



UNIVERSIDAD DE VALLADOLID



Escuela Técnica Superior de Ingenieros de Telecomunicación  
*Universidad de Valladolid*

## TRABAJO FIN DE MÁSTER

### MÁSTER EN INGENIERÍA DE TELECOMUNICACIÓN

# ESTUDIO DE LA UTILIDAD DE DIFERENTES MÉTODOS DE INTERPRETABILIDAD DE REDES PROFUNDAS EN EL DIAGNÓSTICO DEL TDAH

---

AUTOR: GUILLERMO CHICO DELGADO  
TUTOR: PATRICIA AMADO CABALLERO  
J.P. CASASECA DE LA HIGUERA

2 de septiembre de 2025

TÍTULO: ESTUDIO DE LA UTILIDAD DE DIFERENTES MÉTODOS DE INTERPRETABILIDAD DE REDES PROFUNDAS EN EL DIAGNÓSTICO DEL TDAH

AUTOR: GUILLERMO CHICO DELGADO

TUTOR: PATRICIA AMADO CABALLERO  
J.P. CASASECA DE LA HIGUERA

DEPARTAMENTO: TSCIT

### **Miembros del Tribunal**

PRESIDENTE: JUAN IGNACIO ASENSIO PÉREZ

SECRETARIO: BEATRIZ SAINZ DE ABAJO

VOCAL: JAVIER MANUEL AGUIAR PÉREZ

SUPLENTE 1: PEDRO CHAMORRO POSADA

SUPLENTE 2: LUIS MIGUEL SAN JOSÉ REVUELTA

CALIFICACIÓN:

### **Resumen del proyecto**

Este Trabajo de Fin de Máster tiene como objetivo avanzar en la línea de investigación de estudios previos sobre el diagnóstico del Trastorno por Déficit de Atención e Hiperactividad (TDAH), especialmente en su subtipo hiperactivo. Tradicionalmente, el diagnóstico del TDAH se basa en evaluaciones clínicas, las cuales son subjetivas debido a la interpretación de los síntomas y la observación del paciente. El propósito de este estudio es desarrollar un sistema de diagnóstico objetivo utilizando técnicas de aprendizaje profundo e Inteligencia Artificial Explicable (XAI). El trabajo se enfoca principalmente en caracterizar y diferenciar los subtipos de TDAH mediante el análisis de señales actigráficas, procesadas en espectrogramas, utilizando Redes Neuronales Convolucionales y Vision Transformers. Además, se emplean métodos de XAI para interpretar las salidas de las redes neuronales e identificar características clave que distinguen los subtipos de TDAH. La investigación también compara el rendimiento de los diferentes modelos, con el fin de mejorar la precisión del diagnóstico y proporcionar una visión más clara de las características del subtipo hiperactivo. Este trabajo contribuye a reducir la subjetividad en el diagnóstico y ofrece un enfoque prometedor para un diagnóstico más preciso y eficiente del TDAH.

### **Palabras clave**

TDAH, Aprendizaje Profundo, CNN, Transformer, Espectrograma, Mapas de oclusión, Grad-CAM, Mecanismos de Atención

### **Abstract**

This Master's Thesis aims to advance the line of research of previous studies on the diagnosis of Attention Deficit Hyperactivity Disorder (ADHD), especially in its hyperactive subtype.

Traditionally, the diagnosis of ADHD is based on clinical assessments, which are subjective due to the interpretation of symptoms and observation of the patient. The purpose of this study is to develop an objective diagnostic system using deep learning techniques and Explainable Artificial Intelligence (XAI). The work mainly focuses on characterizing and differentiating ADHD subtypes by analyzing actigraphic signals, processed in spectrograms, using Convolutional Neural Networks and Vision Transformers. In addition, XAI methods are used to interpret the outputs of the neural networks and identify key features that distinguish the ADHD subtypes. The research also compares the performance of the different models, in order to improve the accuracy of the diagnosis and provide a clearer view of the characteristics of the hyperactive subtype. This work contributes to reducing subjectivity in diagnosis and offers a promising approach for a more accurate and efficient diagnosis of ADHD.

**Keywords**

ADHD, Deep Learning, CNN, Transformer, Spectrogram, Occlusion Sensitivity, Grad-CAM, Attention Mechanisms

## AGRADECIMIENTOS

Primeramente, quiero agradecer a mis tutores, Juan Pablo Casaseca de la Higuera y Patricia Amado Caballero, por brindarme la oportunidad de realizar este Trabajo Fin de Máster, así como por su constante apoyo y orientación durante todo el proceso.

Asimismo, quiero expresar mi gratitud al Laboratorio de Procesado de Imagen por proporcionarme los recursos necesarios para llevar a cabo este trabajo.

Agradezco también a mi familia por su incondicional apoyo a lo largo de todos estos años de estudios, y a mis amigos y compañeros, quienes han sido parte fundamental de mi trayectoria académica.

# ÍNDICE GENERAL

<b>1. Introducción</b>	<b>1</b>
1.1. Motivación	1
1.2. Objetivos	2
1.3. Fases y Métodos	3
1.4. Medios necesarios empleados para el desarrollo	4
1.5. Estructura del documento	5
<b>2. Antecedentes y estado del arte</b>	<b>6</b>
2.1. Introducción al TDAH	6
2.1.1. Subtipos de TDAH	7
2.1.2. Diagnóstico tradicional del TDAH	7
2.2. Diagnósticos objetivos	8
2.2.1. Resonancia magnética	8
2.2.2. Electroencefalograma	9
2.2.3. Eye-tracking	9
2.2.4. Actimetría	10
2.2.5. Machine Learning	10
2.3. Estado del arte: diagnóstico del TDAH mediante Deep Learning y actimetría	11
2.3.1. Estudios previos que utilizan la actimetría y Deep Learning	11
<b>3. Aprendizaje profundo</b>	<b>15</b>
3.1. Introducción al Deep Learning	15
3.1.1. Jerarquías de la Inteligencia Artificial	16
3.1.2. Redes neuronales	18
3.1.2.1. Neurona artificial	18
3.1.2.2. Redes neuronales multicapa	20
3.1.2.3. Entrenamiento de redes neuronales	21
3.1.2.4. Aprendizaje de redes neuronales	25
3.2. Redes profundas	25
3.2.1. Redes neuronales feed-forward	26
3.2.2. Redes neuronales recurrentes	27
3.2.3. Redes neuronales convolucionales	28
3.2.4. Transformers	30
3.2.4.1. Vision transformer	32
3.3. Inteligencia artificial explicable	34
3.3.1. Métodos de XAI	34
3.3.1.1. Mapas de oclusión	34
3.3.1.2. Grad-CAM	36

3.3.1.3. Mecanismos de atención . . . . .	37
<b>4. Metodología</b>	<b>39</b>
4.1. Procesamiento de los datos . . . . .	40
4.1.1. Señales de actimetría . . . . .	40
4.1.2. Creación de espectrogramas . . . . .	41
4.1.3. Sectorización de pacientes . . . . .	42
4.2. Aplicación de redes neuronales . . . . .	43
4.2.1. Diseño de la CNN . . . . .	43
4.2.2. Diseño del Transformer . . . . .	44
4.2.3. Entrenamiento de las redes . . . . .	45
4.3. Interpretación mediante técnicas XAI . . . . .	47
4.3.1. Creación de mapas . . . . .	47
4.3.2. Enmascaramiento de espectrogramas . . . . .	51
4.3.3. Cálculo de características espectrales . . . . .	56
4.3.4. Análisis estadístico: p-valores . . . . .	58
<b>5. Resultados y discusión</b>	<b>60</b>
5.1. Presentación de los resultados . . . . .	60
5.1.1. Resultados: comparativa Tipo1 vs Tipo3 . . . . .	61
5.1.1.1. Tipo1 vs Tipo3: CNN capa convolucional 3 . . . . .	61
5.1.1.2. Tipo1 vs Tipo3: CNN capa convolucional 2 . . . . .	62
5.1.1.3. Tipo1 vs Tipo3: CNN capa convolucional 1 . . . . .	63
5.1.1.4. Tipo1 vs Tipo3: Transformer . . . . .	64
5.1.2. Resultados: comparativa Tipo2 vs Tipo3 . . . . .	65
5.1.2.1. Tipo2 vs Tipo3: CNN capa convolucional 3 . . . . .	65
5.1.2.2. Tipo2 vs Tipo3: CNN capa convolucional 2 . . . . .	66
5.1.2.3. Tipo2 vs Tipo3: CNN capa convolucional 1 . . . . .	67
5.1.2.4. Tipo2 vs Tipo3: Transformer . . . . .	68
5.2. Análisis de los resultados . . . . .	69
5.2.1. Análisis mediante boxplots . . . . .	71
5.2.1.1. Boxplots: Tipo1 vs Tipo3 . . . . .	72
5.2.1.2. Boxplots: Tipo2 vs Tipo3 . . . . .	73
5.2.2. Comparativa con trabajos previos . . . . .	74
5.2.3. Análisis sin enmascaramiento . . . . .	75
5.3. Discusión de los resultados . . . . .	76
<b>6. Conclusiones y líneas futuras</b>	<b>78</b>
6.1. Conclusiones . . . . .	78
6.2. Limitaciones y Líneas Futuras . . . . .	79
<b>A. Boxplots para características espectrales significativas</b>	<b>80</b>
A.1. Boxplots: Tipo1 vs Tipo3 . . . . .	80
A.2. Boxplots: Tipo2 vs Tipo3 . . . . .	96

