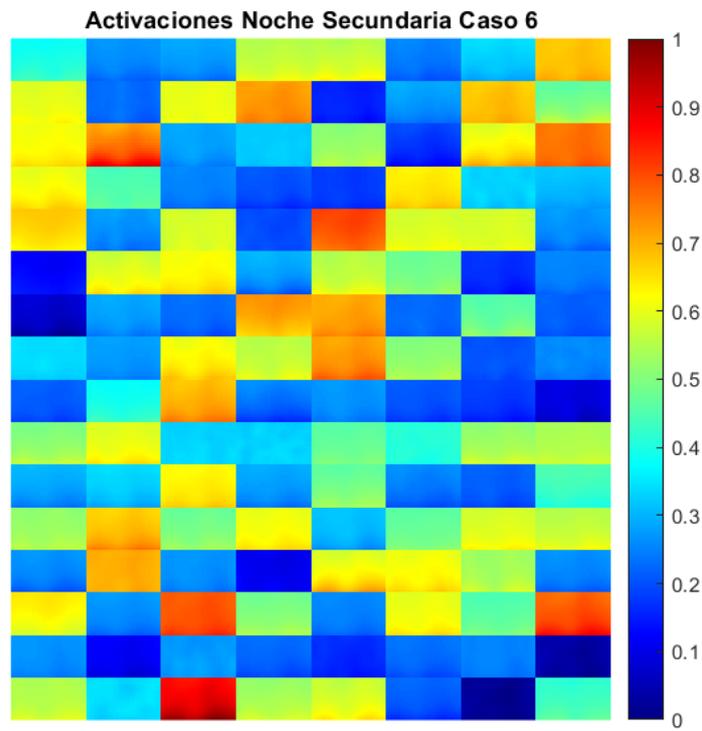


FIGURA 3.6: Mapa de Activación obtenido para un caso

eigenfaces, comúnmente utilizada para la obtención de patrones en redes de reconocimiento facial. Finalmente, tendremos unos mapas de activación caracterizados por los autovectores de las activaciones promedio de la CNN, como los que vemos en la figura 3.8.

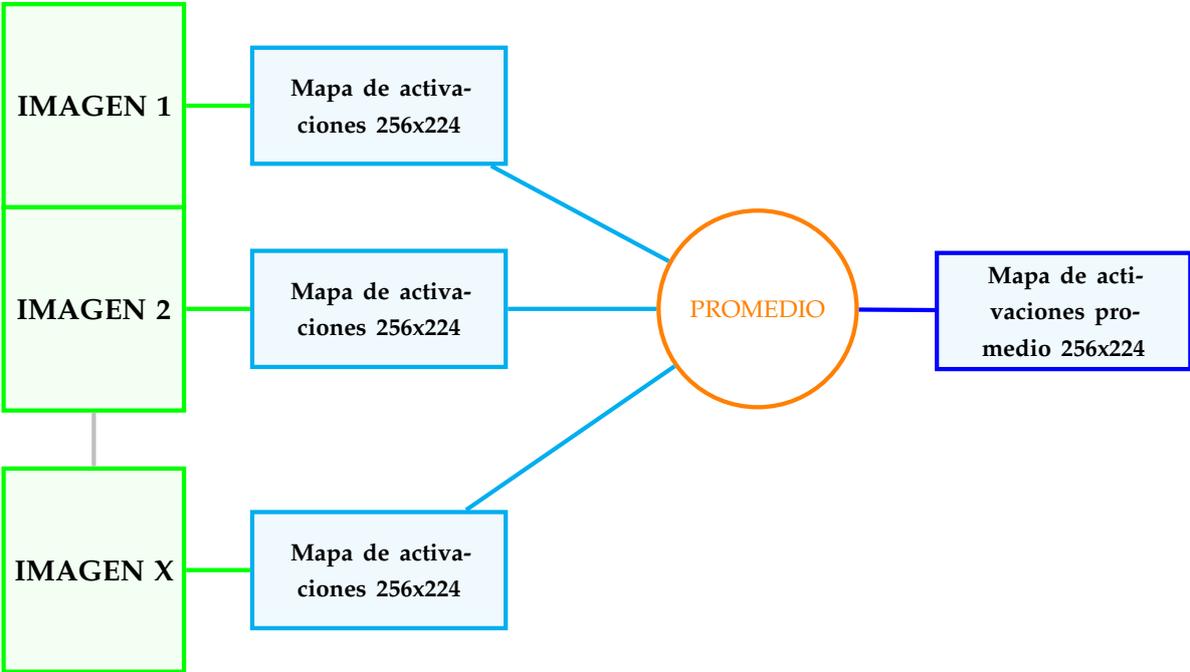


FIGURA 3.7: Esquema del procesado de los diferentes mapas de activación

3.2.1.3 EIGENFACES: CÁLCULO DE AUTOVECTORES PARA EL RECONOCIMIENTO DE PATRONES

Se entiende por **autovectores** de una matriz a aquellos vectores distintos de cero que para cierto escalar λ cumplen;

$$A\mathbf{x} = \lambda\mathbf{x} \quad (3.2)$$

El término *Eigenfaces* o caras propias hace referencia al conjunto de vectores propios utilizado principalmente en problemas de reconocimiento facial. Los autovectores se derivan de la matriz de covarianza de la distribución de probabilidad sobre el espacio vectorial de alta dimensión de las imágenes de entrada. Los *Eigenfaces* permiten crear un conjunto básico de todas las imágenes utilizadas para construir la matriz de covarianza, lo que lleva a una reducción de la dimensionalidad del problema, ya que con un conjunto más pequeño caracterizamos el conjunto original de imágenes de entrenamiento [61].

Para calcular los *Eigenfaces* hemos aplicado la siguiente metodología:

- Primero, se calcula el mapa de activación de cada sujeto del grupo de estudio según el procesado explicado en 3.2.1.2.
- Se calcula la media de los mapas de activación del grupo de estudio y se resta a cada uno de los mapas.
- Vectorizamos estos nuevos mapas, redimensionando para obtener un vector para cada uno de los sujetos de estudio.
- Calculamos la matriz de covarianza y sus autovectores y autovalores, donde cada autovector tendrá la misma dimensionalidad que el mapa de activaciones.
- Por último, se eligen las componentes principales y se genera un nuevo mapa que caracteriza a cada sujeto a partir de sus autovectores. Podemos ver en la figura 3.8 un ejemplo de cómo son estos nuevos mapas.

3.2.2 DEEPDREAM, EL ALGORITMO DE GOOGLE

DeepDream es un algoritmo creado por Google que consiste en la visualización de los patrones aprendidos por una red neuronal. Para ello, envía una imagen a través de la red, calcula el gradiente de la imagen con respecto a las activaciones de una capa en particular y devuelve una imagen que supondría la activación máxima de la red para las clases estudiadas [62].

En nuestro caso concreto, hemos utilizado el algoritmo *DeepDream* de *Matlab* para visualizar cuál sería la activación máxima de la red para casos y controles. Hemos aplicado este algoritmo de dos formas diferentes:

- Directo sobre la capa de clasificación de la CNN, para obtener imágenes que reflejen cómo sería la activación máxima de la red.
- Aplicando el algoritmo sobre las imágenes de entrada que utilizábamos para los mapas de activaciones de neuronas y observando los patrones que se generan a partir de ellas.

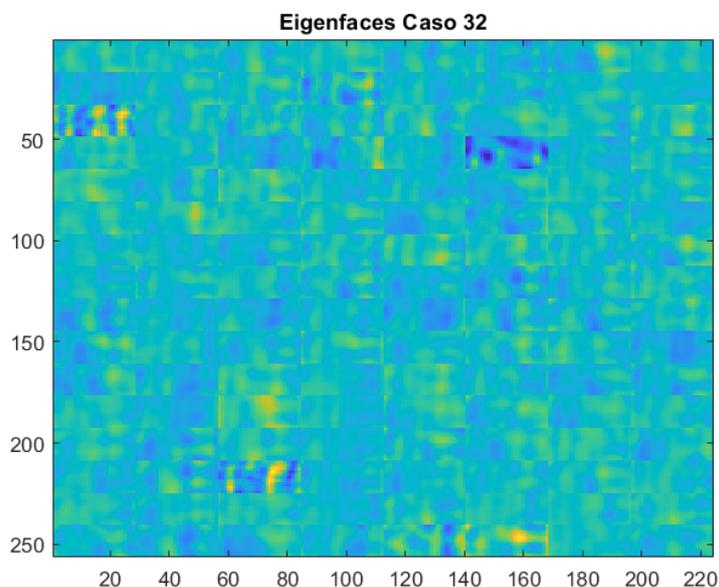


FIGURA 3.8: Eigenfaces obtenidas para un caso

3.2.3 ALGORITMOS PARA LA VISUALIZACIÓN DE ZONAS DE ACTIVACIÓN

A lo largo de los años y con el auge del Deep Learning, se han desarrollado diferentes técnicas que permiten visualizar la importancia que tiene cada píxel de la imagen de entrada en las predicciones realizadas por la red, con el objetivo de conseguir realizar una interpretación del funcionamiento de la red. Entre las numerosas técnicas existentes, en el presente trabajo nos hemos centrado en las siguientes:

- **Métodos basados en *Class Activation Mapping (CAM)***, que son técnicas que evalúan la zona de activación en función de la activación de la capa Softmax mediante la obtención la agrupación promedio global (GAP) de la CNN [63]. En nuestro caso, hemos empleado el método **Grad-CAM**, que es una generalización del método CAM, pero aplicando el cálculo del gradiente para calcular la agrupación promedio.
- **Métodos basados en perturbaciones**, cuyo principio de funcionamiento consiste en someter a perturbaciones a las características de una imagen fija, alimentar el modelo con esta imagen modificada y evaluar cómo cambia el resultado. La técnica de este tipo que hemos implementado en este trabajo se conoce como técnica de oclusión y permite obtener imágenes modificadas empleando una máscara que permite ocultar ciertas zonas de la imagen original.

Los mapas obtenidos mediante estas técnicas son de las dimensiones de las imágenes de entrada, 129x55. Se ha caracterizado a los casos y controles de cada uno de los grupos mediante sus mapas de oclusión y Grad-CAM promedio con un proceso análogo al que se describe en el esquema de la figura 3.7. En los siguientes apartados se explicará el principio de funcionamiento de estos métodos de visualización.

3.2.3.1 MAPAS DE ACTIVACIÓN GRAD-CAM

La generación de mapas de activación de clases (CAM) diseñado por Zhou et al.. [64] es un método que permite visualizar aquellas regiones de la imagen que utiliza la CNN para distinguir entre las diferentes categorías utilizando la agrupación promedio global (GAP) en las CNN. El principio de funcionamiento de esta técnica consiste en dos procedimientos:

- Se realiza un agrupamiento de promedios globales en los mapas de características convolucionales.
- Después, utilizamos estos promedios como características de entrada para una capa completamente conectada que nos devolverá el resultado para una de las clases.

Una vez que extraemos los pesos de la capa completamente conectada, si lo proyectamos sobre la imagen de entrada, podremos visualizar las regiones más importantes para la clasificación. En la figura 3.9 se detalla este proceso. Típicamente se aplica sobre la capa softmax:

$$GAP = \sum_{x,y} f_k(x,y) \quad (3.3)$$

$$S_c = \sum_k \sum_{x,y} \omega_k^c f_k(x,y) \quad (3.4)$$

$$M_c(x,y) = \sum_k \omega_k^c f_k(x,y) \quad (3.5)$$

donde S_c es la entrada al Softmax para una clase c dada, ω_k^c es el peso de la clase c en la unidad k , $f_k(x,y)$ es la activación de la unidad k en la última capa convolucional en la posición (x,y) y $M_c(x,y)$ sería el mapa de activación buscado.

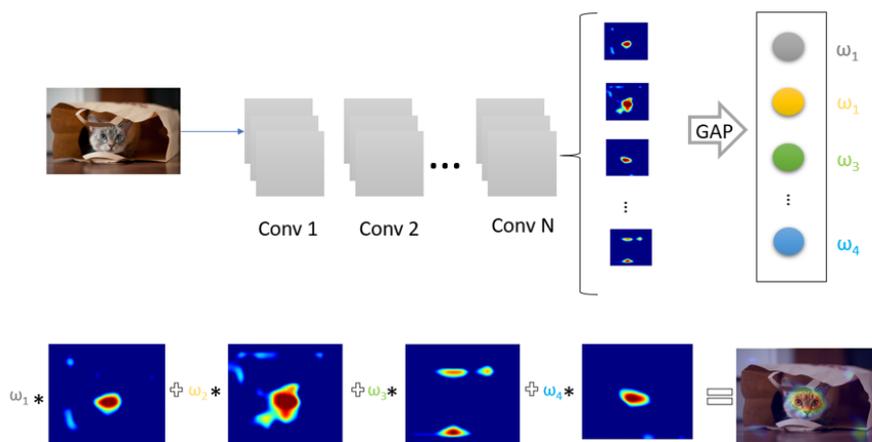


FIGURA 3.9: Algoritmo para la obtención de los Mapas de Activación de clases

3.2.3.2 MAPAS DE OCLUSIÓN

El principal problema de este algoritmo es que no es posible aplicarlo en redes convolucionales que implementen capas no convolucionales para la extracción de características. Selvaraju et al. [65] definen una alternativa más flexible basada en el cálculo del gradiente y que es la que hemos implementado en el presente trabajo.

El método Grad-CAM consiste en una generalización de CAM que emplea la agrupación promedio de los gradientes de las probabilidades de la clase de estudio:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (3.6)$$

donde y^c es la salida de la capa SoftMax para la clase de estudio, A_{ij}^k es el valor del mapa de activación para cada neurona de la última capa convolucional (k) en la posición (i,j) y α_k^c es el vector de pesos cuya longitud es igual al número de neuronas de la última capa convolucional.

Para obtener los mapas de activación, se realiza una combinación lineal ponderada de los pesos; que se obtienen de la misma forma que en el método general y se pasa por una ReLu para eliminar los valores que influyen negativamente en la predicción.

$$L_{grad-CAM}^c = ReLu \left(\sum_k \alpha_k^c * A^k \right) \quad (3.7)$$

Un ejemplo de los mapas promedio que hemos obtenido podemos verlo en la figura 3.10.

3.2.3.3 MAPAS DE OCLUSIÓN

La técnica de oclusión consiste en ocultar diversas zonas de la imagen de entrada de manera sistemática y obtener a partir de dichas zonas información acerca de si la red está identificando el objeto o lo está clasificando por su contexto [66]. El procesado que se realiza en esta técnica puede resumirse en los siguientes pasos:

- Se define una máscara, de tamaño fijo o variable, la cual se desplazará o largo de la imagen, generando en cada iteración una imagen modificada a partir de la entrada.
- Esta imagen se utiliza como entrada de la red, y se evalúa la predicción sobre la clase de estudio, generando una matriz que almacenará diversas puntuaciones en relación a la importancia que tenga el área oculta para la clasificación.
- Una vez calculada la matriz, se redimensiona al tamaño de la imagen original, para obtener así el mapa de oclusión.

En la figura 3.11 podemos ver un gráfico del funcionamiento de dicho algoritmo.

En la figura 3.12 podemos visualizar un ejemplo de los mapas de oclusión promedio que hemos calculado siguiendo el proceso análogo al descrito en 3.7.