# ÍNDICE DE FIGURAS

3.1. Jerarquía entre IA, ML y DL . . . . .	17
3.2. Estructura de una neurona artificial [46] . . . . .	18
3.3. Funciones de activación [48] . . . . .	20
3.4. Funciones de activación [49] . . . . .	21
3.5. Funcionamiento de validación cruzada K-fold . . . . .	23
3.6. Funcionamiento de validación cruzada aleatoria . . . . .	24
3.7. Funcionamiento de validación cruzada Leave-One-Out . . . . .	24
3.8. Estructura de una FNN [58] . . . . .	26
3.9. Estructura de una RNN [60] . . . . .	27
3.10. Estructura de una CNN [62] . . . . .	28
3.11. Funcionamiento de capa convolucional [64] . . . . .	29
3.12. Funcionamiento de capa pooling con los métodos max pooling y average pooling [65] . . . . .	29
3.13. Estructura básica de Transformers [67] . . . . .	31
3.14. Funcionamiento de un Vision Transformer [66] . . . . .	33
3.15. Proceso de generación de mapas de oclusión . . . . .	35
3.16. Proceso de Grad-CAM [76] . . . . .	37
4.1. Flujo de trabajo de la sección . . . . .	40
4.2. Ejemplo de uno de los espectrogramas obtenidos de 129x55 píxeles (imagen ampliada) . . . . .	42
4.3. Arquitectura de la CNN . . . . .	44
4.4. Ejemplo de mapa de oclusión de la CNN . . . . .	48
4.5. Ejemplo de mapa Grad-CAM de la CNN . . . . .	48
4.6. Ejemplo de mapa de atención de la CNN . . . . .	49
4.7. Ejemplo de mapa de oclusión del Transformer . . . . .	49
4.8. Ejemplo de mapa Grad-CAM del Transformer . . . . .	50
4.9. Ejemplo de mapa de atención del Transformer . . . . .	50
4.10. Máscara para diferentes porcentajes de enmascaramiento. De izquierda a derecha: 60 %, 70 % y 80 % . . . . .	51
4.11. Proceso de enmascaramiento a partir de un mapa de oclusión para la CNN . . . . .	52
4.12. Proceso de enmascaramiento a partir de un mapa Grad-CAM para la CNN . . . . .	53
4.13. Proceso de enmascaramiento a partir de un mapa de activación para la CNN . . . . .	53
4.14. Proceso de enmascaramiento a partir de un mapa de oclusión para el Transformer . . . . .	54
4.15. Proceso de enmascaramiento a partir de un mapa Grad-CAM para el Transformer . . . . .	54
4.16. Proceso de enmascaramiento a partir de un mapa de atención para el Transformer . . . . .	55
5.1. Boxplots de las 10 mejores características para Tipo1 vs Tipo3 . . . . .	72
5.2. Boxplots de las 10 mejores características para Tipo2 vs Tipo3 . . . . .	73
A.1. Boxplots relevantes CNN Capa 3 Noche Oclusión Tipo1 vs Tipo3 . . . . .	80
A.2. Boxplots relevantes CNN Capa 3 Noche GradCAM Tipo1 vs Tipo3 . . . . .	81

A.3. <i>Boxplots relevantes CNN Capa 3 Noche Atención Tipo1 vs Tipo3</i> . . . . .	81
A.4. <i>Boxplots relevantes CNN Capa 3 Día Oclusión Tipo1 vs Tipo3</i> . . . . .	82
A.5. <i>Boxplots relevantes CNN Capa 2 Noche Oclusión Tipo1 vs Tipo3</i> . . . . .	83
A.6. <i>Boxplots relevantes CNN Capa 2 Noche GradCAM Tipo1 vs Tipo3</i> . . . . .	84
A.7. <i>Boxplots relevantes CNN Capa 2 Día Oclusión Tipo1 vs Tipo3</i> . . . . .	85
A.8. <i>Boxplots relevantes CNN Capa 2 Día GradCAM Tipo1 vs Tipo3</i> . . . . .	86
A.9. <i>Boxplots relevantes CNN Capa2 Día Atención Tipo1 vs Tipo3</i> . . . . .	87
A.10. <i>Boxplots relevantes CNN Capa 1 Noche Oclusión Tipo1 vs Tipo3</i> . . . . .	88
A.11. <i>Boxplots relevantes CNN Capa 1 Noche Atención Tipo1 vs Tipo3</i> . . . . .	88
A.12. <i>Boxplots relevantes CNN Capa 1 Día Oclusión Tipo1 vs Tipo3</i> . . . . .	89
A.13. <i>Boxplots relevantes CNN Capa 1 Día GradCAM Tipo1 vs Tipo3</i> . . . . .	90
A.14. <i>Boxplots relevantes CNN Capa 1 Día Atención Tipo1 vs Tipo3</i> . . . . .	91
A.15. <i>Boxplots relevantes Transformers Noche Oclusión Tipo1 vs Tipo3</i> . . . . .	91
A.16. <i>Boxplots relevantes Transformers Noche Atención Capa 1 Tipo1 vs Tipo3</i> . . . . .	92
A.17. <i>Boxplots relevantes Transformers Noche Atención Capa 2 Tipo1 vs Tipo3</i> . . . . .	93
A.18. <i>Boxplots relevantes Transformers Día Oclusión Tipo1 vs Tipo3</i> . . . . .	94
A.19. <i>Boxplots relevantes Transformers Día Atención Capa 1 Tipo1 vs Tipo3</i> . . . . .	95
A.20. <i>Boxplots relevantes Transformers Día Atención Capa 2 Tipo1 vs Tipo3</i> . . . . .	96
A.21. <i>Boxplots relevantes CNN Capa 3 Noche Oclusión Tipo2 vs Tipo3</i> . . . . .	96
A.22. <i>Boxplots relevantes CNN Capa 3 Noche GradCAM Tipo2 vs Tipo3</i> . . . . .	97
A.23. <i>Boxplots relevantes CNN Capa 3 Noche Atención Tipo2 vs Tipo3</i> . . . . .	98
A.24. <i>Boxplots relevantes CNN Capa 3 Día Oclusión Tipo2 vs Tipo3</i> . . . . .	99
A.25. <i>Boxplots relevantes CNN Capa 3 Día Atención Tipo2 vs Tipo3</i> . . . . .	100
A.26. <i>Boxplots relevantes CNN Capa 2 Noche Oclusión Tipo2 vs Tipo3</i> . . . . .	101
A.27. <i>Boxplots relevantes CNN Capa 2 Noche GradCAM Tipo2 vs Tipo3</i> . . . . .	102
A.28. <i>Boxplots relevantes CNN Capa 2 Noche Atención Tipo2 vs Tipo3</i> . . . . .	103
A.29. <i>Boxplots relevantes CNN Capa 2 Día Oclusión Tipo2 vs Tipo3</i> . . . . .	104
A.30. <i>Boxplots relevantes CNN Capa 2 Día GradCAM Tipo2 vs Tipo3</i> . . . . .	105
A.31. <i>Boxplots relevantes CNN Capa 2 Día Atención Tipo2 vs Tipo3</i> . . . . .	106
A.32. <i>Boxplots relevantes CNN Capa 1 Noche Oclusión Tipo2 vs Tipo3</i> . . . . .	106
A.33. <i>Boxplots relevantes CNN Capa 1 Noche GradCAM Tipo2 vs Tipo3</i> . . . . .	107
A.34. <i>Boxplots relevantes CNN Capa 1 Noche Atención Tipo2 vs Tipo3</i> . . . . .	108
A.35. <i>Boxplots relevantes CNN Capa 1 Día Oclusión Tipo2 vs Tipo3</i> . . . . .	109
A.36. <i>Boxplots relevantes CNN Capa 1 Día GradCAM Tipo2 vs Tipo3</i> . . . . .	110
A.37. <i>Boxplots relevantes CNN Capa 1 Día Atención Tipo2 vs Tipo3</i> . . . . .	111
A.38. <i>Boxplots relevantes Transformers Noche Oclusión Tipo2 vs Tipo3</i> . . . . .	112
A.39. <i>Boxplots relevantes Transformers Noche Atención Capa 1 Tipo2 vs Tipo3</i> . . . . .	113
A.40. <i>Boxplots relevantes Transformers Noche Atención Capa 2 Tipo2 vs Tipo3</i> . . . . .	114
A.41. <i>Boxplots relevantes Transformers Día Oclusión Tipo2 vs Tipo3</i> . . . . .	115
A.42. <i>Boxplots relevantes Transformers Día Atención Capa 2 Tipo2 vs Tipo3</i> . . . . .	116

# ÍNDICE DE TABLAS

2.1. Síntomas de inatención e hiperactividad/impulsividad del TDAH . . . . .	8
2.2. Estudios que utilizan la actimetría y Deep Learning . . . . .	13
2.3. Estudios que utilizan la actimetría y Deep Learning . . . . .	14
3.1. Diferencias entre IA, ML y DL . . . . .	17
4.1. Sujetos de estudio de cada grupo . . . . .	43
4.2. Número de espectrogramas disponibles para cada tipo . . . . .	45
4.3. Distribución de los folds no balanceados . . . . .	46
4.4. Media de las precisiones del test de las redes entrenadas con 10 folds . . . . .	46
5.1. <i>p</i> -valores de las características frecuenciales para Tipo1 vs Tipo3, en el periodo nocturno para la capa convolucional 3 de la CNN . . . . .	61
5.2. <i>p</i> -valores de las características frecuenciales para Tipo1 vs Tipo3, en el periodo diurno para la capa convolucional 3 de la CNN . . . . .	62
5.3. <i>p</i> -valores de las características frecuenciales para Tipo1 vs Tipo3, en el periodo nocturno para la capa convolucional 2 de la CNN . . . . .	62
5.4. <i>p</i> -valores de las características frecuenciales para Tipo1 vs Tipo3, en el periodo diurno para la capa convolucional 2 de la CNN . . . . .	63
5.5. <i>p</i> -valores de las características frecuenciales para Tipo1 vs Tipo3, en el periodo nocturno para la capa convolucional 1 de la CNN . . . . .	63
5.6. <i>p</i> -valores de las características frecuenciales para Tipo1 vs Tipo3, en el periodo diurno para la capa convolucional 1 de la CNN . . . . .	64
5.7. <i>p</i> -valores de las características frecuenciales para Tipo1 vs Tipo3, en el periodo nocturno para el Transformer . . . . .	64
5.8. <i>p</i> -valores de las características frecuenciales para Tipo1 vs Tipo3, en el periodo diurno para el Transformer . . . . .	65
5.9. <i>p</i> -valores de las características frecuenciales para Tipo2 vs Tipo3, en el periodo nocturno para la capa convolucional 3 de la CNN . . . . .	66
5.10. <i>p</i> -valores de las características frecuenciales para Tipo2 vs Tipo3, en el periodo diurno para la capa convolucional 3 de la CNN . . . . .	66
5.11. <i>p</i> -valores de las características frecuenciales para Tipo2 vs Tipo3, en el periodo nocturno para la capa convolucional 2 de la CNN . . . . .	67
5.12. <i>p</i> -valores de las características frecuenciales para Tipo2 vs Tipo3, en el periodo diurno para la capa convolucional 2 de la CNN . . . . .	67
5.13. <i>p</i> -valores de las características frecuenciales para Tipo2 vs Tipo3, en el periodo nocturno para la capa convolucional 1 de la CNN . . . . .	68
5.14. <i>p</i> -valores de las características frecuenciales para Tipo2 vs Tipo3, en el periodo diurno para la capa convolucional 1 de la CNN . . . . .	68

---

5.15. <i>p</i> -valores de las características frecuenciales para Tipo2 vs Tipo3, en el periodo nocturno para el Transformer . . . . .	69
5.16. <i>p</i> -valores de las características frecuenciales para Tipo2 vs Tipo3, en el periodo diurno para el Transformer . . . . .	69
5.17. 10 características con mejores <i>p</i> -valores para Tipo1 vs Tipo3 . . . . .	70
5.18. 10 características con mejores <i>p</i> -valores para Tipo2 vs Tipo3 . . . . .	70
5.19. <i>P</i> -valores más bajos por parámetro para las comparaciones Tipo1 vs Tipo3 y Tipo2 vs Tipo3 obtenidos en trabajo anterior . . . . .	74
5.20. <i>p</i> -valores de las características frecuenciales de los mapas de oclusión de la CNN obtenidos en la capa convolucional 3 de la CNN sin enmascarar para Tipo1 vs Tipo3 . . . . .	75
5.21. <i>p</i> -valores de las características frecuenciales de los mapas de oclusión de la CNN obtenidos en la capa convolucional 3 de la CNN sin enmascarar para Tipo2 vs Tipo3 . . . . .	76

## Capítulo 1

# INTRODUCCIÓN

En este capítulo, se busca establecer el contexto del presente Trabajo Fin de Máster, el cual pretende avanzar en el estudio y diagnóstico de los subtipos de trastorno por déficit de atención e hiperactividad (TDAH), en especial del subtipo hiperactivo. Tanto el diagnóstico del TDAH como de su subtipo se basan en evaluaciones clínicas, ligadas a la subjetividad ya que implican la interpretación de síntomas y la observación del paciente. Así pues, el objetivo del trabajo es ofrecer un sistema capaz de identificar los subtipos de TDAH de manera objetiva mediante el uso de herramientas de *Deep Learning* e Inteligencia Artificial Explicable.

En las siguientes secciones, se presentan la motivación y los objetivos del trabajo, así como los medios y métodos empleados para la realización del trabajo y la estructura del mismo.

## 1.1 MOTIVACIÓN

---

El trastorno por déficit de atención e hiperactividad (TDAH) es un desorden neuroconductual frecuente que puede presentarse en niños y adultos. Se manifiesta por problemas constantes en áreas como la atención, la hiperactividad y la impulsividad, afectando tanto al desempeño cotidiano como al desarrollo personal. Según los criterios del *Diagnostic and Statistical Manual of Mental Disorders* (DSM-V), el diagnóstico del TDAH se realiza mediante una evaluación exhaustiva que incluye entrevistas clínicas, recopilación de información y observación del comportamiento del individuo durante un periodo mínimo de 6 meses. Durante este tiempo deben evidenciarse síntomas incompatibles con su nivel de desarrollo, que influyen negativamente en sus actividades sociales y académicas [1].

Según los criterios diagnósticos del DSM-V [1], el trastorno por déficit de atención e hiperactividad (TDAH) se clasifica en tres subtipos principales: predominantemente hiperactivo-impulsivo (TDAH-H), predominantemente inatento (TDAH-I) y combinado (TDAH-C). El subtipo hiperactivo-impulsivo se caracteriza por una mayor presencia de hiperactividad e impulsividad, mientras que en el subtipo inatento los problemas de atención son predominantes. Por su parte, el subtipo combinado incluye síntomas significativos tanto de inatención como de hiperactividad/impulsividad. La determinación del subtipo específico se basa en la cantidad de síntomas observados en cada área, diagnosticándose como TDAH-I o TDAH-H cuando

predominan los síntomas de inatención o de hiperactividad/impulsividad, respectivamente. Si ambos grupos de síntomas están presentes de manera significativa, el diagnóstico corresponde al TDAH combinado.

El correcto diagnóstico de subtipo de TDAH es fundamental ya que cada uno presenta características y necesidades particulares que influyen en el tratamiento y manejo del trastorno. Como se ha visto, el diagnóstico tradicionalmente se ha basado en criterios subjetivos, dependiendo enormemente de la veracidad de la información aportada por el entorno del paciente y de la adecuada evaluación de los especialistas médicos. Por ello, resulta crucial contar con métodos de diagnóstico objetivos y fiables que apoyados en datos cuantitativos y tecnologías avanzadas reduzcan la subjetividad y mejoren la precisión en la identificación de los subtipos de TDAH.

Con este propósito, se han desarrollado estudios que buscan elaborar un sistema objetivo de diagnóstico del TDAH basado en *Deep Learning* como [2] o [3]. En ellos, se utilizaron registros de actimetría tomados durante la vida diaria de los pacientes, aplicando una red neuronal convolucional (CNN) para clasificar entre sujetos con y sin TDAH, con resultados prometedores en términos de precisión y aplicabilidad clínica. Por su parte, en [4], se realizó un análisis centrado en la diferenciación de subtipos, concretamente entre el subtipo inatento y el combinado, también utilizando CNNs y registros de comportamiento.

En cuanto al subtipo hiperactivo, en [5] se desarrolló un estudio cuyo objetivo fue caracterizar específicamente este subtipo, comparándolo con los subtipos inatento y combinado. Para ello, se analizaron los resultados del entrenamiento de una red neuronal convolucional (CNN) mediante técnicas de interpretación del modelo, como los mapas declusión y el modelado por gaussianas, permitiendo identificar patrones distintivos de forma visual y estadística. Métodos similares de interpretación también fueron empleados en [6] y [7].

Sin embargo, uno de los principales desafíos en este tipo de estudios es la escasa disponibilidad de muestras correspondientes al subtipo hiperactivo (TDAH-H). Esta limitación dificulta tanto el entrenamiento robusto de modelos como la generalización de los resultados. Ante esta dificultad, la motivación del presente trabajo es avanzar en la caracterización del TDAH-H mediante la aplicación de nuevas metodologías de interpretación de modelos (XAI) y la exploración de arquitecturas de red más recientes y complejas, con el fin de mejorar la comprensión de las diferencias entre subtipos desde una perspectiva tanto clínica como computacional.

## 1.2 OBJETIVOS

---

El trastorno por déficit de atención e hiperactividad de tipo hiperactivo (TDAH-H) es el subtipo menos frecuente en los diagnósticos de TDAH. De acuerdo con a [8], se calcula que solo entre el 13 % y el 17 % de las personas con TDAH cumplen los requisitos para recibir este diagnóstico, lo que lo convierte en el subtipo menos diagnosticado dentro del espectro del TDAH.

El objetivo principal de este trabajo es desarrollar un sistema de análisis basado en registros de actimetría y técnicas de *Deep Learning*, con el fin de caracterizar el subtipo hiperactivo del TDAH (TDAH-H). Para ello, se busca avanzar respecto al enfoque planteado en [5], tanto a nivel de arquitectura como de interpretabilidad del modelo.

En concreto, se utilizarán redes neuronales convolucionales (CNN), así como una nueva arquitectura basada en *Transformers*. Además, se incorporarán herramientas más avanzadas de Inteligencia Artificial Explicable (XAI), incluyendo nuevos tipos de mapas de activación y análisis mediante características espectrales extraídas de los datos, con el fin de mejorar la interpretación de los resultados y facilitar la diferenciación entre subtipos.

Este estudio tiene como objetivos principales:

- Crear un sistema basado en *Deep Learning* capaz de establecer y reconocer las características distintivas del subtipo hiperactivo de TDAH, mediante comparativas frente a los subtipos inatento y combinado.
- Interpretar el comportamiento de la salida de las redes neuronales profundas.
- Comparar el rendimiento y los resultados de cada modelo de red neuronal utilizado.
- Obtener diferencias y características relevantes entre subtipos de TDAH mediante las herramientas de XAI utilizadas.