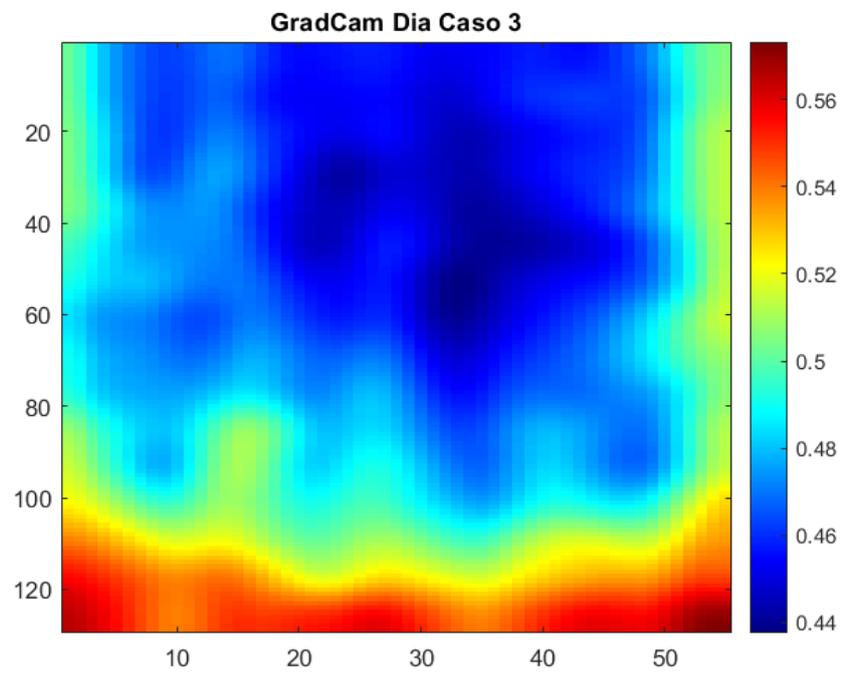


FIGURA 3.10: Mapa Grad-CAM obtenido para un caso

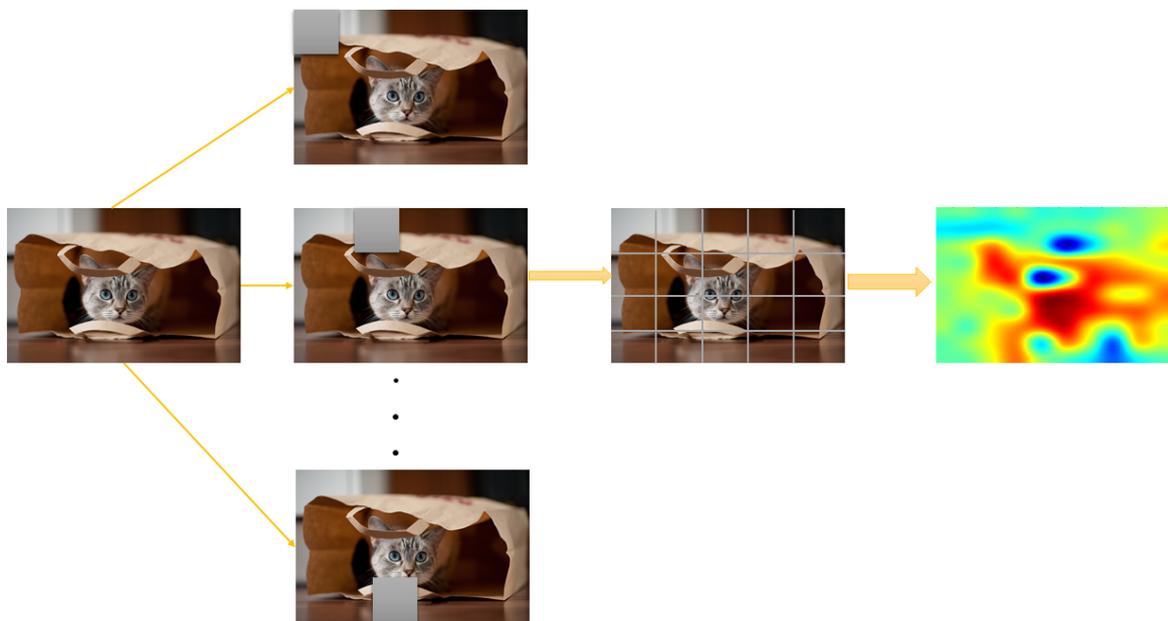


FIGURA 3.11: Esquema de funcionamiento de la técnica de oclusión

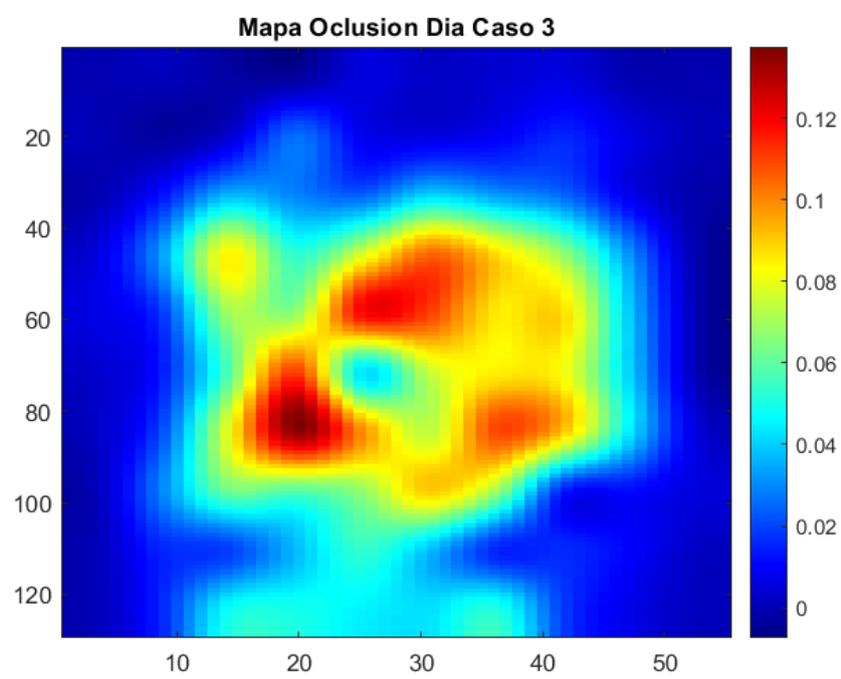


FIGURA 3.12: Mapa de Oclusión obtenido para un caso

RESULTADOS Y DISCUSIÓN

En esta sección se discutirán los resultados obtenidos tanto en la clasificación sectorizada como para la interpretabilidad. Para ello:

- Mostraremos los resultados obtenidos para los cuatro grupos sobre los que se ha realizado el estudio en términos de clasificación de ventanas y de pacientes. Nos centraremos en mostrar un análisis para las métricas habituales.
- Se analizarán algunos de los resultados obtenidos mediante las técnicas presentadas en el capítulo 3 para la interpretación de los resultados extraídos de la CNN utilizada en este estudio.

4.1 PRESTACIONES DE LA RED CONVOLUCIONAL PARA LA CLASIFICACIÓN SECTORIZADA

Para analizar la eficiencia de la red en la clasificación sectorizada, se definieron en el capítulo 2.2 diferentes parámetros que permiten evaluar el comportamiento de nuestro sistema de clasificación:

- **Exactitud, Sensibilidad y Especificidad:** Nos permiten medir el porcentaje de acierto global del clasificador y la probabilidad de que el sistema clasifique correctamente un paciente diagnosticado con la patología o sano respectivamente.
- **Valores Predictivos:** Permiten medir la eficacia real de una prueba diagnóstica, ya que nos proporcionan la probabilidad de padecer o no la enfermedad conociendo los resultados de la prueba diagnóstica.
- **Razón de Verosimilitud (LR):** Se refiere a la razón entre la posibilidad de observar un resultado en los pacientes diagnosticados con la enfermedad en cuestión frente a la posibilidad de ese resultado en pacientes sin la patología. Este parámetro es de gran utilidad en términos clínicos.

En la tabla 4.1 podemos ver los resultados obtenidos para los cuatro grupos de estudio según la longitud de ventana.

Secundaria		
Longitud de ventana	Exactitud Dia	Exactitud Noche
60	0.5913	0.5323
300	0.62151	0.56735
1800	0.70963	0.63854
Primaria		
Longitud de ventana	Exactitud Dia	Exactitud Noche
60	0.5838	0.6041
300	0.6016	0.6831
1800	0.6045	0.7183
Chicas		
Longitud de ventana	Exactitud Dia	Exactitud Noche
60	0.5493	0.6031
300	0.5698	0.6296
1800	0.5787	0.6782
Chicos		
Longitud de ventana	Exactitud Dia	Exactitud Noche
60	0.6059	0.6747
300	0.6020	0.6782
1800	0.6089	0.7119

TABLA 4.1: Exactitud obtenida del entrenamiento de la CNN para el conjunto de test.

En cuanto a la clasificación de las ventanas, aunque no existe una diferencia sustancial en los valores obtenidos para los conjuntos sectorizados respecto a los que se obtuvieron para la clasificación conjunta [5], sí se observan diferencias. En el caso del grupo de secundaria, la red presenta una exactitud mayor en el conjunto de día respecto a la noche, al contrario de lo que ocurre con el conjunto de primaria, lo que parece indicar que la actividad diurna tiene mayor relevancia en este grupo de estudio. Otro dato a destacar es que en el caso de la clasificación por sexo, aunque ambos grupos presentan el patrón de clasificación que se obtuvo en [5], la exactitud en el grupo de chicas es significativamente más baja, lo que puede deberse al tamaño del conjunto de test.

Para poder evaluar la eficiencia real de la clasificación sectorizada, es necesario comprobar cómo se comporta el sistema diseñado en la clasificación de pacientes. Si la clasificación de pacientes es satisfactoria, podremos utilizar los resultados obtenidos de la red neuronal como base para interpretar el funcionamiento de dicha red.

4.1.1 CLASIFICACIÓN GLOBAL DE PACIENTES

La clasificación de pacientes se ha llevado a cabo a partir de los porcentajes de ventanas clasificadas como caso o control obtenidos de la clasificación de ventanas. En las tablas siguientes se muestran los resultados obtenidos para las diferentes métricas en el mejor de los clasificadores utilizados.

Conjunto	Sensibilidad	Precisión	Especificidad	VPP	NVP	LR+	LR-
1800	1	0.9	0.75	0.8750	1	4	0
300	1	0.86	0.65	0.8450	0.8	2.8571	0
60	0.9667	0.92	0.85	0.9167	0.95	6.444	0.0392

TABLA 4.2: Resultados obtenidos para el conjunto de secundaria.

Conjunto	Sensibilidad	Precisión	Especificidad	VPP	NVP	LR+	LR-
1800	0.9235	0.9486	0.9722	0.9705	0.9322	32.2	0.0787
300	0.9471	0.96	0.9722	0.9721	0.9541	39.0941	0.0545
60	0.9529	0.9743	0.9944	0.9941	0.9581	171.529	0.0473

TABLA 4.3: Resultados obtenidos para el conjunto de primaria.

Conjunto	Sensibilidad	Precisión	Especificidad	VPP	NVP	LR+	LR-
1800	1	0.9571	0.90	0.9705	0.9467	10	0
300	1	0.9429	0.8667	0.9267	1	7.5	0
60	0.85	0.9143	1	1	0.8850		0.15

TABLA 4.4: Resultados obtenidos para el conjunto de chicas.

Conjunto	Sensibilidad	Precisión	Especificidad	VPP	NVP	LR+	LR-
1800	0.9125	0.9438	0.9750	0.9757	0.9227	36.5	0.08970
300	0.9313	0.9531	0.9750	0.9753	0.9362	37.25	0
60	0.9625	0.9688	0.9750	0.9757	0.9646	38.5	0.0385

TABLA 4.5: Resultados obtenidos para el conjunto de chicos.

Cabe destacar que en el caso de los grupos de secundaria y las chicas, el conjunto de test es tan pequeño que un error en la clasificación de un único sujeto puede significar una gran variación en el resultado. Es por esto que para estos conjuntos se ha realizado también una clasificación *Leave One Out*.

En las tablas 4.6 y 4.7 se muestran los resultados obtenidos para esta clasificación.

Conjunto	Exactitud	σ
1800	0.9789	0.0713
300	0.9789	0.035
60	0.9684	0.0820

TABLA 4.6: Exactitud para la clasificación *Leave One Out* de secundaria.

Conjunto	Exactitud	σ
1800	0.9741	0.0656
300	0.9667	0.0679
60	0.9593	0.0747

TABLA 4.7: Exactitud para la clasificación *Leave One Out* de chicas.

En los resultados obtenidos para la clasificación *Leave One Out* se puede observar que la tasa de acierto a la hora de clasificar cada sujeto individualmente es elevada, lo que indica que a pesar de estar trabajando en grupos de test pequeños, la respuesta del sistema de clasificación es buena para dichos grupos.

Los valores predictivos y las razones de verosimilitud son parámetros que tienen aplicación clínica ya que permiten hacernos una idea de la eficacia de la prueba diagnóstica. En nuestro caso, todos los grupos presentan valores predictivos elevados, por lo que podemos afirmar que nuestro sistema se comporta de manera eficiente, con una cantidad muy pequeña de falsos negativos y falsos positivos. Las razones de verosimilitud nos permiten medir la validez de una prueba diagnóstica. Para que la utilidad de dicha prueba sea elevada, los valores de LR+ y LR- deben ser superiores a 10 e inferiores a 0.1 respectivamente, como ya pusimos de manifiesto en la tabla 2.4. En nuestro caso se cumplen estas condiciones en la mayoría de los grupos, con excepción de los conjuntos de secundaria y las chicas, debido principalmente a que el conjunto de test es muy pequeño. En relación con el resto de métricas que hemos evaluado, el sistema otorga buenas prestaciones en términos de sensibilidad, exactitud y especificidad, aunque son ligeramente inferiores a las que obteníamos al trabajar con el conjunto completo. Esta ligera disminución se debe a que, al trabajar con grupos más pequeños, clasificar erróneamente perjudica más a los porcentajes globales. Para comparar de una forma más equitativa con los resultados obtenidos anteriormente, hemos recogido en la tabla 4.8 la media de fallos que se producen en cada uno de los grupos de entrenamiento y la media que obteníamos para los entrenamientos del grupo completo.

Observamos que, en general, la sectorización que hemos llevado a cabo mejora la clasificación global, ya que aunque los porcentajes de exactitud, sensibilidad y especificidad han variado lige-

Ventana	Secundaria	Primaria	Chicas	Chicos	Conjunto Global
1800 s	<1	1.8	<1	1.5	<1
300 s	<1	1.4	<1	1.3	2.5
60 s	<1	<1	<1	1	1.2

TABLA 4.8: Media de fallos que se producen en la clasificación.

ramente, el sistema falla en clasificar menos pacientes que los que fallaba al hacer la clasificación conjunta.

Para ilustrar mejor estas conclusiones, vamos a visualizar los diagramas de dispersión (*Scatter Plot*) extraídos de los conjuntos de clasificación.

4.1.1.1 SCATTER PLOT PARA LOS GRUPOS DE ESTUDIO

En las figuras 4.1 a 4.6 podemos ver los *Scatter plot* obtenidos para los diferentes tamaños de ventana.

Estos diagramas reafirman lo que observábamos en los resultados de la clasificación global , los *Scatter Plot* nos muestran un bajo número de sujetos mal clasificados, en particular cuando se trata de casos.

Con los resultados obtenidos, podemos afirmar que la CNN diseñada es un buen punto de partida para aplicar los algoritmos de interpretabilidad.