### 1.3 FASES Y MÉTODOS

---

En esta sección se describe la metodología empleada para alcanzar los objetivos establecidos:

- Generación de espectrogramas a partir de las señales actigráficas de los pacientes y sectorización de los espectrogramas según el subtipo de TDAH diagnosticado previamente.
- Elaboración de conjuntos de datos enfocados en el estudio del subtipo hiperactivo: TDAH-I vs TDAH-H y TDAH-C vs TDAH-H.
- Creación de las redes neuronales empleadas: una red neuronal convolucional (CNN) y un *Vision Transformer* (ViT) y entrenamiento de ambas para cada combinación de subtipos y periodo de actividad.
- Generación de mapas de activación, a partir de técnicas de Inteligencia Artificial Explicable (XAI): mapas de oclusión, Grad-CAM y mecanismos de atención.
- Enmascaramiento de los espectrogramas en función de los mapas de activación obtenidos y cálculo de sus características espectrales.
- Presentación y análisis de los resultados obtenidos con el fin de de caracterizar los subtipos de TDAH, concretamente el subtipo hiperactivo, y discusión de los resultados.
- Obtención de resultados y conclusiones que permitan avanzar en la caracterización del subtipo hiperactivo, mejorando el desempeño de trabajos anteriores.

## 1.4 MEDIOS NECESARIOS EMPLEADOS PARA EL DESARROLLO

---

Durante la realización del trabajo, se han empleado las siguientes herramientas de *hardware* y *software*:

### Hardware

- PC de sobremesa con distribución Linux del Laboratorio de Procesado de Imagen (LPI) de la Universidad de Valladolid (UVa) con las siguientes características:
  - Procesador 6xIntel®Core™i7-7700 3.60 GHz.
  - 16 GB de memoria RAM.
  - Disco duro HDD de 250 GB de capacidad.
- Conexión a servidores del Laboratorio de Procesado de Imagen (LPI) de la Universidad de Valladolid (UVa) para procesamiento de GPU.
- PC de sobremesa personal con Windows 11 con las siguientes características:
  - Procesador 6xIntel®Core™i7-8700 3.20 GHz.
  - 16 GB de memoria RAM.
  - Disco duro HDD de 1 TB de capacidad.
  - GPU NVIDIA GeForce GTX 1050 Ti con 4 GB de memoria RAM

### Software

- MATLAB R2023a [9]: Software de lenguaje de alto nivel que permite el desarrollo de algoritmos, el análisis de datos y computación numérica.
- L<sup>A</sup>T<sub>E</sub>X [10]: Software utilizado para la composición de documentos de texto de alta calidad, ampliamente utilizado en el ámbito académico para la creación de artículos científicos. En LaTeX, los documentos se escriben en un lenguaje de marcado que incluye comandos y estructuras específicas para controlar la apariencia del texto, el diseño de páginas, la creación de tablas, ecuaciones matemáticas, referencias bibliográficas, etc.
- Python: Lenguaje de programación ampliamente utilizado en el ámbito del *Deep Learning*, junto con diversas librerías como:
  - TensorFlow [11]: Biblioteca de código abierto para computación numérica y construcción de modelos de *Deep Learning*.
  - Keras [12] : API de alto nivel para construir y entrenar modelos de *Deep Learning*, integrada con TensorFlow.
  - NumPy [13] : Librería fundamental para la manipulación de arrays y operaciones matemáticas.
  - Matplotlib [14]: Librería para la creación de gráficos y visualizaciones.

- scikit-learn [15] : Librería para *Machine Learning* que incluye herramientas de preprocesamiento y análisis de datos.

## 1.5 ESTRUCTURA DEL DOCUMENTO

---

El presente trabajo se compone de un total de seis capítulos, incluyendo el capítulo de introducción. A continuación, se detallan los capítulos que conforman este documento:

**Capítulo 1- Introducción.** Se expone el tema del documento, acompañado de la motivación, los objetivos planteados, las fases y métodos necesarios para su consecución, y la estructura del presente informe.

**Capítulo 2- Antecedentes y estado del arte.** En este capítulo se describen los distintos métodos de diagnóstico del TDAH y sus subtipos, centrándose específicamente en la detección mediante herramientas de *Deep Learning*.

**Capítulo 3- Aprendizaje profundo.** A lo largo de este capítulo, se presenta una descripción detallada sobre el aprendizaje profundo, destacando su importancia en el análisis de datos complejos. Se abordarán las redes neuronales, con especial énfasis en las redes neuronales convolucionales y los modelos basados en *Transformers*, que han demostrado ser herramientas clave en tareas de clasificación y detección. Además, se explicarán las técnicas de inteligencia artificial explicable (XAI) empleadas para comprender el funcionamiento interno de estos modelos.

**Capítulo 4- Métodos** En este capítulo se describen la metodología empleada y las etapas llevadas a cabo para el desarrollo el trabajo. Incluye la creación de los espectrogramas y su sectorización por subtipos; el diseño y entrenamiento de las redes utilizadas; la creación de los diferentes mapas de activación; el enmascaramiento de los espectrogramas y la obtención de sus características espectrales.

**Capítulo 5- Resultados y discusión** En este capítulo se expondrán los resultados obtenidos durante la realización del trabajo. Asimismo, se realizará una interpretación y discusión de los mismos.

**Capítulo 6- Conclusiones y líneas futuras** En el último apartado de este trabajo, se ofrece un resumen general del Trabajo de Fin de Máster, destacando las conclusiones alcanzadas. También se discuten las dificultades encontradas a lo largo del proceso y se proponen posibles líneas de investigación basadas en los resultados obtenidos.

## Capítulo 2

# ANTECEDENTES Y ESTADO DEL ARTE

## 2.1 INTRODUCCIÓN AL TDAH

---

El Trastorno por Déficit de Atención e Hiperactividad (TDAH) es un trastorno del neurodesarrollo que afecta a millones de personas en todo el mundo. Aunque fue identificado clínicamente en el siglo XX, los primeros registros históricos de comportamientos relacionados con el TDAH se remontan al siglo XVIII, cuando médicos como Alexander Crichton describieron por primera vez síntomas de inatención y conductas impulsivas en sus pacientes. No obstante, no fue hasta 1980, con la publicación del DSM-III [16], que el TDAH obtuvo un reconocimiento formal como trastorno diferenciado. Desde entonces, su comprensión y tratamiento han evolucionado significativamente gracias a los avances en la investigación neurocientífica y psicológica [17, 18].

El TDAH se caracteriza por un patrón persistente de inatención y/o hiperactividad-impulsividad que interfiere con el funcionamiento diario y el desarrollo. Los síntomas de inatención incluyen dificultades para mantener la concentración en tareas, cometer errores por descuido, no escuchar cuando se les habla directamente y problemas de organización. Por otro lado, los síntomas de hiperactividad-impulsividad abarcan inquietud motora, hablar de manera excesiva, dificultad para esperar y tendencia a interrumpir o invadir el espacio de otros [1].

En el ámbito educativo, el TDAH es una de las condiciones más comunes diagnosticadas en niños en edad escolar. Según [19], su prevalencia global se estima entre el 2.6 % y 4.5 % en menores de 18 años. Estos niños suelen enfrentar dificultades significativas en el aprendizaje y la adaptación al entorno escolar debido a la incapacidad para mantener la atención, controlar impulsos o gestionar la hiperactividad. Estas dificultades pueden derivar en problemas académicos, conflictos interpersonales y una menor autoestima.

La intervención temprana y el desarrollo de estrategias adecuadas son factores clave para mitigar los efectos del TDAH y potenciar las capacidades de las personas afectadas. Un enfoque integral que combine estrategias educativas, terapias conductuales y apoyo social puede facilitar la adaptación y el desarrollo pleno de las habilidades de estos individuos. El impacto del TDAH no se limita al individuo, sino que también afecta al entorno familiar, escolar y social. Por ello, resulta fundamental comprender a fondo este trastorno y garantizar un diagnóstico preciso para desarrollar intervenciones efectivas. Además, el continuo avance en la investigación científica

permite explorar nuevas perspectivas sobre sus causas y las estrategias de apoyo más eficaces.

### 2.1.1 SUBTIPOS DE TDAH

---

El Trastorno por Déficit de Atención e Hiperactividad (TDAH) se clasifica en tres subtipos principales según la predominancia de los síntomas, de acuerdo con los criterios de la quinta edición del *Diagnostic And Statistical Manual Of Mental Disorders* (DSM-V) [1]. Estos subtipos son los siguientes:

- **Predominantemente inatento** (TDAH-I): este subtipo se caracteriza por un predominio de dificultades en el mantenimiento de la atención y concentración.
- **Predominantemente hiperactivo/impulsivo** (TDAH-H): este subtipo está definido por un predominio de síntomas relacionados con la hiperactividad y la impulsividad.
- **Combinado** (TDAH-C): en este subtipo, se observan síntomas tanto de inatención como de hiperactividad e impulsividad.

La prevalencia de los diferentes subtipos de TDAH varía según las muestras estudiadas, los métodos de diagnóstico y las características demográficas. En general, el subtipo predominantemente inatento tiene una prevalencia algo menor en comparación con el subtipo combinado, que es el más frecuente en la población infantil diagnosticada con TDAH. Por otro lado, el subtipo predominantemente hiperactivo/impulsivo es menos común, representando una porción más pequeña de los casos en comparación con los otros dos subtipos [8, 20]

### 2.1.2 DIAGNÓSTICO TRADICIONAL DEL TDAH

---

Tradicionalmente, el diagnóstico del TDAH y su subtipo requieren de una evaluación completa que integra información clínica, observación del comportamiento y los criterios definidos en *Diagnostic and Statistical Manual of Mental Disorders* (DSM-V). Este proceso suele incluir entrevistas con el afectado, sus padres, docentes o profesionales de la salud, además de la recopilación de información sobre el sujeto, así como evaluaciones médicas y psicológicas [1].

El diagnóstico del TDAH y su subtipo, según los criterios del DSM-V, se realiza seleccionando los síntomas más relevantes de una lista de síntomas relacionados con la inatención y otros con la hiperactividad/impulsividad. Estos síntomas deben haber persistido durante al menos 6 meses, con un grado que no corresponde al nivel de desarrollo y que afecta negativamente a las actividades sociales y académicas. A continuación, en la tabla 2.1 se muestra de manera resumida estos síntomas de inatención e hiperactividad/impulsividad propuestos por el DSM-V [1].

Inatención	Hiperactividad/Impulsividad
Falla en prestar atención a detalles o comete errores.	Juguetea o golpea las manos o pies.
Dificultad para mantener la atención en tareas o actividades.	Se levanta en situaciones donde debe permanecer sentado.
Parece no escuchar cuando se le habla directamente.	Corretea o trepa en situaciones inapropiadas.
No sigue instrucciones ni termina tareas.	Incapaz de jugar o relajarse tranquilamente.
Dificultad para organizar tareas y actividades.	Está ocupado, con mucha energía.
Evita o muestra desgana en tareas que requieren esfuerzo mental.	Habla excesivamente.
Pierde cosas necesarias para tareas o actividades.	Responde antes de que se termine la pregunta.
Se distrae fácilmente por estímulos externos.	Le cuesta esperar su turno.
Olvida actividades cotidianas.	Interrumpe o se inmiscuye con otros.

TABLA 2.1: Síntomas de inatención e hiperactividad/impulsividad del TDAH

Una vez vistos los síntomas, si se detectan 6 o más síntomas de inatención, el sujeto será diagnosticado con TDAH predominantemente inatento (TDAH-I), mientras que si se detectan 6 o más síntomas de hiperactividad/impulsividad, se le diagnosticará TDAH predominantemente hiperactivo/impulsivo (TDAH-H). Por último, si se le detectan 6 o más síntomas tanto de inatención como de hiperactividad/impulsividad, será diagnosticado con TDAH combinado (TDAH-C) [1].

Para que el diagnóstico sea válido, los síntomas deben causar un deterioro significativo en el funcionamiento social, académico o laboral del individuo. Además, los síntomas deben haber estado presentes antes de los 12 años y en al menos dos entornos diferentes, por ejemplo, el hogar o la escuela.

## 2.2 DIAGNÓSTICOS OBJETIVOS

En este apartado, se presentan los principales métodos de diagnóstico objetivos del TDAH y sus subtipos, los cuales buscan complementar la evaluación clínica mediante el uso de herramientas basadas en mediciones estandarizadas y datos cuantificables.

### 2.2.1 RESONANCIA MAGNÉTICA

La **resonancia magnética (MRI)** es una técnica no invasiva de diagnóstico por imagen que utiliza campos magnéticos para capturar representaciones detalladas del interior del cuerpo humano. Esta herramienta permite visualizar en tres dimensiones estructuras como órganos y tejidos blandos, siendo especialmente útil para analizar el cerebro en el contexto de este trabajo.

De manera complementaria, la resonancia magnética funcional (fMRI) es una tecnología de neuroimagen que se enfoca en el estudio de la actividad cerebral. Además de ofrecer imágenes precisas de la estructura del cerebro, la fMRI detecta variaciones en el flujo sanguíneo y la concentración de oxígeno, indicadores directamente relacionados con la actividad neuronal. Esto hace posible observar el funcionamiento cerebral durante la realización de tareas específicas o en diferentes estados mentales.

Aunque la MRI no detecta el TDAH por sí sola, su capacidad para ofrecer imágenes detalladas de la anatomía cerebral ayuda a proporcionar información adicional en el proceso diagnóstico [21].

### 2.2.2 ELECTROENCEFALOGRAMA

---

El **electroencefalograma (EEG)** es una técnica utilizada para registrar la actividad eléctrica del cerebro mediante la medición de las señales generadas por las neuronas. Para realizar esta prueba, se colocan electrodos en el cuero cabelludo del paciente, que detectan y registran dichas señales. Estas mediciones se traducen en patrones gráficos que reflejan las ondas cerebrales y permiten analizar la actividad eléctrica del cerebro [22].

El EEG es una herramienta fundamental en el diagnóstico de diversas condiciones neurológicas, como la epilepsia, los trastornos del sueño, las encefalopatías, los tumores cerebrales y, en el contexto de este trabajo, el Trastorno por Déficit de Atención e Hiperactividad (TDAH).

Los **potenciales evocados** son una extensión del EEG que permite medir las respuestas eléctricas cerebrales ante estímulos específicos, como sonidos, luces o toques. Estos potenciales reflejan cómo el cerebro procesa la información sensorial y cognitiva. En el caso del TDAH, se utilizan principalmente los potenciales evocados cognitivos (como el P300 y N200), que son respuestas a estímulos cognitivos relacionados con la atención y la toma de decisiones [23].

Por consiguiente, tanto el EEG como los potenciales evocados son herramientas valiosas y complementarias para evaluar las alteraciones cerebrales en el TDAH, ofreciendo datos objetivos sobre la actividad eléctrica y el procesamiento cognitivo en el cerebro.

### 2.2.3 *Eye-tracking*

---

El ***eye tracking*** o seguimiento ocular es una técnica que permite registrar y analizar los movimientos oculares durante la observación de un estímulo visual. Utiliza cámaras infrarrojas de alta precisión para medir parámetros como fijaciones, movimientos sacádicos (*saccades*, movimientos rápidos del ojo) y patrones de exploración visual [24].

Esta tecnología ofrece información valiosa sobre cómo las personas procesan visualmente la información y mantienen su atención. En el caso del TDAH, los estudios han revelado patrones anómalos en los movimientos oculares, como mayor número de *saccades* o menor duración de fijaciones, que están asociados con problemas de inatención e impulsividad.

El *eye tracking* se emplea cada vez más como herramienta objetiva para evaluar y caracterizar la atención en personas con TDAH. Además, su integración con modelos de inteligencia artificial

permite la clasificación automática y precisa de este trastorno.

#### 2.2.4 ACTIMETRÍA

---

La **actimetría** es una técnica que mide de forma objetiva los patrones de actividad y descanso de una persona a lo largo del tiempo, utilizando dispositivos portátiles como pulseras o acelerómetros. Estos dispositivos registran el movimiento corporal durante periodos prolongados, generalmente días o semanas, permitiendo obtener información sobre los niveles de hiperactividad y los ciclos de sueño-vigilia [25].

A diferencia de técnicas más intrusivas como la resonancia magnética o el electroencefalograma, la actimetría resulta más accesible y menos molesta para el paciente. Además, es una opción más económica, lo que la convierte en una herramienta valiosa para el diagnóstico de diversas patologías y trastornos, proporcionando un diagnóstico objetivo sin necesidad de procedimientos invasivos y a un coste considerablemente más bajo [25, 26].

La actimetría se ha consolidado como una herramienta valiosa en el diagnóstico del TDAH, destacando por su accesibilidad y coste frente a técnicas como la resonancia magnética o el EEG. Su eficacia ha mejorado gracias a los avances en aprendizaje automático, como se muestra en [3, 7] donde se emplea en el diagnóstico del TDAH y en [4, 5], donde se utiliza para diferenciar entre los distintos subtipos. Por estas razones, se ha elegido la actimetría como método para identificar el subtipo hiperactivo de TDAH en este trabajo.

#### 2.2.5 *Machine Learning*

---

El aprendizaje automático (*Machine Learning*) se ha utilizado como herramienta para el diagnóstico y caracterización del TDAH a partir de datos neurofisiológicos y de neuroimagen. Varios estudios han demostrado su eficacia en la extracción de patrones relevantes y la clasificación de subtipos de TDAH.

Mao et al. [27] y Kuang et al. [28] aplicaron redes profundas y convolucionales a datos de fMRI para mejorar la precisión en la clasificación de TDAH y sus subtipos. De manera similar, Ahmadi et al. [29] y Vahid et al. [30] emplearon señales EEG y redes neuronales convolucionales, alcanzando altos niveles de precisión en la diferenciación entre niños con TDAH y controles. Además, Tenev et al. [31] y Yamashita et al. [32] mostraron que los enfoques supervisados y no supervisados permiten identificar subtipos de TDAH y explorar correlaciones con características cognitivas y cerebrales.

En conjunto, estos estudios evidencian que las técnicas de *Machine Learning* facilitan tanto la clasificación como la detección de patrones complejos, aportando información valiosa sobre la heterogeneidad del TDAH.

## 2.3 ESTADO DEL ARTE: DIAGNÓSTICO DEL TDAH MEDIANTE *Deep Learning* Y ACTIMETRÍA

El uso combinado de actimetría y *Deep Learning* resulta especialmente adecuado debido a que la actimetría proporciona datos objetivos y continuos sobre la actividad motora, mientras que las redes profundas permiten capturar patrones complejos y sutiles en estos datos. Esta sinergia facilita no solo la clasificación precisa de individuos con TDAH, sino también la identificación de diferencias entre subtipos, ofreciendo una herramienta prometedora para la evaluación clínica y la investigación del trastorno.

En esta sección se presentan las publicaciones más recientes que combinan técnicas de *Deep Learning* con la actimetría para el diagnóstico del TDAH. El objetivo es contextualizar los avances recientes en esta línea de investigación y destacar las contribuciones más relevantes en la literatura.

Para ello, se revisan estudios que utilizan actimetría como fuente de datos y *Deep Learning* como metodología de análisis, enfocándose en cómo esta combinación permite mejorar la precisión del diagnóstico y caracterizar los subtipos de TDAH.

### 2.3.1 ESTUDIOS PREVIOS QUE UTILIZAN LA ACTIMETRÍA Y *Deep Learning*

En este caso, en la tabla 2.3 se presentan los estudios más relevantes que utilizan tanto la actimetría como *Deep Learning* en el estudio del TDAH y sus tipos.