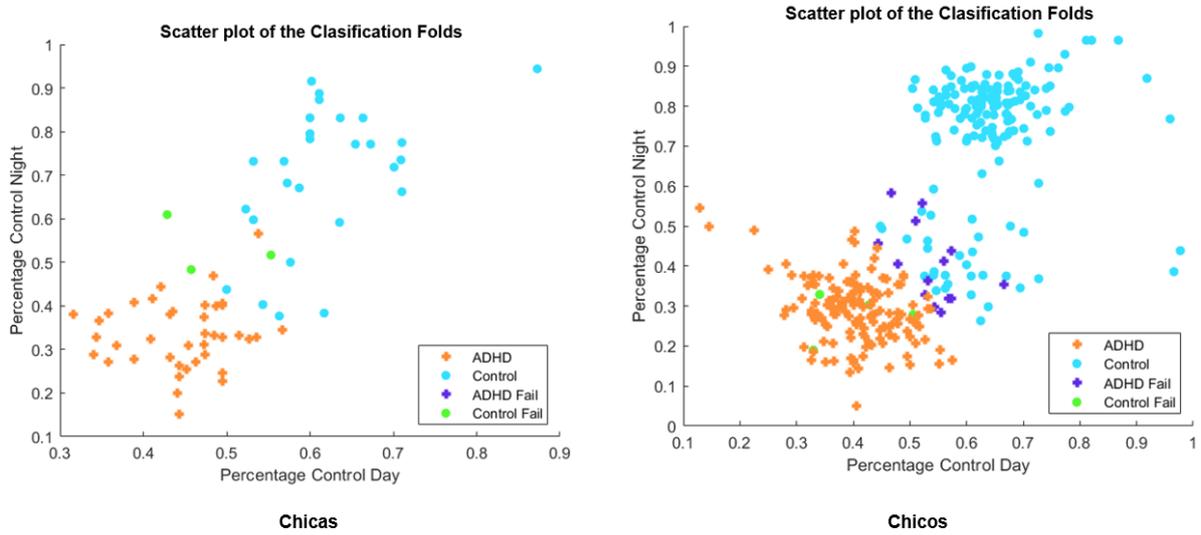


FIGURA 4.1: Scatter Plot para las ventanas de 1800s de Chicas (izquierda) y Chicos (derecha).

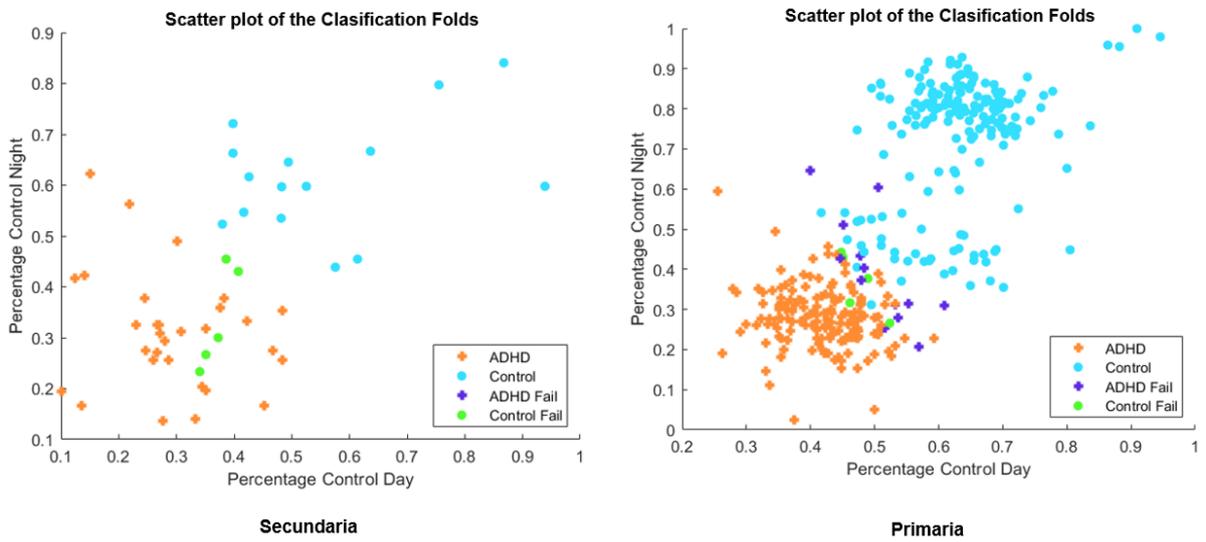


FIGURA 4.2: Scatter Plot para las ventanas de 1800s de secundaria(izquierda) y primaria(derecha).

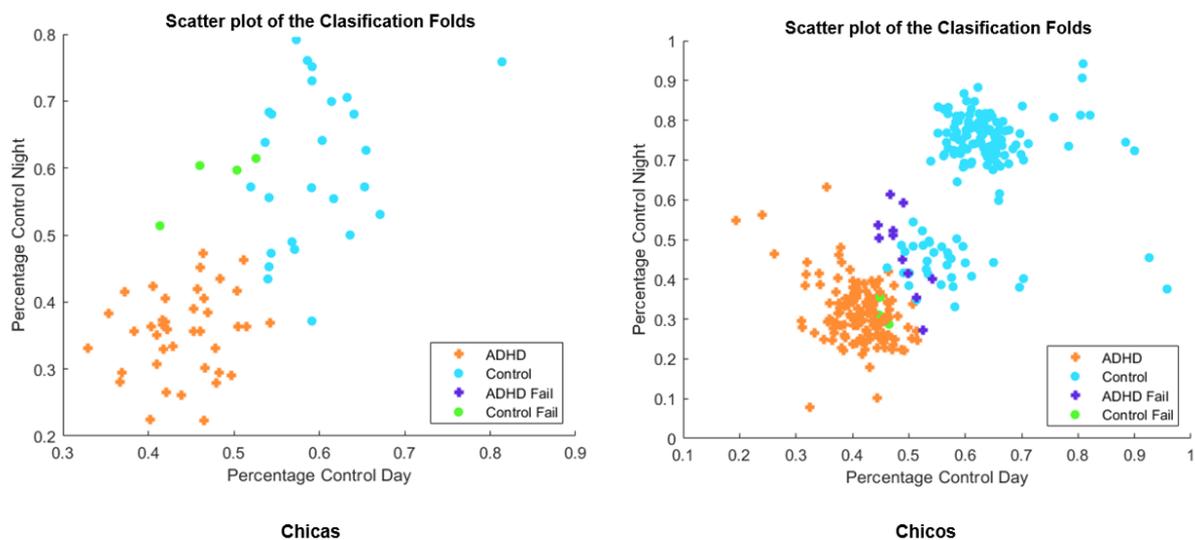


FIGURA 4.3: *Scatter Plot* para las ventanas de 300s de chicas (izquierda) y chicos (derecha).

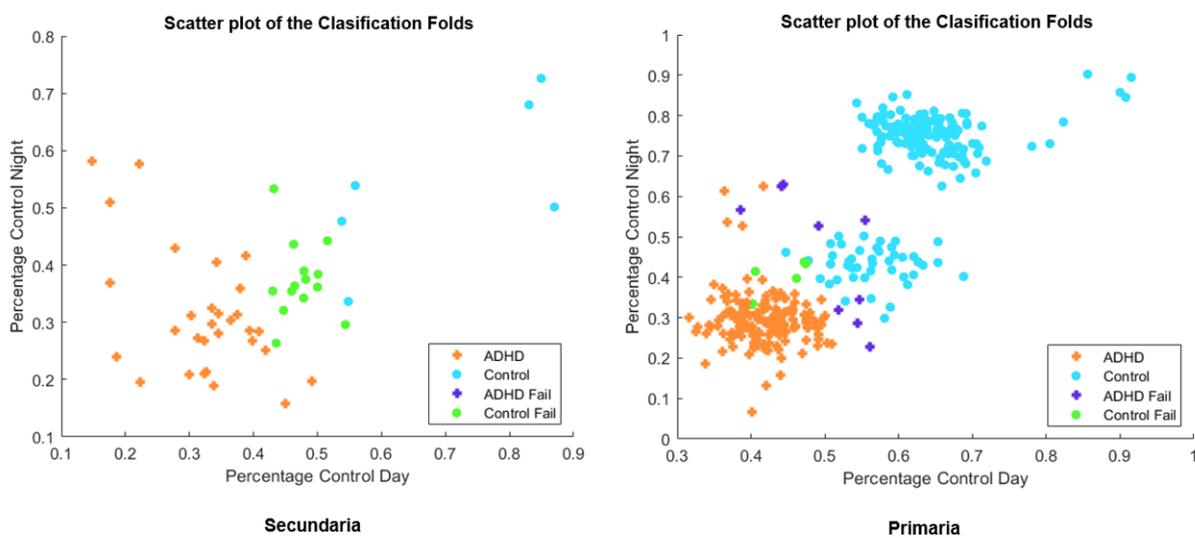


FIGURA 4.4: *Scatter Plot* para las ventanas de 300s de secundaria(izquierda) y primaria(derecha).

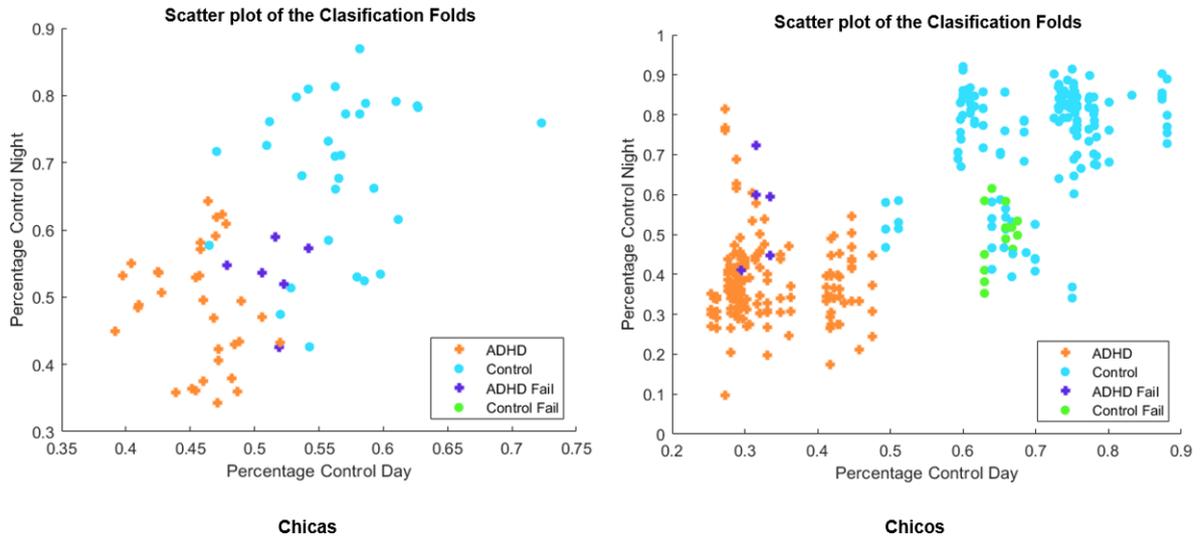


FIGURA 4.5: *Scatter Plot* para las ventanas de 60s de de chicas (izquierda) y chicos (derecha).

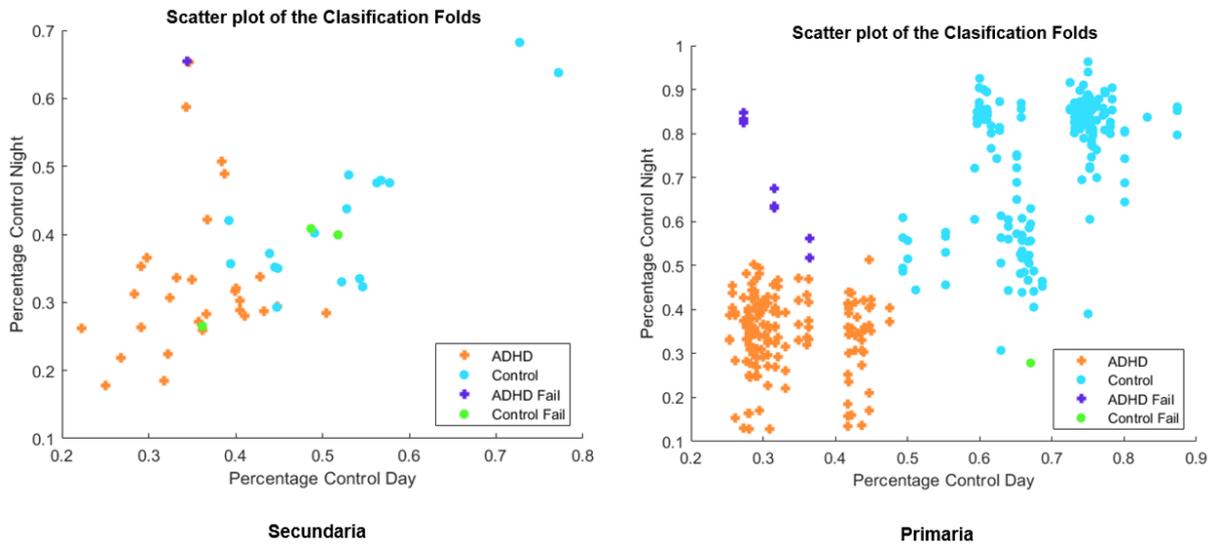


FIGURA 4.6: *Scatter Plot* para las ventanas de 60s de secundaria(izquierda) y primaria(derecha).

4.2 PRESTACIONES Y RESULTADOS OBTENIDOS PARA LOS DIFERENTES MÉTODOS DE INTERPRETABILIDAD DE LA CNN

En el capítulo 3 se explicaron varias técnicas que sirven para extraer de la red neuronal información útil a la hora de entender e interpretar su funcionamiento. En esta sección vamos a ver algunos de los resultados obtenidos tras adaptar esas técnicas a nuestro problema concreto. Presentaremos los resultados en dos grupos:

- Métodos que se centran en el funcionamiento de la CNN.
- Métodos que permiten la visualización de las zonas más relevantes de la imagen para su posterior clasificación.

Con ello, se pretende llegar a un método que nos permita entender visualmente las diferencias entre aquellos sujetos que están diagnosticados con TDAH (casos) y los sujetos de control.

4.2.1 MAPAS DE ACTIVACIONES

En la sección 3.2.1.2 se detalló el proceso de obtención de estos mapas de activaciones. Cuando se obtiene el mapa de las activaciones, lo que se obtiene es la activación que cada una de las imágenes de entrada provoca en las neuronas de la capa seleccionada. En nuestro caso, se ven las 128 neuronas pertenecientes a la tercera capa convolucional, que es la capa más profunda de la red y la que creemos que puede darnos información más relevante.

Para obtener cada uno de los mapas para cada sujeto se ha hecho primero una media de las activaciones en cada uno de los 10 folds entrenados (sección 3.1.3) y después se han medido las activaciones de neuronas que provocan cada una de las imágenes de acierto máximo que hemos obtenido para cada sujeto. Finalmente, cada sujeto tendrá asociados dos mapas de activación: uno relacionado con las redes separadas en grupos de edad y otro para los grupos separados por sexo.

Vamos a visualizar en las figuras 4.7 y 4.8 algunos de los mapas de activación obtenidos

Los mapas de activación muestran ligeras variaciones en las neuronas que se activan en función de si es caso o control, pero no se ve de una forma clara cuál puede ser el origen de dichas variaciones.

4.2.1.1 EIGENFACES SOBRE LOS MAPAS DE ACTIVACIÓN

Debido a que los mapas de activación por sí solos no proporcionan una interpretación clara, vamos a proceder a realizar sobre ellos el cálculo de los *Eigenfaces*, o conjunto de vectores propios empleados en problemas de reconocimiento facial (sección 3.2.1.3).