Artículo	Resumen
Amado-Caballero [3]	Este estudio presenta un sistema experto para el diagnóstico rápido, fiable y no intrusivo del TDAH infantil, combinando registros de actigrafía con técnicas de aprendizaje profundo. Los datos se procesan en ventanas temporales y se transforman en espectrogramas que sirven como entrada a una red neuronal convolucional. A partir de la clasificación de las ventanas, se determina el diagnóstico final del paciente. El enfoque busca superar la subjetividad del diagnóstico clínico tradicional mediante el uso de <i>hardware</i> moderno y análisis automatizado.
Amado-Caballero et al. [2]	Se propone un método automático para diagnosticar el TDAH combinado en niños y adolescentes utilizando registros de actividad de 24 horas y redes neuronales convolucionales. Los resultados muestran una sensibilidad promedio del 97.62 %, especificidad del 99.52 % y valores de AUC superiores al 99 %. El artículo demuestra que combinar técnicas de aprendizaje profundo con actimetría puede llevar a un sistema eficaz y confiable para el diagnóstico objetivo del TDAH de manera económica y no invasiva.

Amado-Caballero [7]	Este estudio combina registros de actigrafía y técnicas de aprendizaje profundo para analizar las diferencias en la manifestación del TDAH según la edad y el sexo, dada la mayor dificultad para diagnosticar la patología en mujeres y la atenuación de los síntomas con el tiempo. Los pacientes se agrupan por edad y género, y cada grupo es clasificado mediante una red neuronal convolucional específica. Para interpretar los resultados, se utilizan mapas de oclusión que permiten identificar las características relevantes de la señal asociadas al diagnóstico.
Amado-Caballero et al. [6]	En este artículo, se propone una metodología para interpretar los resultados de diagnóstico de TDAH en poblaciones divididas por edad y género basada en inteligencia artificial. Utiliza redes neuronales convolucionales para analizar registros de actividad de 24 horas y destacar patrones en tiempo y frecuencia a través de mapas de oclusión. Los resultados muestran diferencias significativas en los patrones de frecuencia entre pacientes con TDAH y controles. Estas técnicas de interpretación podrían mejorar la precisión diagnóstica y proporcionar una comprensión más profunda del TDAH.
Álvarez-Tavera [4]	Este estudio analiza el Trastorno por Déficit de Atención e Hiperactividad (TDAH), centrándose en la clasificación de los subtipos inatento y combinado mediante registros de actigrafía y redes neuronales convolucionales. Se han considerado grupos según edad y sexo, aplicando un análisis cualitativo con mapas de oclusión y un análisis cuantitativo basado en modelos gaussianos. Los resultados permiten identificar patrones diferenciales entre los subtipos. Además, se han determinado parámetros gaussianos estadísticamente relevantes para mejorar el diagnóstico.
Chico-Delgado [5]	Este estudio aborda la caracterización del TDAH de tipo hiperactivo, el menos diagnosticado y con menor información disponible. Se propone un sistema basado en actimetría y aprendizaje profundo para identificar patrones de actividad específicos. Se han generado espectrogramas a partir de señales actigráficas y se han entrenado redes neuronales convolucionales para su clasificación. Además, se aplican técnicas de interpretabilidad como mapas de oclusión y modelos de mezcla gaussiana. Los resultados permiten un análisis más preciso del subtipo hiperactivo, facilitando su diagnóstico.

Muñoz-Organero et al. [33]	En este estudio, se analizaron datos de acelerómetros en 22 niños, 11 con TDAH y 11 controles. Los niños usaron dos acelerómetros durante la escuela, situados en la muñeca y en el tobillo de la mano y pierna dominante respectivamente. Se convirtieron las señales en imágenes de aceleración 2D y se empleó una red neuronal convolucional (CNN) para distinguir entre los grupos. Los resultados mostraron diferencias significativas en la actividad de la muñeca en ambos grupos y en el tobillo solo en los niños no medicados. La CNN para extraer información acerca de los patrones de aceleración, se logró una precisión del 87.5 % en la muñeca y del 93.75 % en el tobillo.
Mahony et al. [34]	Este estudio propone el uso de sensores inerciales inalámbricos para diagnosticar el TDAH objetivamente. Estos sensores han sido colocados en la cintura y tobillo dominantes y miden el movimiento lineal y rotacional y durante una consulta psiquiátrica. Utilizando un SVM para clasificar sujetos como TDAH o no TDAH se logra una precisión del 95 % o más. Se concluyó que el movimiento registrado durante una tarea de concentración forzada proporciona una mejor clasificación que durante el tiempo libre.
Wiebe et al. [35]	Se explora el diagnóstico del TDAH en adultos mediante datos multimodales recogidos durante una tarea de atención en un entorno de realidad virtual, empleando un modelo de aprendizaje automático. Se utilizaron registros de actigrafía, seguimiento ocular, índices conductuales y EEG. El modelo, basado en máquinas de vectores soporte y entrenado con criterios de selección óptima de características, alcanzó una precisión del 81 % en un conjunto de prueba independiente. La actimetría, combinada con otras señales no invasivas, mostró un valor diagnóstico destacable, reforzando el papel del análisis automático del movimiento en entornos realistas como herramienta complementaria para el diagnóstico del TDAH.
Feretzakis et al. [36]	Se evalúa el uso de datos de actigrafía diaria junto con técnicas avanzadas de aprendizaje automático para la clasificación del TDAH. A partir de registros de 45 participantes, se extrajeron nuevas características relacionadas con patrones temporales, transiciones de actividad y ritmos circadianos. Entre los modelos evaluados, el de máquinas de vectores soporte obtuvo el mejor rendimiento (F1 score: 0.779). El análisis de importancia de características destacó la utilidad de las transiciones día-noche y las tasas de ráfagas de actividad como indicadores clave. A pesar del tamaño limitado de la muestra, los resultados respaldan la validez de la actimetría como herramienta diagnóstica objetiva y como base para futuras investigaciones con cohortes más amplias.

TABLA 2.2: Estudios que utilizan la actimetría y *Deep Learning*

Jiang et al. [37]	Este estudio analiza el uso de datos recogidos por dispositivos portables para el diagnóstico del TDAH en adolescentes. Se emplearon registros de actigrafía pasiva (aceleración del movimiento, frecuencia cardíaca y patrones de sueño) combinados con modelos de aprendizaje automático basados en XGBoost. A pesar del tamaño reducido de la muestra (30 participantes), los modelos que utilizaron exclusivamente medidas objetivas lograron una AUC de 0.844 para clasificar el TDAH y de 1.000 para predecir el uso de medicación. Estos resultados respaldan el potencial de la monitorización basada en actigrafía como herramienta no invasiva y precisa para apoyar el diagnóstico multidimensional del TDAH en adolescentes.
-------------------	--

TABLA 2.3: Estudios que utilizan la actimetría y *Deep Learning*

## Capítulo 3

# APRENDIZAJE PROFUNDO

El presente trabajo está centrado en el empleo de diversas técnicas de aprendizaje profundo, concretamente redes neuronales artificiales en el contexto del análisis de subtipos de TDAH. En el presente capítulo se busca abordar el tema del *Deep Learning* o aprendizaje profundo así como sus aplicaciones y técnicas. Además, se analizarán los modelos y arquitecturas específicas utilizadas y diferentes métodos de interpretación.

### 3.1 INTRODUCCIÓN AL *Deep Learning*

---

El *Deep Learning* (DL) o aprendizaje profundo es una técnica avanzada dentro del campo de la Inteligencia Artificial (IA) que utiliza redes neuronales artificiales con múltiples capas para modelar representaciones complejas de datos. Estas redes están diseñadas para simular la forma en que el cerebro humano procesa la información, permitiendo a las máquinas aprender de manera automática a partir de grandes volúmenes de datos. Esta capacidad de aprender por sí mismas permite a las máquinas realizar tareas complejas, como reconocer patrones en imágenes, interpretar texto o prever resultados a partir de datos históricos. A diferencia de otros métodos de aprendizaje automático, el *Deep Learning* se caracteriza por su capacidad para abordar problemas en los que las relaciones subyacentes entre los datos son extremadamente complejas o difíciles de definir de forma explícita [38].

El *Deep Learning* ha sido fundamental para el avance de muchas aplicaciones actuales, como la visión por computadora, el procesamiento de lenguaje natural o la conducción autónoma, entre otras. La clave del éxito del *Deep Learning* radica en su capacidad para extraer automáticamente características relevantes de los datos, lo que mejora la precisión y eficiencia de los modelos. Los algoritmos de *Deep Learning* pueden ser entrenados en grandes volúmenes de datos y hacer uso de potentes recursos computacionales, como las unidades de procesamiento gráfico (GPU), lo que facilita su aplicación en entornos prácticos.

### 3.1.1 JERARQUÍAS DE LA INTELIGENCIA ARTIFICIAL

---

Como hemos visto, el *Deep Learning* consiste en una subdisciplina de la Inteligencia Artificial. Esto se debe a que la IA se clasifica en varias ramas, de las cuales el *Machine Learning* (ML) y el *Deep Learning* son las más destacadas. La jerarquía que se establece entre ellas depende de la complejidad de las tareas que pueden abordar ya que cada uno de estos conceptos tiene su propio enfoque y propósito dentro del ámbito de la computación [39, 40]:

- **Inteligencia Artificial (IA)**: es el campo más amplio que engloba todas las técnicas y métodos que intentan simular o emular la inteligencia humana en las máquinas. La Inteligencia Artificial incluye aquellos sistemas que realizan tareas que normalmente requieren procesos de razonamiento humanos, como la percepción, el razonamiento, la toma de decisiones y el aprendizaje. De hecho, la IA ha revolucionado completamente nuestra sociedad, mediante aplicaciones en diferentes campos como la robótica, el procesamiento de lenguaje natural o el aprendizaje automático, entre otros.
- ***Machine Learning* (ML)** o aprendizaje automático: es un subcampo de la Inteligencia Artificial (IA) que se enfoca en el desarrollo de algoritmos y modelos capaces de aprender de los datos sin la necesidad de ser explícitamente programados para realizar tareas específicas. El *Machine Learning* (ML) utiliza grandes volúmenes de datos para entrenar modelos que identifican patrones y relaciones subyacentes. Esto permite que las máquinas realicen predicciones, clasificaciones y tomen decisiones sin depender de reglas predefinidas, imitando la capacidad humana de aprender de la experiencia. Su capacidad de adaptarse y mejorar con los datos lo convierte en una herramienta poderosa para diversas aplicaciones, desde el análisis de datos hasta la toma de decisiones automatizadas.
- ***Deep Learning* (DL)** o aprendizaje profundo: es una rama del *Machine Learning* cuyo propósito es imitar el procesamiento de información y el pensamiento humano mediante redes neuronales profundas. De esta manera, el *Deep Learning* es capaz de aprender representaciones más complejas de los datos, lo que le permite abordar tareas más abstractas. Destaca especialmente en tareas como el reconocimiento de imágenes, la comprensión del lenguaje natural y el procesamiento de voz, que son simples para los seres humanos, pero extremadamente difíciles para las máquinas. Su capacidad para aprender de grandes volúmenes de datos no estructurados ha transformado diversos campos de la Inteligencia Artificial, permitiendo a las máquinas realizar tareas anteriormente exclusivas del ser humano.

En la figura 3.1 se muestra la relación que hay entre los 3 conceptos vistos [41]:



FIGURA 3.1: Jerarquía entre IA, ML y DL

Aunque estos conceptos están relacionados, tienen diferencias clave. La IA abarca un conjunto amplio de técnicas para emular la inteligencia humana. Por otro lado, el *Machine Learning* permite a las máquinas aprender de los datos sin intervención humana. Mientras que el *Deep Learning* es una técnica avanzada que utiliza algoritmos más complejos para abordar problemas sofisticados, especialmente con grandes volúmenes de datos no estructurados, y requiere mayor capacidad computacional. En la tabla 3.1 se muestra a modo de resumen las diferencias entre IA, ML y DL:

Característica	Inteligencia Artificial	<i>Machine Learning</i>	<i>Deep Learning</i>
Alcance	Campo general	Subcampo de la IA	Subcampo del ML
Enfoque	Imitar la inteligencia humana	Aprender de los datos	Aprender de grandes datos complejos
Métodos	Reglas, algoritmos, heurísticas	Algoritmos estadísticos	Redes neuronales profundas
Datos requeridos	Indiferente	Grandes cantidades de datos	Grandes volúmenes de datos no estructurados
Complejidad	Varía según el enfoque	Moderada	Alta, mayor capacidad computacional
Aplicaciones	Robótica, juegos, sistemas expertos	Clasificar, predecir, reconocer patrones	Clasificación de imágenes, NLP, traducción

TABLA 3.1: Diferencias entre IA, ML y DL

En el contexto del trabajo, una de las diferencias más importantes entre *Machine Learning* y *Deep Learning* radica en el proceso de extracción de características. En los métodos tradicionales

de Machine Learning, la eficacia del modelo depende en gran medida de una buena selección y diseño manual de características relevantes, lo cual requiere conocimiento experto del dominio del problema. Por otro lado, el Deep Learning es capaz de aprender representaciones complejas directamente a partir de los datos en bruto. Esto simplifica el trabajo de diseño de los modelos y mejora su capacidad para adaptarse a datos complejos [38, 42].

### 3.1.2 REDES NEURONALES

Las **redes neuronales artificiales** son un modelo de cálculo inspirado en la estructura y funcionamiento del cerebro humano. Estas redes se componen de elementos interconectados llamados neuronas, que trabajan conjuntamente para procesar información y resolver problemas complejos como clasificación, reconocimiento de patrones o predicción.

#### 3.1.2.1 NEURONA ARTIFICIAL

Uno de los modelos más simples es el **perceptrón**, que fue el precursor de las redes neuronales modernas y está diseñado para realizar tareas de clasificación binaria mediante combinaciones lineales de sus entradas.

La neurona artificial o perceptrón, propuesto por Frank Rosenblatt en 1958, es la unidad o bloque básico de una red neuronal y establece las bases para redes neuronales más complejas. Su diseño está inspirado en la neurona biológica del cerebro humano. El perceptrón: recibe múltiples entradas, las pondera mediante pesos asignados, aplica una función de activación y genera una salida binaria. Así, la neurona, como bloque, tiene un estado interno que puede cambiar en función de las entradas que recibe. [43, 44, 45].

En la figura 3.2, se representa la arquitectura de una neurona artificial [46]:

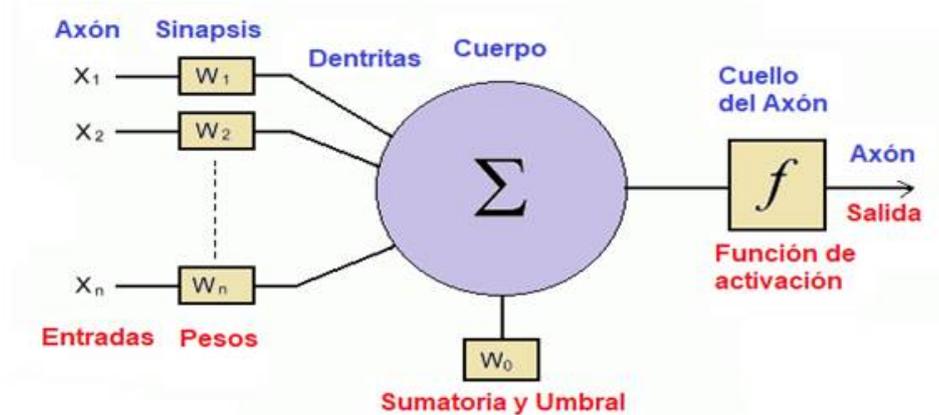


FIGURA 3.2: Estructura de una neurona artificial [46]

Como vemos, los elementos principales que conforman la arquitectura de la neurona artificial son los siguientes:

- **Entradas** ( $x_1, x_2, \dots, x_n$ ): datos que recibe como entrada la neurona.
- **Pesos** ( $w_1, w_2, \dots, w_n$ ): valores que asignan un nivel de importancia a cada entrada, ponderando su contribución.
- **Sumatorio ponderado** ( $\sum w_i x_i$ ): operación que combina las entradas multiplicadas por sus respectivos pesos.
- **Umbral o bias** ( $w_0$ ): valor constante que se suma al cálculo del sumatorio ponderado para ajustar el resultado.
- **Función de activación** ( $f$ ): procesa el valor neto para introducir no linealidad y determina la salida.
- **Salida**: resultado generado por la neurona, que puede servir como entrada para otras neuronas o como salida final del sistema.

Matemáticamente, el funcionamiento del perceptrón o neurona artificial puede expresarse como [43]:

$$y = f\left(\sum_{i=1}^n w_i x_i + w_0\right) \quad (3.1)$$

Las funciones de activación son esenciales en las redes neuronales porque introducen no linealidad, permitiendo que las redes aprendan patrones complejos. Sin ellas, la red solo podría realizar combinaciones lineales de entradas y salidas, limitando su capacidad. Estas funciones determinan si una neurona se activa, influenciando el rendimiento de la red. En la figura 3.3, se muestran las funciones de activación más comunes, siendo estas la función lineal, escalón, sigmoide y ReLU (*Rectified Linear Unit*) [47, 48]:

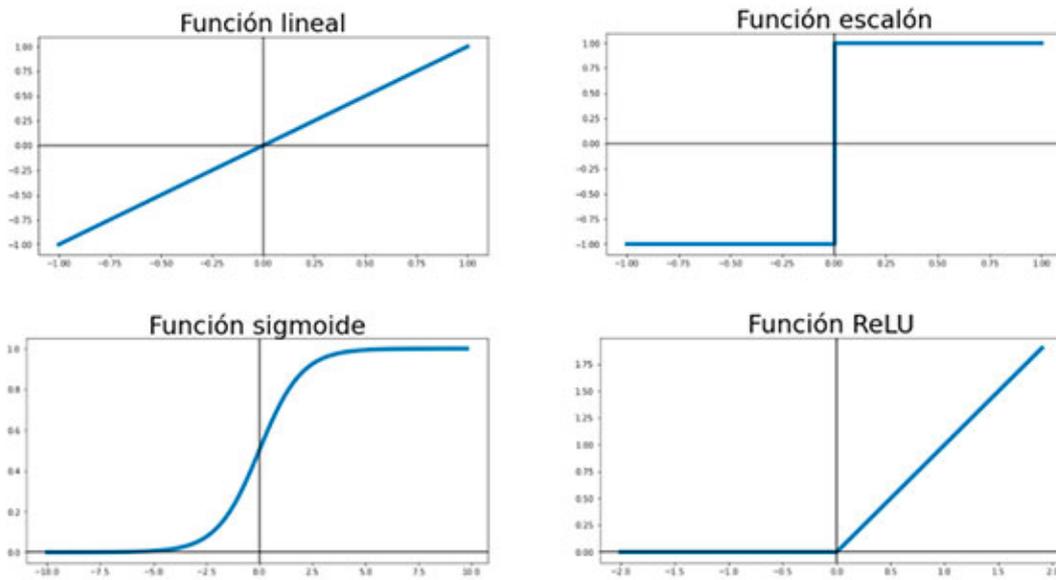


FIGURA 3.3: Funciones de activación [48]