Adaptando esta técnica a nuestro caso particular, se han obtenido mapas de *Eigenfaces* como los que podemos ver en la figura 4.9, en la que se muestran los *Eigenfaces* de dos chicas de control y dos casos. Vemos que las imágenes características obtenidas sí que presentan diferencias pero,

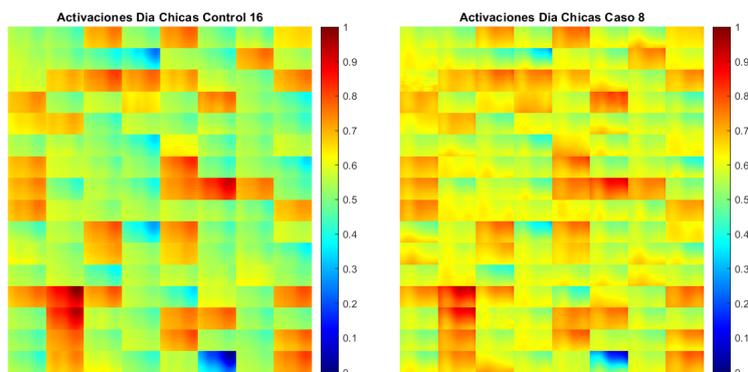


FIGURA 4.7: Mapas de activación para dos chicas de secundaria en el tramo de actividad diurna.

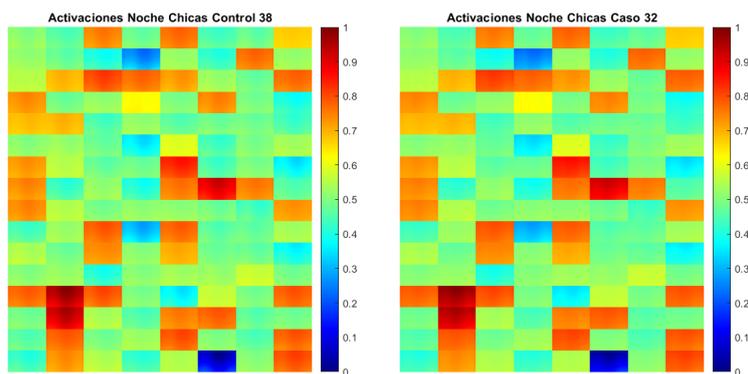


FIGURA 4.8: Mapas de activación para dos chicas de primaria en el tramo de actividad nocturna.

al igual que pasaba con los mapas de activaciones, no nos clarifican a qué se deben dichas diferencias.

El análisis a partir de los mapas de activaciones y los *Eigenfaces* no parece proporcionar una explicación clara de los pormenores de la CNN, pero sí nos permite ver ligeras diferencias en función de si el sujeto que estamos evaluando es un caso o un control. Por tanto, resultará interesante probar otros métodos alternativos.

4.2.1.2 APLICACIÓN DEL ALGORITMO DEEP DREAM PARA LOS CUATRO SUBGRUPOS DE ESTUDIO

Este algoritmo, que se explicó en el capítulo 3.2.2, nos puede servir para comprobar que en la red:

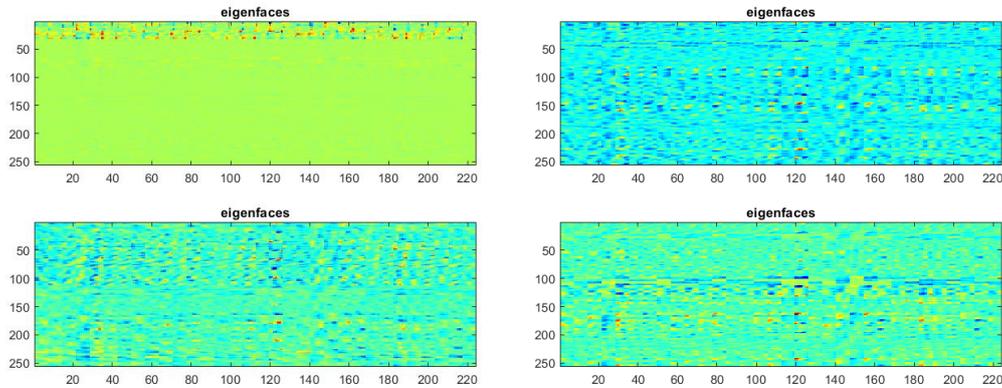


FIGURA 4.9: Mapas de Eigenfaces obtenidos para controles(izquierda) y casos(derecha) en la red de secundaria.

- Se activan diferentes zonas de la imagen en función de si es un caso o control.
- Existen diferencias sustanciales en las activaciones en función del grupo de estudio.

Las figuras 4.10 a 4.13 representan las imágenes artificiales creadas por el algoritmo que representarían imágenes idealizadas de entrada para un caso y un control. Estas imágenes nos muestran diferencias importantes, por lo que podemos entender que la red genera excitaciones diferentes de sus neuronas en función de si la imagen de entrada es un caso o control.

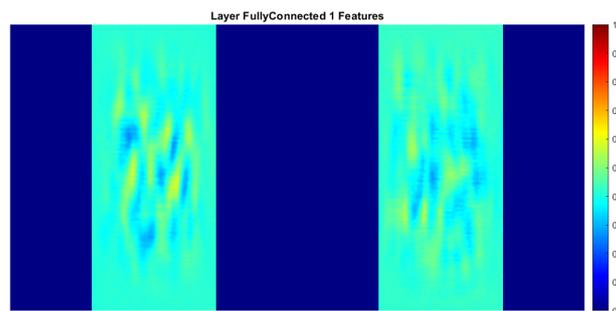


FIGURA 4.10: Imágenes de activación máxima para un caso (izquierda) y un control(derecha) generadas por el algoritmo DeepDream en la red de secundaria para la actividad nocturna.

No obstante, si intentamos aplicar este algoritmo partiendo de las imágenes de entrada de máximo acierto que definíamos en la sección 3.2.1.2 este algoritmo no nos permite apreciar dichas diferencias.

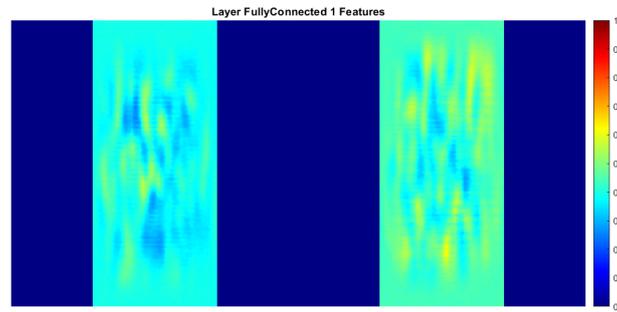


FIGURA 4.11: Imágenes de activación máxima para un caso (izquierda) y un control(derecha) generadas por el algoritmo DeepDream en la red de chicas para la actividad nocturna.

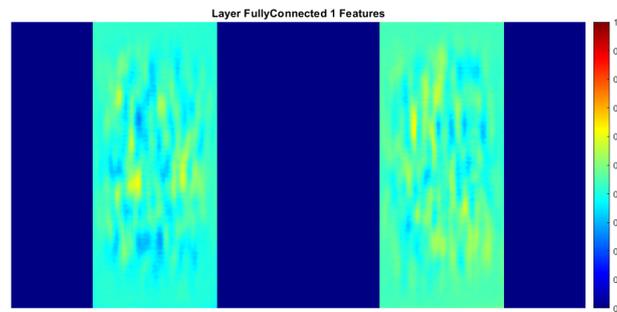


FIGURA 4.12: Imágenes de activación máxima para un caso (izquierda) y un control(derecha) generadas por el algoritmo DeepDream en la red de primaria para la actividad nocturna.

4.2.2 VISUALIZACIÓN DE LAS ZONAS MÁS RELEVANTES PARA LA CLASIFICACIÓN

En los problemas de clasificación, resulta de especial interés poder visualizar aquellas zonas de la imagen de entrada que la red utiliza para hacer la discriminación por clases. En nuestro caso, como las imágenes que utilizamos para alimentar la CNN son espectrogramas de las señales de actividad, este análisis por zonas nos indicará el rango frecuencial que la red está utilizando en la clasificación. Como se mencionó en la sección 3.2.3, en este estudio hemos empleado dos de ellos:

- **Mapas Grad-Cam:** Devuelve el mapa de activación de clase ponderado por gradiente.
- **Mapas de Oclusión:** Se obtienen midiendo el cambio en la activación total para la capa y el canal especificados cuando algunas partes de los datos de entrada están ocultos con una máscara.

En las siguientes secciones vamos a visualizar y explicar los resultados obtenidos mediante estas técnicas.

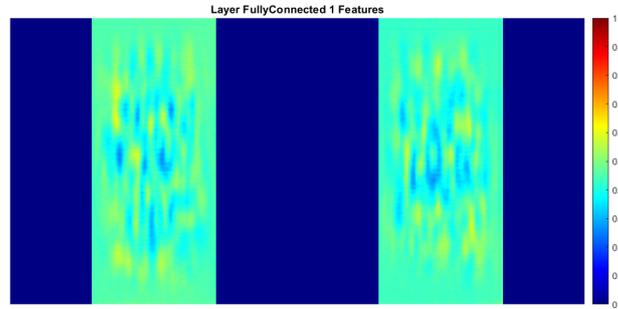


FIGURA 4.13: Imágenes de activación máxima para un caso (izquierda) y un control(derecha) generadas por el algoritmo DeepDream en la red de chicos para la actividad nocturna.

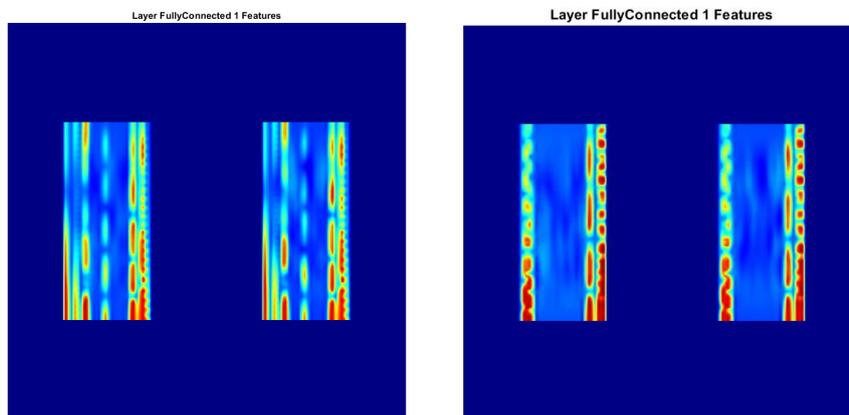


FIGURA 4.14: Imágenes generadas por el algoritmo DeepDream para nuestras imágenes de entrada para la red de chicas (izquierda) y la red de secundaria, mostrando el resultado para un caso y un control en ambas ejecuciones.

4.2.2.1 MAPAS GRAD-CAM

Los métodos de activación de clase (CAM) permiten obtener mapas de calor resultantes de la combinación lineal ponderada de las capas más profundas de la red. En los ejemplos mostrados en las figuras 4.15 y 4.16 podemos ver los mapas obtenidos para chicas con TDAH y de control en la etapa de secundaria. Aunque se aprecian diferencias entre los mapas de las dos clases planteadas, es muy complejo hallar un patrón que se repita en los mapas de activación de una de las clases.

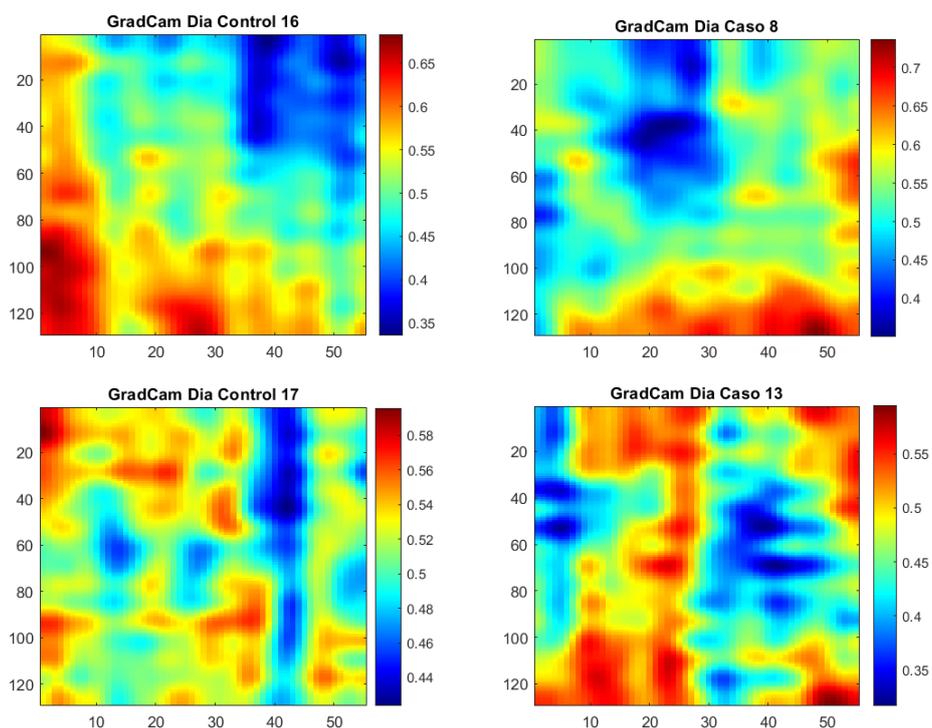


FIGURA 4.15: Mapas de Grad-CAM obtenidos para las señales de actividad diurna.

4.2.2.2 MAPAS DE OCLUSIÓN

Con los métodos basados en oclusión podemos obtener mapas de activaciones por zonas que nos ayudan comprender de forma intuitiva las partes de una imagen que son más importantes en la clasificación. En nuestro caso concreto, como las imágenes de entrada representan espectrogramas de las señales de actimetría, la visualización de dichas zonas nos indicará qué rangos de frecuencia resultan importantes en la clasificación.

En la figura 4.17 podemos ver un ejemplo de cómo son estos espectrogramas.

Para realizar el estudio, se han generado los mapas de oclusión para el caso de ventanas de tamaño grande (1800s) ya que la clasificación por ventanas responde mejor para esta franja de

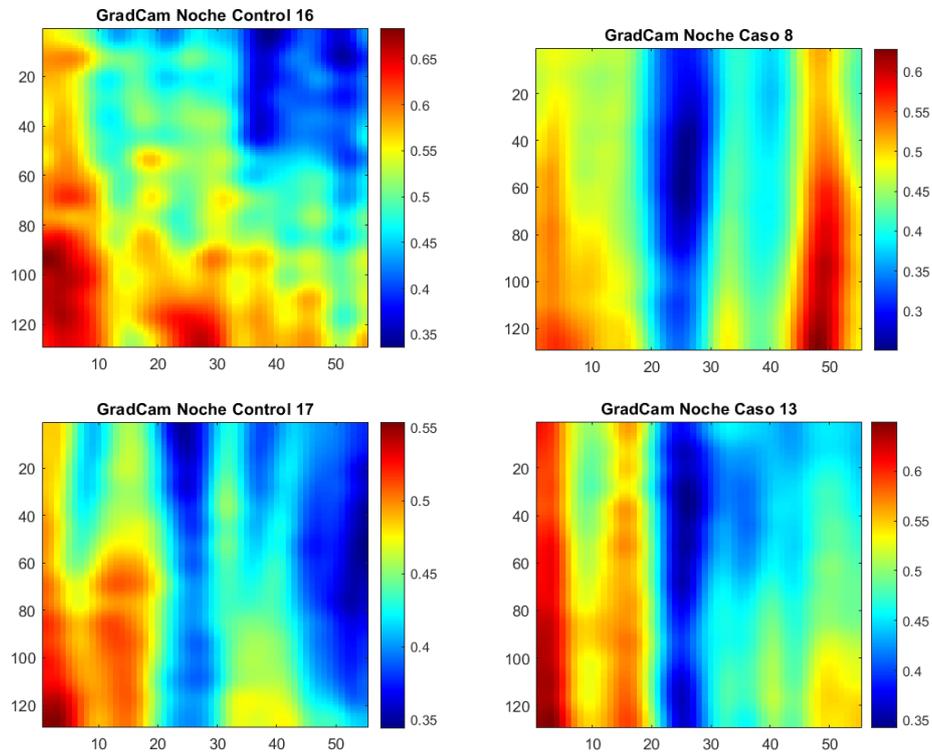


FIGURA 4.16: Mapas de Grad-CAM obtenidos para las señales de actividad nocturna.

actividad. Vamos a organizar estos mapas de oclusión en diferentes subgrupos, organizados simultáneamente por edad y sexo para poder hacer un mejor análisis. A continuación se muestran los resultados más relevantes encontrados en el estudio.

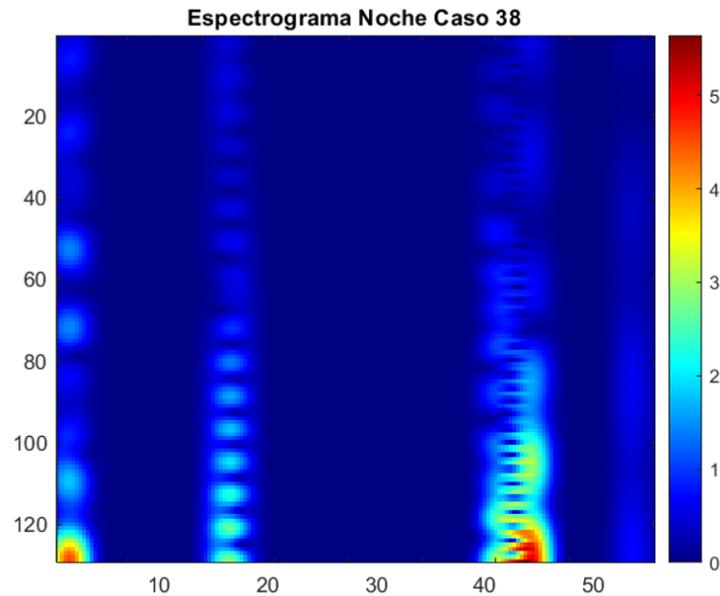


FIGURA 4.17: Ejemplo de espectrograma utilizado para alimentar la CNN.

Chicas de Secundaria

En la figura 4.18 podemos observar que los mapas para casos y controles de la red de chicas para la actividad diurna presentan zonas de activación máximas diferentes entre sí. Para los controles, la red focaliza en la zona central de la ventana de actividad, no dando casi importancia a lo que sucede al final del periodo de tiempo. Presenta cierta bimodalidad a lo largo del tiempo, a diferencia de para los casos, donde tienen algo más de relevancia los periodos finales y la zona de activación máxima es más pequeña, pero el área de influencia se extiende en el tiempo. En cuanto a las frecuencias, en ambos casos las zonas de activación están en las frecuencias medias, por lo que podemos deducir que para este caso tiene más relevancia la evolución a lo largo del tiempo para hacer la diferenciación entre sujetos.

En el caso de la actividad nocturna, que podemos ver en la figura 4.19, las diferencias son más evidentes: mientras que la actividad a lo largo de la ventana que tiene peso en los controles es la del principio, en los casos vuelve a centrarse en el periodo central de actividad. Además, en la actividad nocturna el patrón es contrario al del día y la bimodalidad se observa en los casos. En cuanto a las frecuencias, en los casos parecen tenerse más en cuenta las bajas frecuencias y en los controles, aunque toma en cuenta las bajas frecuencias, tiene importancia todo el rango frecuencial de la primera mitad de la ventana.

Con el objetivo de ver cómo varía el patrón en función de la edad, vamos a mostrar a continuación los mapas de oclusión obtenidos para las chicas de secundaria que veíamos en las figuras 4.18 y 4.19 pero, en esta ocasión, para la red del grupo de secundaria.

Podemos ver en la figura 4.20 de actividad diurna que el mapa de oclusión muestra zonas claramente diferenciadas para los casos y controles; se aprecia de nuevo bimodalidad para los controles y una mayor influencia de la evolución temporal en los casos, con mayor peso para las

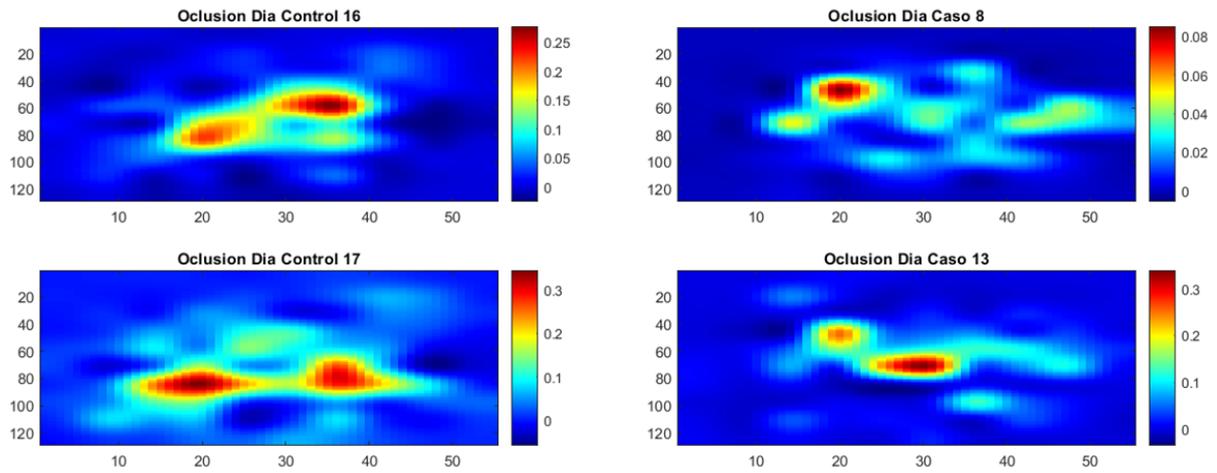


FIGURA 4.18: Mapas de oclusión de actividad diurna para las chicas de secundaria generados para la red entrenada con chicas.

frecuencias medias-altas.

En los mapas de actividad nocturna (figura 4.21), las diferencias son menos evidentes pero se observa una mayor dispersión en tiempo para los casos.

Chicas de Primaria

Realizaremos ahora el mismo análisis que para las chicas de secundaria, pero en el grupo de chicas con edades comprendidas entre 5 y 11 años.

En los mapas de la actividad diurna, vemos en la figura 4.22 bimodalidad en tiempo para los controles. Para la actividad nocturna, la bimodalidad aparece en los casos (figura 4.25), por lo que volvemos a observar patrones contrarios entre casos y controles para la actividad diurna y nocturna.

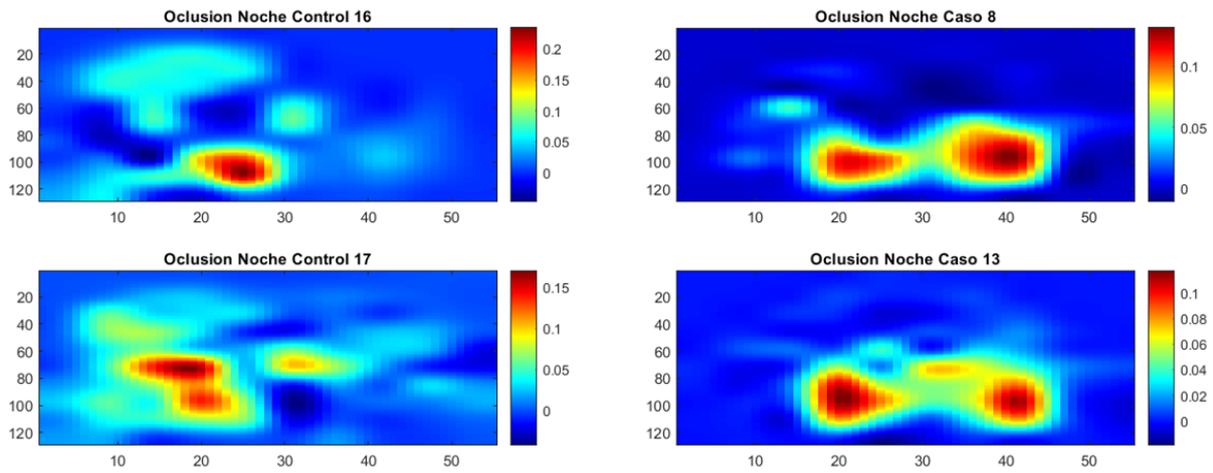


FIGURA 4.19: Mapas de oclusión de actividad nocturna para las chicas de secundaria generados para la red entrenada con chicas.

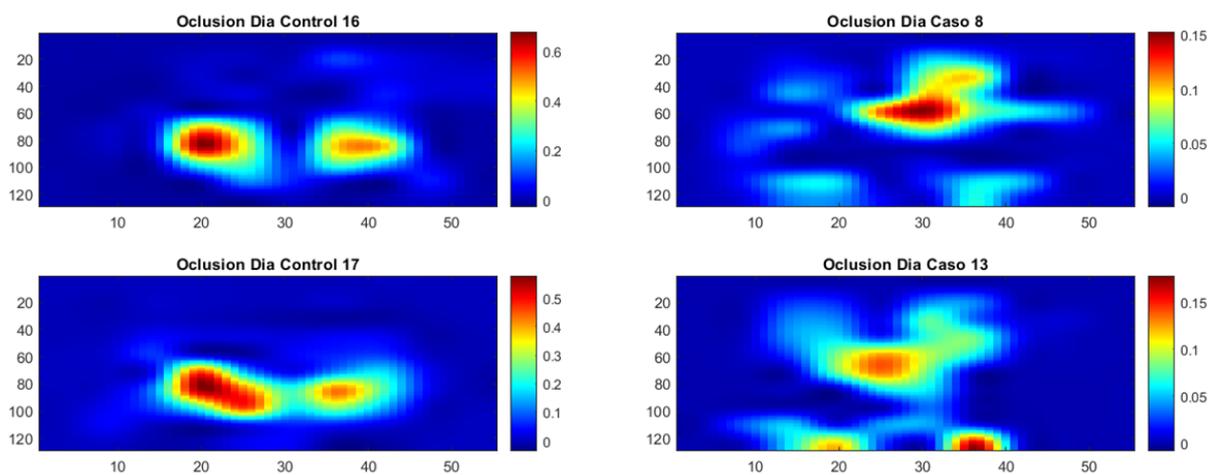


FIGURA 4.20: Mapas de oclusión de actividad diurna para las chicas de secundaria generados para la red entrenada con registros de secundaria.

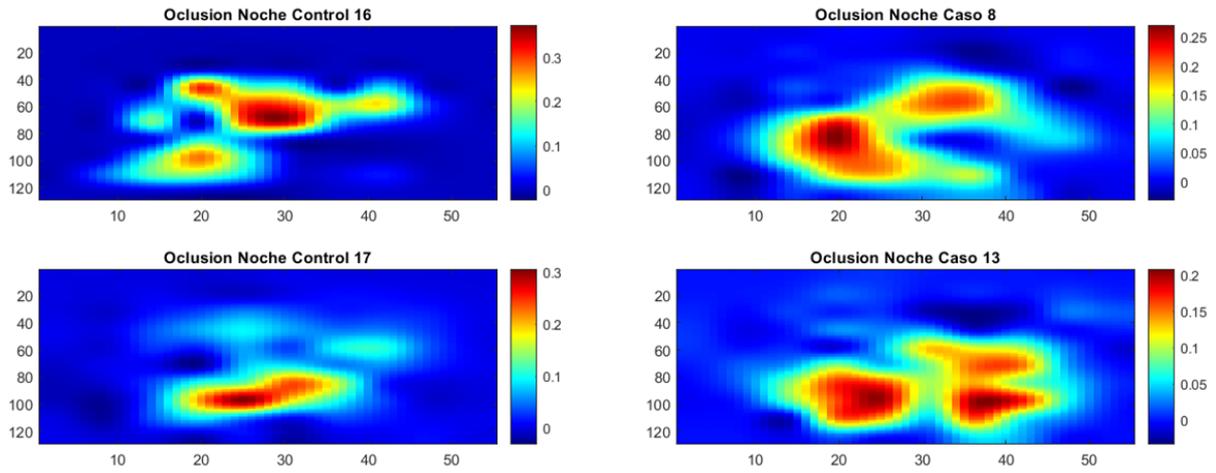


FIGURA 4.21: Mapas de oclusión de actividad nocturna para las chicas de secundaria generados para la red entrenada con registros de secundaria.

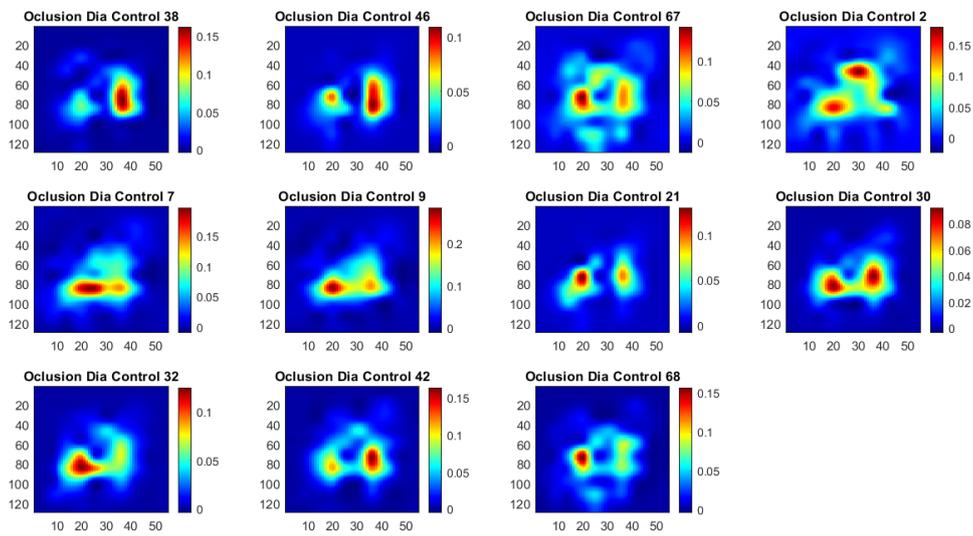


FIGURA 4.22: Mapas de oclusión de actividad diurna para las chicas de primaria para la red entrenada con chicas(CONTROLES).

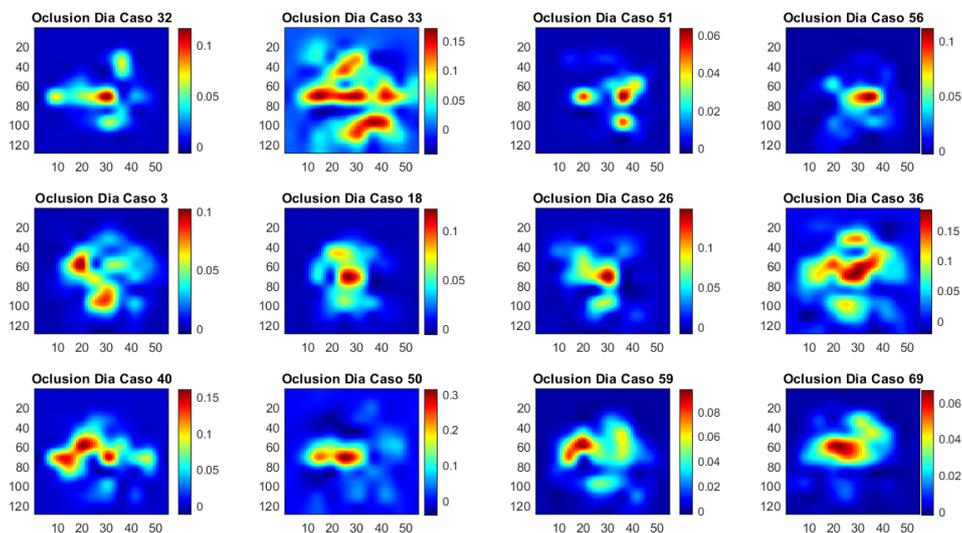


FIGURA 4.23: Mapas de oclusión de actividad diurna de las chicas en primaria para la red entrenada con chicas (CASOS).

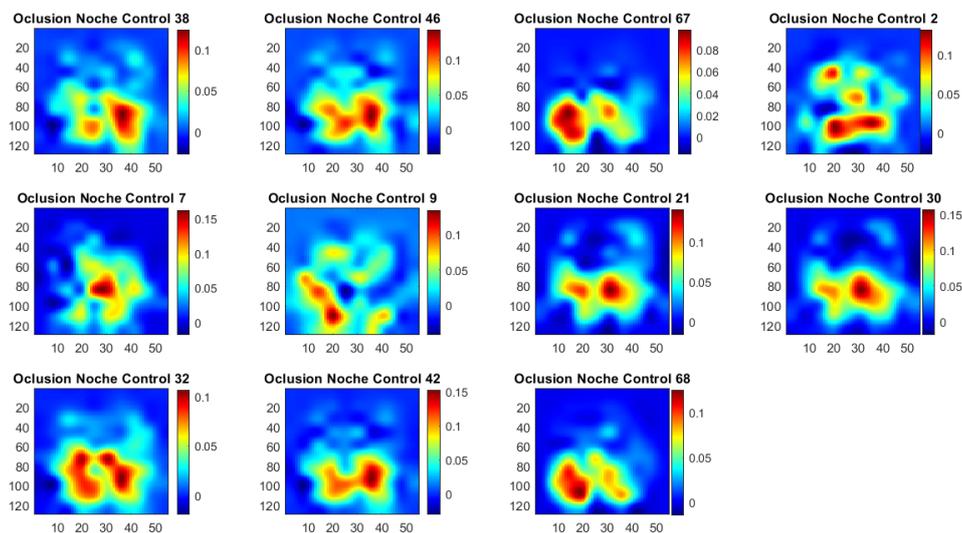


FIGURA 4.24: Mapas de oclusión de actividad nocturna de las chicas de primaria para la red entrenada con chicas (CONTROLES).

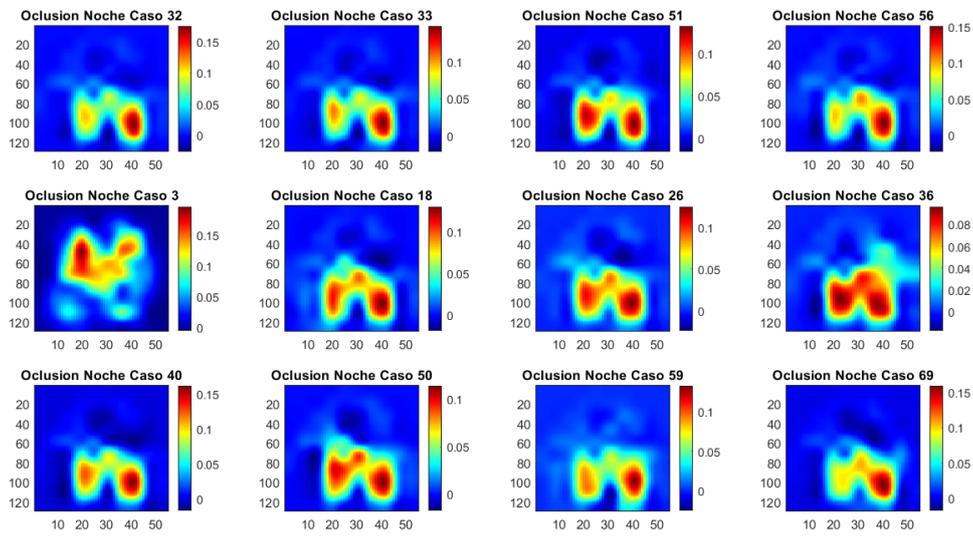


FIGURA 4.25: Mapas de oclusión de actividad nocturna de las chicas de primaria para la red entrenada con chicas (CASOS).

Chicos de Secundaria

En este grupo, como podemos ver en la figura 4.26, ocurre lo mismo que para las chicas si comparamos con los mapas de los casos de la figura 4.27. En la figura 4.29 podemos ver cómo en los casos la actividad está más extendida a lo largo de la ventana temporal que en los controles. En cuanto al estudio solo por sexo, podemos ver en las figuras 4.30 y 4.31 que tienen especial relevancia los mapas de oclusión de actividad diaria, donde se ve que para los controles se tiende a concentrar la información en la zona central de la ventana de actividad, mientras que la región para los casos es más extensa.

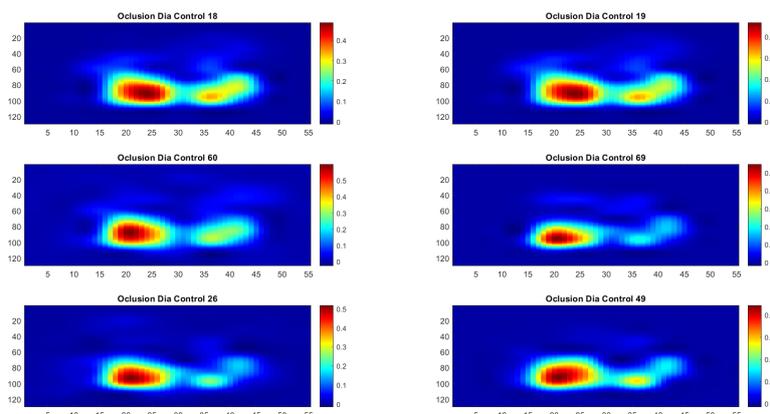


FIGURA 4.26: Mapas de oclusión de actividad diaria para los chicos de secundaria generados para la red entrenada con registros de secundaria.(CONTROLES)

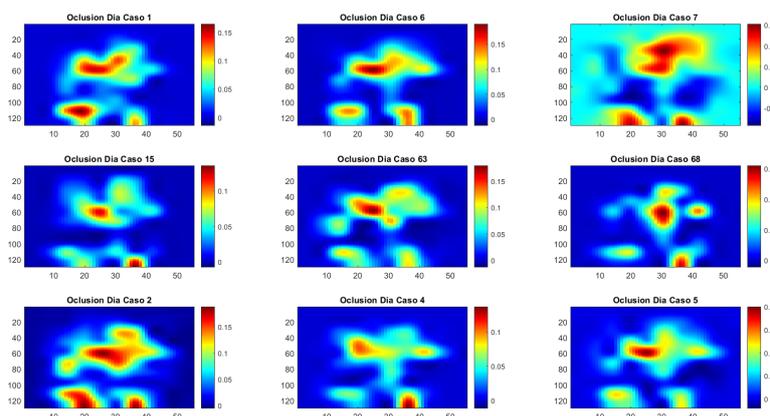


FIGURA 4.27: Mapas de oclusión de actividad diaria para los chicos de secundaria generados para la red entrenada con registros de secundaria.(CASOS)

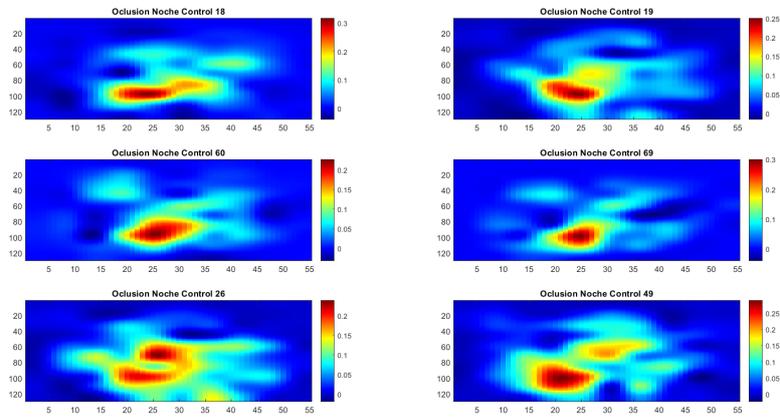


FIGURA 4.28: Mapas de oclusión de actividad nocturna para los chicos de secundaria generados para la red entrenada con registros de secundaria.(CONTROLES)

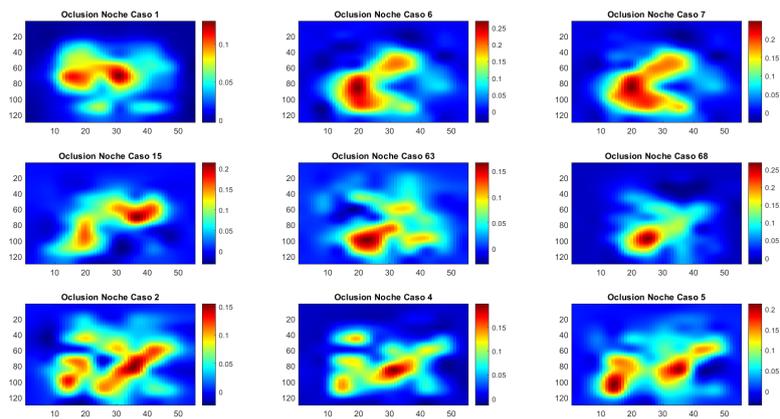


FIGURA 4.29: Mapas de oclusión de actividad nocturna para los chicos de secundaria generados para la red entrenada con registros de secundaria.(CASOS)

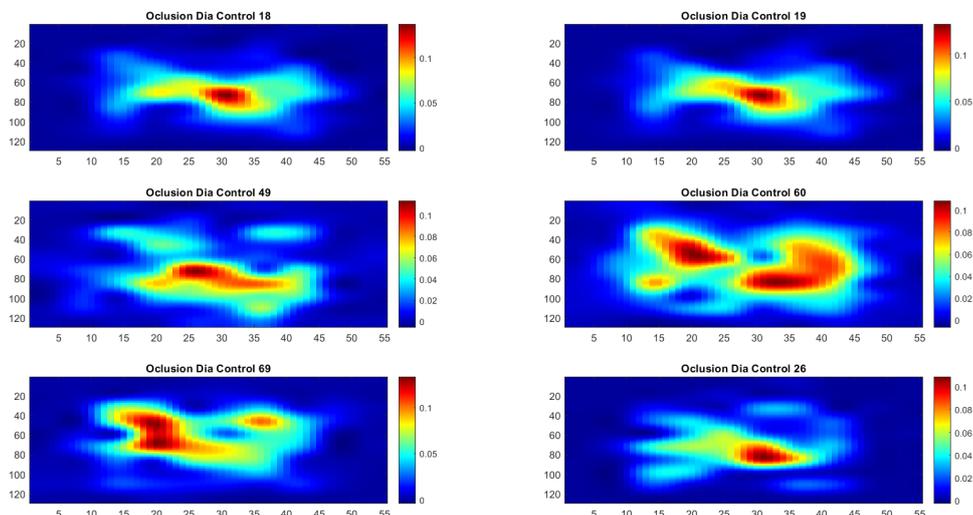


FIGURA 4.30: Mapas de oclusión de actividad diurna para los chicos de secundaria (controles) en la red de chicos.

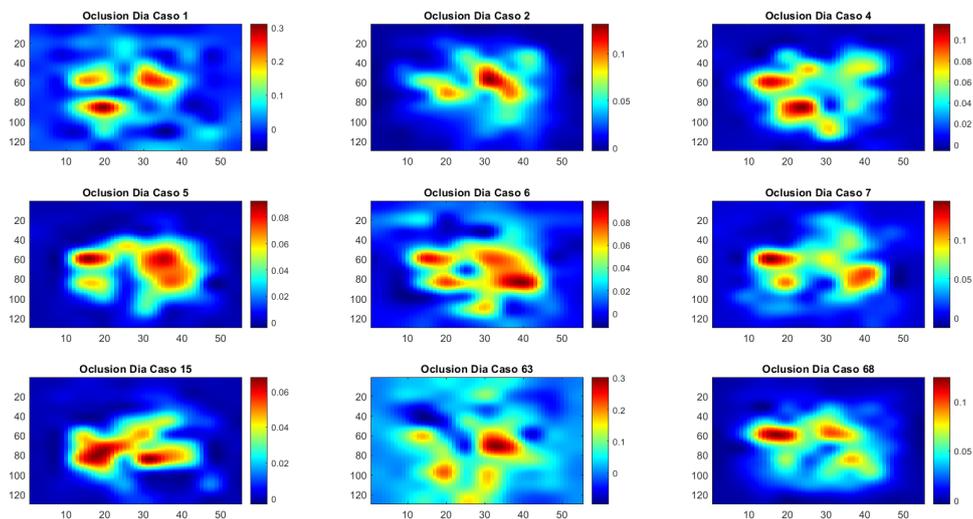


FIGURA 4.31: Mapas de oclusión de actividad diurna para los chicos de secundaria (casos) en la red de chicos.

Chicos de Primaria

En este subgrupo, para la actividad diurna (figura 4.32), nos encontramos con que los casos tienen una zona de activación más extendida en el rango de frecuencias, con varias áreas de valor máximo, mientras que para los controles se concentra en un punto en torno a la mitad de la ventana, tanto en tiempo como en frecuencia. En cuanto a la actividad nocturna (figura 4.33), nos encontramos en ambos casos con una área de activación grande en la zona central, pero en el caso de los controles el máximo se encuentra en zona de frecuencias mayores.

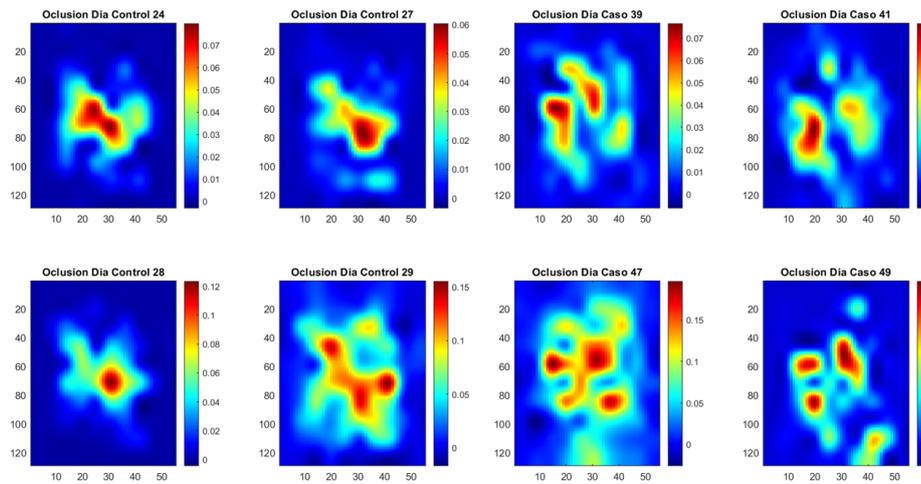


FIGURA 4.32: Mapas de oclusión de actividad diurna para algunos de los casos (derecha) y controles (izquierda) de los chicos de primaria en la red entrenada con chicos.

Si vemos los mapas de oclusión para estos mismos sujetos en la red entrenada con registros de primaria (figura 4.34), encontramos relevantes los mapas de oclusión para la noche, donde se observa que las áreas de activación máxima son muy diferentes para casos y controles. Para los casos, nos encontramos con máximos de tamaño muy pequeño en zonas de baja frecuencia, mientras que para los controles estas áreas son mucho más grandes y a frecuencias más altas. En cuanto a la disposición temporal, los casos presentan los máximos más cercanos al final de la ventana y los controles, más cerca del comienzo.

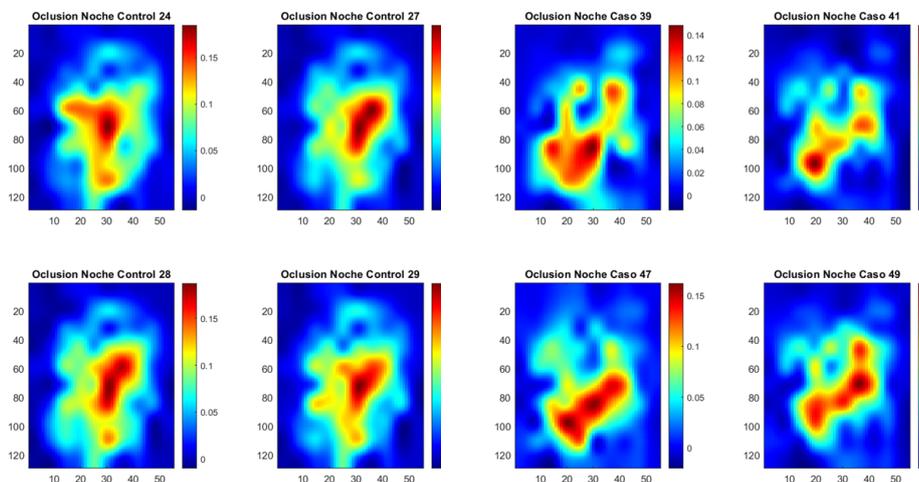


FIGURA 4.33: Mapas de oclusión de actividad nocturna para algunos de los casos (derecha) y controles (izquierda) en los chicos de primaria en la red de chicos.

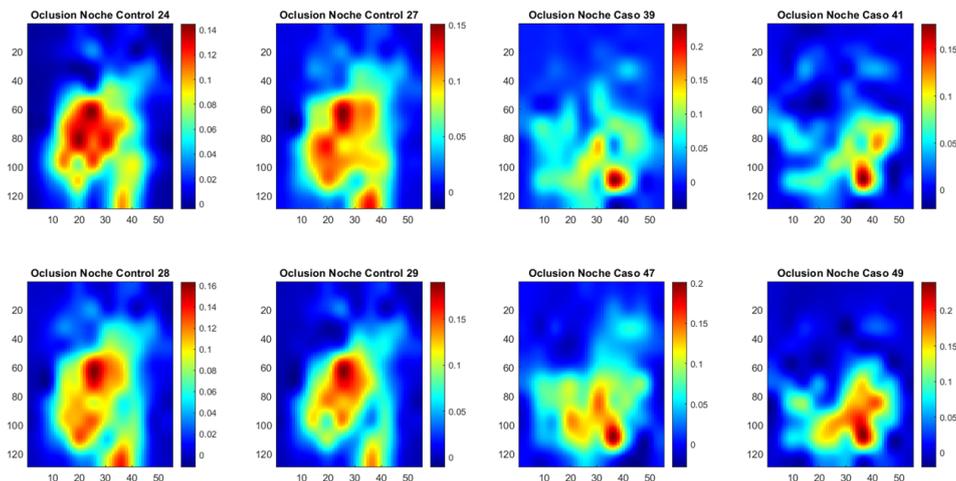


FIGURA 4.34: Mapas de oclusión de actividad nocturna para algunos de los casos (derecha) y controles (izquierda) en los chicos de primaria en la red de primaria.

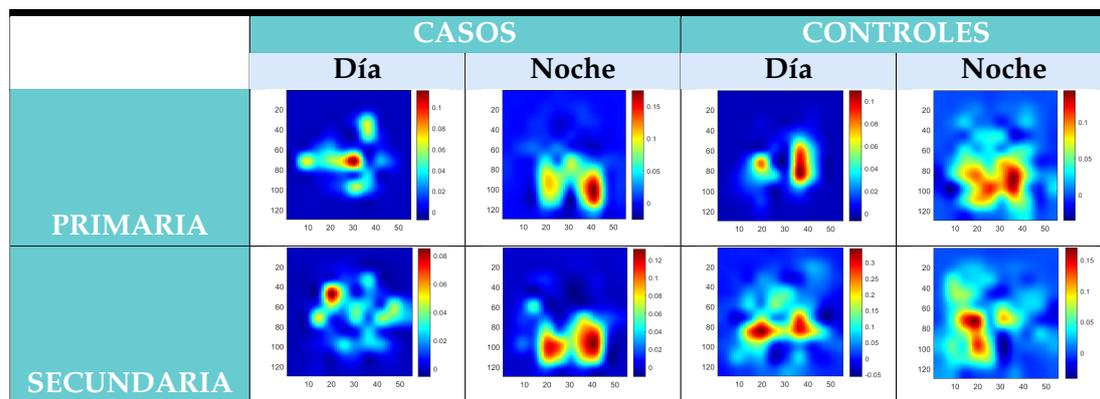


TABLA 4.9: Resumen de los Mapas de oclusión obtenidos para las CNN de chicas.

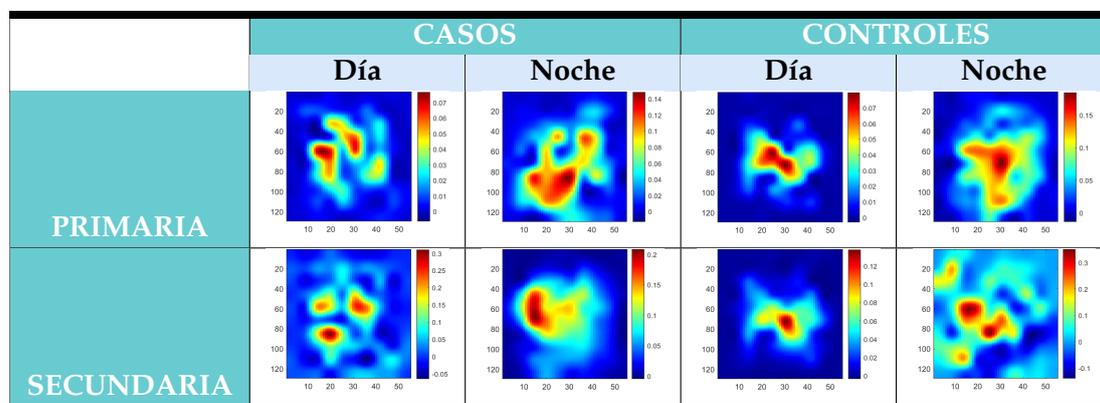


TABLA 4.10: Resumen de los Mapas de oclusión obtenidos para la CNN de chicos.

4.3 DISCUSIÓN DE LOS RESULTADOS CUALITATIVOS OBTENIDOS PARA LOS MAPAS DE OCLUSIÓN

En esta sección resumiremos las conclusiones extraídas del estudio de los mapas de oclusión. Para ello, mostraremos en varias tablas los mapas que hemos presentado en la sección 4.2.2.2 y después presentaremos las conclusiones que hemos extraído de ellos.

4.3.1 RESUMEN DE LOS MAPAS DE OCLUSIÓN OBTENIDOS EN EL ESTUDIO

En las tablas 4.9 a 4.12 se resumen los mapas de oclusión detallados en la sección 4.2.2.2. Para ello, se presentan mapas de oclusión típicos de cada grupo de estudio en función de la CNN que se esté evaluando. Las tablas 4.9 y 4.10 muestran los resultados de las redes entrenadas con registros de sexo y las tablas 4.11 y 4.12, los resultados de las redes entrenadas con registros de edad.

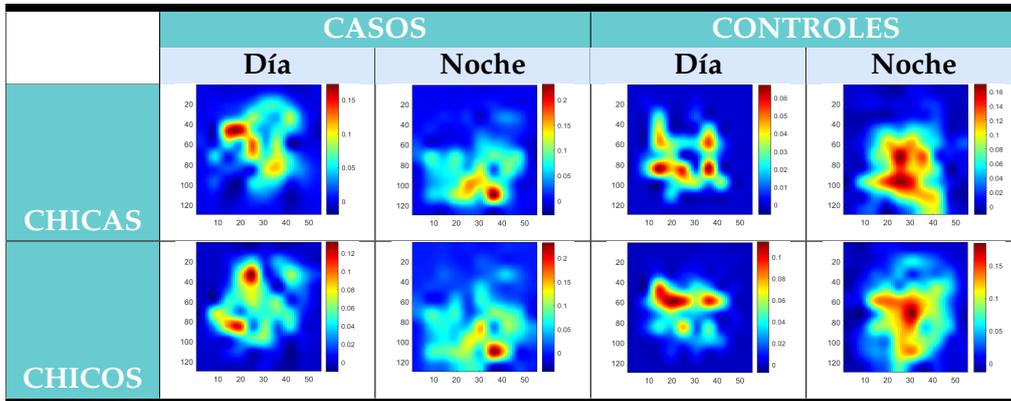


TABLA 4.11: Resumen de los Mapas de ocusión para la CNN de primaria.

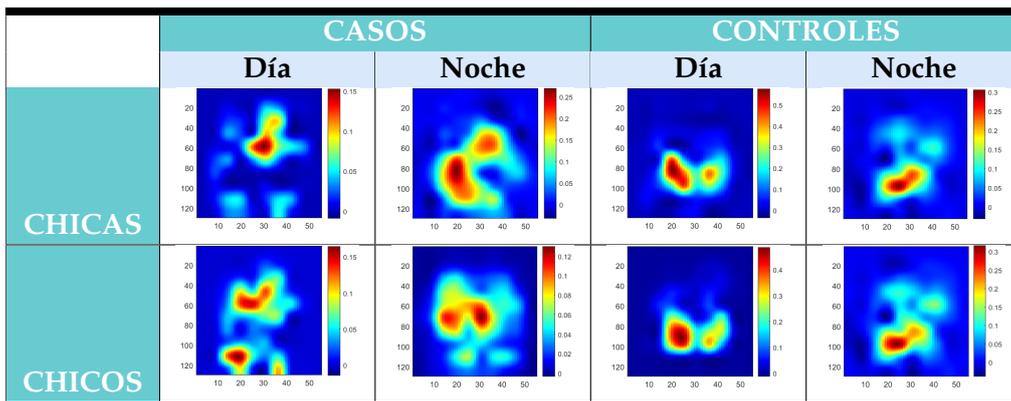


TABLA 4.12: Resumen de los Mapas de ocusión obtenidos para la CNN de secundaria.

4.3.2 DISCUSIÓN

En general, los mapas para los casos presentan una actividad mucho más extendida durante el periodo temporal, mientras que para los controles suele estar más concentrada en un área concreta de la ventana, presentando periodos de baja actividad, lo que podría traducirse en un movimiento más continuo para los casos. Según han mostrado los mapas de oclusión, su zona central es el área donde se concentra la mayor parte de la información, tanto en los casos como en los controles y, en consecuencia, las zonas periféricas tienen, salvo excepciones, menos relevancia. En términos de tiempo y frecuencia, esto se traduce en que para discriminar entre casos y controles la red se centrará en el comportamiento del sujeto durante la realización de una actividad de larga duración y menos relevancia al comienzo y el final de la misma, a excepción del caso de las chicas de secundaria en su actividad nocturna, donde los controles presentan máximos de activación durante la primera mitad del periodo analizado.

Centrándonos en la sectorización solo por sexo (chicos y chicas) y por edad (primaria y secundaria) podemos determinar que:

- En la red entrenada para las chicas (tabla 4.9) podemos observar que el patrón es contrario en la actividad diurna y nocturna para los casos y controles. Los casos presentan bimodalidad en la noche y los controles en el día. Se observa algo más de dispersión si el caso o control estudiado es de primaria.
- Los resultados obtenidos para la red de chicos se resumen en la tabla 4.10. La diferencia más relevante se encuentra en la actividad diurna, siendo para los controles más concentrada que para los casos. No se aprecian diferencias destacables en los patrones de movimiento en función de la edad, en ambos casos se observa bastante dispersión temporal.
- En cuanto a la red de primaria, podemos ver en la tabla 4.11 que los patrones entre chicos y chicas son muy similares, con bastante dispersión temporal. En el caso de los controles, se ve una actividad mucho más concentrada y zonas de baja actividad.
- En la red de secundaria (tabla 4.12) se observa que la actividad diurna de los controles presenta bimodalidad, con la actividad muy concentrada en esa zona, mientras que los casos presentan actividad en toda la ventana. El patrón se repite de forma similar para la actividad nocturna, pero sin bimodalidad para los controles. Se observa además que para los casos de secundaria, las zonas más relevantes son las zonas de alta frecuencia localizadas en la segunda parte de la ventana temporal.
- Los chicos en general presentan mayor actividad que las chicas, aunque para ambos casos se observa dispersión en el patrón de actividad (tablas 4.9 y 4.10).
- Ocurre algo parecido entre primaria y secundaria (tablas 4.11 y 4.12), los patrones que se obtienen de la red de secundaria reflejan una actividad menor que en la red de primaria.

CONCLUSIONES Y LÍNEAS FUTURAS

5.1 CONCLUSIONES

Al comienzo de este documento se estableció como objetivo principal del presente trabajo el desarrollo de un sistema experto que permitiera identificar diferencias en la manifestación del TDAH para grupos sectorizados, empleando técnicas de visualización de mapas de activaciones para redes neuronales convolucionales y utilizando como entradas para la red neuronal señales de actimetría transformadas en un espacio 2D, según la metodología descrita en [4].

Tras realizar la clasificación sectorizada, se estudiaron diversas técnicas de visualización de activaciones concluyendo en la conveniencia de emplear mapas de oclusión. Para la obtención de dichos mapas, se utilizaron imágenes que maximizaran las activaciones de la red y, una vez obtenidos los mapas para cada imagen, se promediaron para obtener un mapa único para cada sujeto de estudio.

Una vez obtenidos los mapas, se realizó una comparación a nivel cualitativo de dichos mapas, para evaluar las diferencias entre casos y controles, así como las diferencias en función del grupo de estudio.

Teniendo en cuenta los resultados obtenidos al finalizar tanto la clasificación como el análisis cualitativo, podemos extraer las siguientes conclusiones:

- Las prestaciones del sistema de diagnóstico basado en la CNN diseñada en [5] para grupos reducidos son buenas, aunque los resultados se ven perjudicados debido a la escasez de sujetos de estudio en alguno de los grupos.
- El empleo de mapas de oclusión para obtener las activaciones de la CNN diseñada, resulta el método más adecuado entre los analizados, para las señales de entrada.
- Los mapas de oclusión han permitido extraer patrones relacionados con la patología de estudio de una forma subjetiva por apreciación visual.

5.2 LIMITACIONES Y LÍNEAS FUTURAS

A lo largo del desarrollo de este Trabajo Fin de Máster, se han encontrado algunas limitaciones. La escasez de sujetos en los grupos de chicas y secundaria hace que la generalización sea una tarea difícil. En relación a la interpretación de las características de la CNN, encontrar un método que proporcionara un resultado interpretable ha sido una tarea compleja. Se han empleado numerosas técnicas y procesados previos hasta llegar a los mapas finales que han permitido obtener las respuestas esperadas.

Los mapas de oclusión han permitido realizar un estudio cualitativo, pero resultaría interesante desarrollar un método que permitiera realizar un análisis cuantitativo de los patrones extraídos de los resultados.

En conclusión, los resultados obtenidos en este TFM han resultado adecuados a los objetivos que se perseguían; no obstante, aunque los resultados son prometedores aún queda un largo camino por delante para conseguir que sean de utilidad clínica para un diagnóstico objetivo del TDAH.

BIBLIOGRAFÍA

- [1] American Psychiatric Association. Diagnostic And Statistical Manual Of Mental Disorders, (DSM-V). *American Journal of Psychiatry*, 5ª edición, 2013.
- [2] Erik G Willcutt. The prevalence of dsm-iv attention-deficit/hyperactivity disorder: a meta-analytic review. *Neurotherapeutics*, 9(3):490–499, 2012.
- [3] Fiorella Rusca-Jordán and Carla Cortez-Vergara. Trastorno por déficit de atención con hiperactividad (tdah) en niños y adolescentes. una revisión clínica. *Revista de Neuro-Psiquiatría*, 83(3):148–156, 2020.
- [4] Patricia Amado-Caballero, Pablo Casaseca-de-la Higuera, Susana Alberola-Lopez, Jesus Maria Andres-de Llano, Jose Antonio Lopez Villalobos, Jose Ramon Garmendia-Leiza, and Carlos Alberola-Lopez. Objective adhd diagnosis using convolutional neural networks over daily-life activity records. *IEEE journal of biomedical and health informatics*, 24(9):2690–2700, 2020.
- [5] Patricia Amado Caballero et al. Ayuda al diagnóstico de tdah en la infancia mediante técnicas de procesamiento de señal y aprendizaje. 2019.
- [6] MathWorks MATLAB for Deep Learning. <https://es.mathworks.com/>, Último acceso Diciembre 2018.
- [7] D Martín Martínez, P Casaseca de la Higuera, S Alberola López, J Andrés de Llano, JA López Villalobos, J Ardua Fernández, and C Alberola López. Nonlinear Analysis Of Actigraphic Signals For The Assessment Of The Attention-Deficit/Hyperactivity Disorder (ADHD). *Medical Engineering & Physics*, 34(9):1317–1329, 2012.
- [8] S. Cortese, S. Faraone, E. Konofal, and M. Lecendreux. Sleep in Children With Attention-Deficit/Hyperactivity Disorder: Meta-Analysis of Subjective and Objective Studies. *J. Am. Acad. Child. Adolesc. Psychiatry*, 48:894–908, 2009.
- [9] R.W. Tuthill. Hair Lead Levels Related to Children’s Classroom Attention-deficit Behavior. *Arch. Environ. Health*, 52:214–220, 1996.
- [10] R.E. Weiss, M.A. Stein, B. Trommer, and S. Refetoff. Attention-Deficit Hyperactivity Disorder and Thyroid Function. *J. Pediatr.*, 123:539–545, 1993.

- [11] M. El-Faddagh, M. Laucht, A. Maras, L. Vöhringer, and M. H. Schmidt. Association of Dopamine D4 Receptor (DRD4) Gene with Attention-Deficit/Hyperactivity Disorder (ADHD) in a High-Risk Community Sample: a Longitudinal Study from Birth to 11 Years of Age. *J. Neural. Transm.*, 111:883–889, 2004.
- [12] M.M Van-Veen, J.J.S. Kooij, M.M. Boonstra, M.C.M.Gordijn, and E.J.W. Van-Someren. Delayed Circadian Rhythm in Adults with Attention-Deficit/Hyperactivity Disorder and Chronic Sleep-Onset Insomnia. *Biol. Psychiatry*, 67:1091 – 1096, 2010.
- [13] S. Kurtek, E. Klassen, Zhaohua Ding, S.W. Jacobson, J.L. Jacobson, M.J. Avison, and A. Srivastava. Parameterization-Invariant Shape Comparisons of Anatomical Surfaces. *IEEE Trans. Med. Imag.*, 30:849 –858, 2011.
- [14] M. Ashtari, S. Kumra, S.L. Bhaskar, T. Clarke, E. Thaden, K.L. Cervellione, J. Rhinewine, J.M. Kane, A. Adesman, R. Milanaik, J. Maytal, A. Diamond, P. Szeszko, and B.A. Ardekani. Attention-Deficit/Hyperactivity Disorder: A Preliminary Diffusion Tensor Imaging Study. *Biol. Psychiatry*, 57:448–455, 2005.
- [15] T.J. Silk, N. Vance, A. Rinehart, J.L. Bradshaw, and R. Cunningtin. White-Matter Abnormalities in Attention Deficit Hyperactivity Disorder: A Diffusion Tensor Imaging Study. *Hum. Brain. Map*, 30:2757–2765, 2009.
- [16] L. Zimmer. Positron Emission Tomography Neuroimaging for a Better Understanding of the Biology of ADHD. *Neuropharmacology*, 57:601–607, 2009.
- [17] Atif Riaz, Muhammad Asad, Eduardo Alonso, and Greg Slabaugh. Fusion of fmri and non-imaging data for adhd classification. *Computerized Medical Imaging and Graphics*, 65: 115–128, 2018.
- [18] Huaiqiang Sun, Ying Chen, Qiang Huang, Su Lui, Xiaoqi Huang, Yan Shi, Xin Xu, John A. Sweeney, and Qiyong Gong. Psychoradiologic utility of mr imaging for diagnosis of attention deficit hyperactivity disorder: A radiomics analysis. *Radiology*, 287(2):620–630, May 2018.
- [19] M. Modarres-Zadeh, Xueyan Xu, A. Jaromczyk, and S. Senders. A Neuro-Behavioral Test and Algorithms for Screening and Evaluating Therapy in ADHD. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 2486–2489, 2005.
- [20] J.A. Gutierrez Gnechi, J.C. Herrera Garcia, and J. de Dios Ortiz Alvarado. Auxiliary Neurofeedback System for Diagnostic of Attention Deficit Hyperactivity Disorder. In *Electronics, Robotics and Automotive Mechanics Conference, 2007. CERMA 2007*, pages 135 –138, 2007.
- [21] S. Hyun-Lee, B. Abibullaev, Won-Seok Kang, Yunhee Shin, and Jinung An. Analysis of Attention Deficit Hyperactivity Disorder in EEG Using Wavelet Transform and Self Organizing Maps. In *Control Automation and Systems (ICCAS), 2010 International Conference on*, pages 2439 –2442, 2010.

- [22] D. Ming, M. Zhang, Y. Xi, H. Qi, Y. Hu, and K.D.K. Luk. Multiscale entropy analysis of attention related EEG based on motor imaginary potential. In *Computational Intelligence for Measurement Systems and Applications, 2009. CIMSAS '09. IEEE International Conference on*, pages 24–27, 2009.
- [23] Y. Nan, J. Wang, S.A. Xue, H. Sheng, Y. Jiao, and J. Wang. Analysis of Propagation of Multi-Channel EEG in the Test of Sustained Attention. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 1666–1669, 2010.
- [24] T.S. Hale, S.L. Smalley, P.D. Walshaw, G. Hanada, J. Macion, J.T. McCracken, J.J. McGough, and S.K. Loo. Atypical EEG Beta Asymmetry in Adults with ADHD. *Neuropsychologia*, 48: 3532–3539, 2010.
- [25] F. Ghassemi, M.H. Moradi, M. Tehrani-Doost, and V. Abootalebi. Classification of ADHD/Normal Participants Using Frequency Features of ERP's Independent Components. In *Biomedical Engineering (ICBME), 2010 17th Iranian Conference of*, pages 1–4, 2010.
- [26] P. Castro-Cabrera, J. Gómez-García, F. Restrepo, O. Moscoso, and G. Castellanos-Dominguez. Evaluation Of Feature Extraction Techniques On Event-Related Potentials For Detection Of Attention-Deficit/Hyperactivity Disorder. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 851–854, 2010.
- [27] C. Demanuele, C.J. James, E.J.S. Sonuga-Barke, and A. Capilla. Low Frequency Phase Synchronisation Analysis of MEG Recordings from Children with ADHD and Controls Using Single Channel ICA. In *Advances in Medical, Signal and Information Processing, 2008. MEDSIP 2008. 4th IET International Conference on*, pages 1–4, 2008.
- [28] Hüseyin Öztoprak, Mehmet Toycan, Yaşar Kemal Alp, Orhan Arıkan, Elvin Doğutepe, and Sirel Karakaş. Machine-based classification of adhd and nonadhd participants using time/frequency features of event-related neuroelectric activity. *Clinical Neurophysiology*, 128(12):2400–2410, 2017.
- [29] R. Gruber, A. Sadeh, and A. Raviv. Instability of Sleep Patterns in Children With Attention-Deficit/Hyperactivity Disorder. *J. Am. Acad. Child. Adolesc. Psychiatry*, 39:495–501, 2000.
- [30] R. Silvestri, A. Gagliano, I. Aricò, T. Calarese, C. Cedro, O. Bruni, R. Conduro, E. Germanó, G. Gervasi, R. Siracusano, G. Vita, and P. Bramanti. Sleep Disorders in Children with Attention-Deficit/Hyperactivity Disorder (ADHD) Recorded Overnight by Video-Polysomnography. *Sleep Med.*, 10:1132–1138, 2009.
- [31] M.H. Teicher, Y. Ito, C.A. Glod, and P. Wallace. Quantitative Assessment of Fidgeting in Attention-Deficit Hyperactivity Disorder (ADHD). *Am. Acad. Child Adolescence Psychiatry*, 9:42–43, 1993.
- [32] M.H. Teicher, Y. Ito, C.A. Glod, and N.I. Barber. Objective Measurement of Hyperactivity and Attentional Problems in ADHD. *J. Am. Acad. Child. Adolesc. Psychiatry*, 35:334–342, 1996.

- [33] A.V. Dane, R.J. Schachar, and R. Tannok. Does Actigraphy Differentiate ADHD Subtypes in a Clinical Research Setting? *J. Am. Acad. Child. Adolesc. Psychiatry*, 39:752–760, 2000.
- [34] M.H. Teicher, C. A. Glod, K. Pahlavan, D. Harper, E. Magnus, and F. Wren. Locomotor Activity and the Diagnosis of ADHD. *Am. Psychiatr Assoc. Abstr.*, 141:98, 1991.
- [35] H.J. Kam, K. Lee, S.M. Cho, Y.M. Shin, and R.W. Park. High-Resolution Actigraphic Analysis of ADHD: A Wide Range of Movement Variability Observation in Three School Courses – A Pilot Study. *Healthc. Inform. Res.*, 17(19–37), 2011.
- [36] P. Casaseca-de-la-Higuera, D. Martín-Martínez, S. Alberola-López, J. María Andrés-de-Llano, J. Antonio López-Villalobos, J. Ramón-Garmendia Leiza, and C. Alberola-López. Automatic diagnosis of adhd based on multichannel nonlinear analysis of actimetry registries. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4204–4207, Aug 2012. doi: 10.1109/EMBC.2012.6346894.
- [37] Mario Muñoz-Organero, Lauren Powell, Ben Heller, Val Harpin, and Jack Parker. Automatic extraction and detection of characteristic movement patterns in children with adhd based on a convolutional neural network (cnn) and acceleration images. *Sensors*, 18(11):3924, 2018.
- [38] Niamh O’Mahony, Blanca Florentino-Liano, Juan J Carballo, Enrique Baca-García, and Antonio Artés Rodríguez. Objective diagnosis of adhd using imus. *Medical engineering & physics*, 36(7):922–926, 2014.
- [39] Aja Louise Murray, Tom Booth, Manuel Eisner, Bonnie Auyeung, George Murray, and Denis Ribeaud. Sex differences in adhd trajectories across childhood and adolescence. *Developmental science*, 22(1):e12721, 2019.
- [40] Maryanne Loyer Carbonneau, Martin Demers, Marc Bigras, and Marie-Claude Guay. Meta-analysis of sex differences in adhd symptoms and associated cognitive deficits. *Journal of Attention Disorders*, 25(12):1640–1656, 2021.
- [41] Daniel F Hermens, Michael R Kohn, Simon D Clarke, Evian Gordon, and Leanne M Williams. Sex differences in adolescent adhd: findings from concurrent eeg and eda. *Clinical neurophysiology*, 116(6):1455–1463, 2005.
- [42] I Lazzaro, E Gordon, W Li, CL Lim, M Plahn, S Whitmont, S Clarke, RJ Barry, A Dosen, and R Meares. Simultaneous eeg and eda measures in adolescent attention deficit hyperactivity disorder. *International Journal of Psychophysiology*, 34(2):123–134, 1999.
- [43] Adam R Clarke, Robert J Barry, Rory McCarthy, and Mark Selikowitz. Age and sex effects in the eeg: differences in two subtypes of attention-deficit/hyperactivity disorder. *Clinical neurophysiology*, 112(5):815–826, 2001.
- [44] Keri S Rosch, Stewart H Mostofsky, and Mary Beth Nebel. Adhd-related sex differences in fronto-subcortical intrinsic functional connectivity and associations with delay discounting. *Journal of neurodevelopmental disorders*, 10(1):1–14, 2018.

- [45] Stefan Jonas, Andrea O Rossetti, Mauro Oddo, Simon Jenni, Paolo Favaro, and Frederic Zubler. Eeg-based outcome prediction after cardiac arrest with convolutional neural networks: Performance and visualization of discriminative features. *Human brain mapping*, 40(16):4606–4617, 2019.
- [46] Thomas Uyttenhove, Aren Maes, Tom Van Steenkiste, Dirk Deschrijver, and Tom Dhaene. Interpretable epilepsy detection in routine, interictal eeg data using deep learning. In *Machine Learning for Health*, pages 355–366. PMLR, 2020.
- [47] Jinzheng Cai, Fuyong Xing, Abhinandan Batra, Fujun Liu, Glenn A Walter, Krista Vandenberg, and Lin Yang. Texture analysis for muscular dystrophy classification in mri with improved class activation mapping. *Pattern recognition*, 86:368–375, 2019.
- [48] Hwei Geok Ng, Matthias Kerzel, Jan Mehnert, Arne May, and Stefan Wermter. Classification of mri migraine medical data using 3d convolutional neural network. In *International Conference on Artificial Neural Networks*, pages 300–309. Springer, 2018.
- [49] Sarah Eskreis-Winkler, Natsuko Onishi, Katja Pinker, Jeffrey S Reiner, Jennifer Kaplan, Elizabeth A Morris, and Elizabeth J Sutton. Using deep learning to improve nonsystematic viewing of breast cancer on mri. *Journal of Breast Imaging*, 3(2):201–207, 2021.
- [50] Fernando Andreotti, Huy Phan, and Maarten De Vos. Visualising convolutional neural network decisions in automatic sleep scoring. In *CEUR Workshop Proceedings*, pages 70–81. CEUR Workshop Proceedings, 2018.
- [51] He Chen, Yan Song, and Xiaoli Li. Use of deep learning to detect personalized spatial-frequency abnormalities in eegs of children with adhd. *Journal of neural engineering*, 16(6):066046, 2019.
- [52] Christopher M Florkowski. Sensitivity, specificity, receiver-operating characteristic (roc) curves and likelihood ratios: communicating the performance of diagnostic tests. *The Clinical Biochemist Reviews*, 29(Suppl 1):S83, 2008.
- [53] Anne B Arnett, Bruce F Pennington, Erik G Willcutt, John C DeFries, and Richard K Olson. Sex differences in adhd symptom severity. *Journal of Child Psychology and Psychiatry*, 56(6):632–639, 2015.
- [54] Florence D Mowlem, Mina A Rosenqvist, Joanna Martin, Paul Lichtenstein, Philip Asherson, and Henrik Larsson. Sex differences in predicting adhd clinical diagnosis and pharmacological treatment. *European child & adolescent psychiatry*, 28(4):481–489, 2019.
- [55] José-Ramón Valdizán Usón, Elvira Mercado Val, and Alberto Mercado Undanivia. Características y variabilidad clínica del tdah en niñas.
- [56] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1):145–151, 1999.
- [57] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

- [58] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.
- [59] Ruth Fong, Mandela Patrick, and Andrea Vedaldi. Understanding deep networks via extremal perturbations and smooth masks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2950–2958, 2019.
- [60] Christoph Molnar. *Interpretable machine learning*. Lulu. com, 2020.
- [61] Matthew A Turk and Alex P Pentland. Face recognition using eigenfaces. In *Proceedings. 1991 IEEE computer society conference on computer vision and pattern recognition*, pages 586–587. IEEE Computer Society, 1991.
- [62] Alexander Mordvintsev, Christopher Olah, and Mike Tyka. Inceptionism: Going deeper into neural networks. 2015.
- [63] Chunpeng Wu, Wei Fan, Yuan He, Jun Sun, and Satoshi Naoi. Handwritten character recognition by alternately trained relaxation convolutional neural network. In *2014 14th International Conference on Frontiers in Handwriting Recognition*, pages 291–296. IEEE, 2014.
- [64] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016.
- [65] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.
- [66] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.