### 3.1.2.2 REDES NEURONALES MULTICAPA

Una vez vista la neurona artificial, estas neuronas son agrupadas formando redes neuronales. Las **redes neuronales** se construyen combinando múltiples neuronas artificiales en una estructura jerárquica, donde cada neurona se conecta con otras, formando capas intermedias entre la entrada y la salida. Esta organización permite que las redes neuronales sean mucho más poderosas que un solo perceptrón, ya que pueden aprender representaciones más complejas de los datos. Al combinar varias neuronas, cada capa en una red neuronal realiza una parte del procesamiento y, en conjunto, permiten la resolución de problemas complejos que no podrían ser resueltos por una sola neurona [43]. En la figura 3.4 se muestra un ejemplo de red neuronal multicapa [49]:

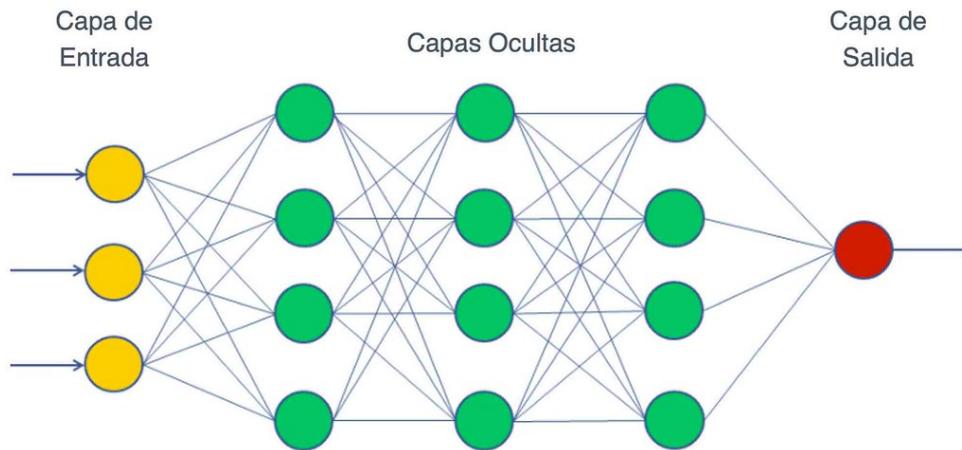


FIGURA 3.4: Funciones de activación [49]

Como se puede ver en la figura 3.4, una red neuronal multicapa está compuesta por tres tipos de capas principales [50]:

- **Capa de entrada:** capa que recibe los datos iniciales del sistema.
- **Capas ocultas:** capas intermedias que realizan el procesamiento computacional de los datos. Cada capa oculta recibe las salidas de la capa anterior y pasa esa información a la siguiente capa. Las capas ocultas son responsables de aprender representaciones más abstractas de los datos, y cuanto mayor sea el número de estas capas, mayor será la capacidad de la red para aprender relaciones complejas.
- **Capa de salida:** última capa de la red neuronal y proporciona el resultado final del modelo.

Estas capas están conectadas por pesos que determinan la intensidad de la influencia de una neurona sobre la siguiente. El proceso de aprendizaje de la red neuronal implica ajustar estos pesos para minimizar el error en la salida.

El uso de varias capas y neuronas, junto con funciones de activación no lineales, permite que las redes neuronales resuelvan tareas mucho más complejas que las que un perceptrón simple podría abordar. Este enfoque de redes neuronales profundas, también conocido como Deep Learning, es la base de muchos avances recientes en Inteligencia Artificial, como el reconocimiento de voz, imágenes y otros patrones complejos.

### 3.1.2.3 ENTRENAMIENTO DE REDES NEURONALES

El **entrenamiento** de redes neuronales es el proceso mediante el cual la red aprende a partir de un conjunto de datos. Durante este proceso, la red ajusta sus parámetros internos (pesos) para minimizar el error entre sus predicciones y los resultados esperados. El entrenamiento se realiza a través de ciclos iterativos, donde los datos de entrada se pasan por las capas de la red y se calcula una salida. Esta salida se compara con la salida deseada, y se ajustan los pesos para

reducir la diferencia. Este ajuste de pesos se lleva a cabo mediante algoritmos de optimización, siendo el más común el gradiente descendente.

El **gradiente descendente** es un algoritmo que ajusta los pesos de una red neuronal para minimizar la función de pérdida, es decir, el error entre la salida predicha y la salida esperada. Esto se logra calculando el gradiente de la función de pérdida respecto a los pesos y actualizándolos en la dirección opuesta al gradiente para reducir el error. El tamaño de las actualizaciones depende de la tasa de aprendizaje, un hiperparámetro que controla la magnitud de los pasos. Este proceso se repite durante varias iteraciones, permitiendo que la red aprenda los patrones en los datos [51].

Matemáticamente, el ajuste de los pesos en el gradiente descendente se expresa como:

$$\theta_{t+1} = \theta_t - \alpha_t \nabla f(\theta_t). \quad (3.2)$$

Donde  $\theta_t$  es el peso en la iteración  $t$ ,  $\alpha$  es la tasa de aprendizaje en la iteración  $t$  y  $\nabla f(\theta_t)$  es el gradiente de la función de objetivo, esto es, la función de pérdida que se quiere minimizar, con respecto a los pesos  $\theta_t$ .

Con el objetivo de entrenar adecuadamente una red neuronal, se realiza una **división de los datos** en conjuntos [52]:

- Conjunto de **entrenamiento**: Es el conjunto de datos utilizado para entrenar la red. La red aprende a partir de estos datos, ajustando sus pesos en cada iteración.
- Conjunto de **validación**: Se utiliza para evaluar el rendimiento de la red durante el entrenamiento y ajustar los hiperparámetros, como la tasa de aprendizaje o la estructura de la red, sin que la red haya visto estos datos durante el entrenamiento. Esto ayuda a prevenir el overfitting.
- Conjunto de **test**: Es un conjunto de datos separado que la red no ve durante el entrenamiento ni la validación. Se utiliza para evaluar el rendimiento final del modelo después de entrenarlo, proporcionando una estimación realista de cómo se comportará la red con datos nuevos e inéditos.

Uno de los principales desafíos durante el entrenamiento es el **overfitting** o sobreajuste. Este fenómeno ocurre cuando la red aprende demasiado bien los detalles del conjunto de datos de entrenamiento, en lugar de generalizar patrones que puedan aplicarse a datos no vistos. Como resultado, el modelo tiene un rendimiento excelente en los datos de entrenamiento pero falla en generalizar a nuevos datos, mostrando un mal desempeño en el conjunto de test. El sobreajuste puede ocurrir cuando el modelo es excesivamente complejo o cuando se entrena durante demasiado tiempo sin técnicas de regularización [53].

Para evitar el sobreajuste, se emplean varias técnicas, como la regularización, que penaliza pesos grandes; el *early stopping*, que detiene el entrenamiento cuando el rendimiento en el conjunto de validación deja de mejorar; el *dropout* que desactiva una fracción de neuronas durante el entrenamiento; o la validación cruzada, que ayuda a evaluar el modelo de manera más robusta y prevenir que se ajuste demasiado a un solo conjunto de datos.

La **validación cruzada** es una técnica que busca mejorar la evaluación del rendimiento de una red neuronal y ayuda a prevenir el overfitting. En validación cruzada, los datos se dividen en subconjuntos llamados pliegues o  *folds* . Dependiendo del método, estos  *folds*  se combinan de una manera u otra. A continuación se explican detalladamente los métodos de validación cruzada más comunes [54]:

- **Validación cruzada *K-fold***: uno de los métodos más utilizados, donde el conjunto de datos se divide en  $K$   *folds* . En cada iteración, uno de los  *folds*  se utiliza como conjunto de validación y el resto como conjunto de entrenamiento. Este proceso se repite  $K$  veces, asegurando que cada  *fold*  se utilice tanto para el entrenamiento como para la validación. En la figura 3.5 se muestra un ejemplo de validación cruzada *K-fold* [55]:

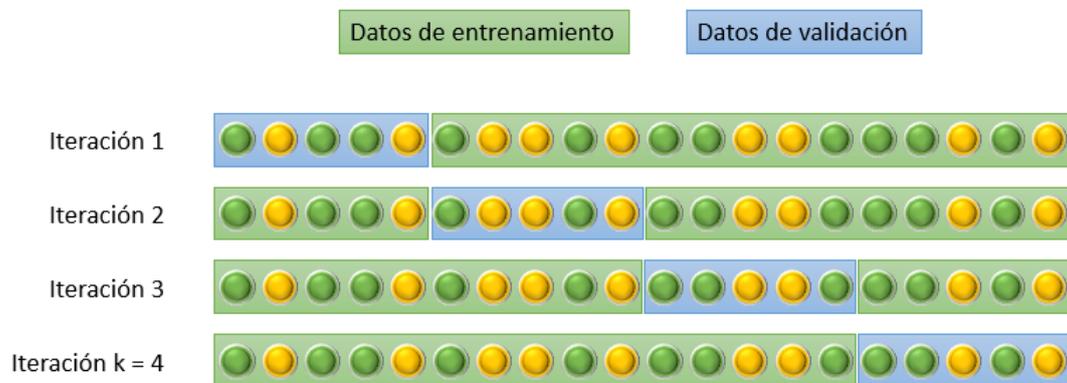


FIGURA 3.5: Funcionamiento de validación cruzada *K-fold*

- **Validación cruzada aleatoria**: los datos se dividen aleatoriamente en subconjuntos para crear el conjunto de entrenamiento y el conjunto de validación en varias iteraciones. Los subconjuntos no están predefinidos, lo que introduce una variabilidad adicional en los  *folds* . La ventaja de este método es que el número de iteraciones no depende del tamaño de los subconjuntos lo que puede ser útil para evaluaciones rápidas. Sin embargo, su desventaja es que no todas las muestras serán usadas en los conjuntos de entrenamiento o validación, y algunas podrían repetirse en varias iteraciones. En la figura 3.6 se muestra un ejemplo de validación cruzada aleatoria [56]:

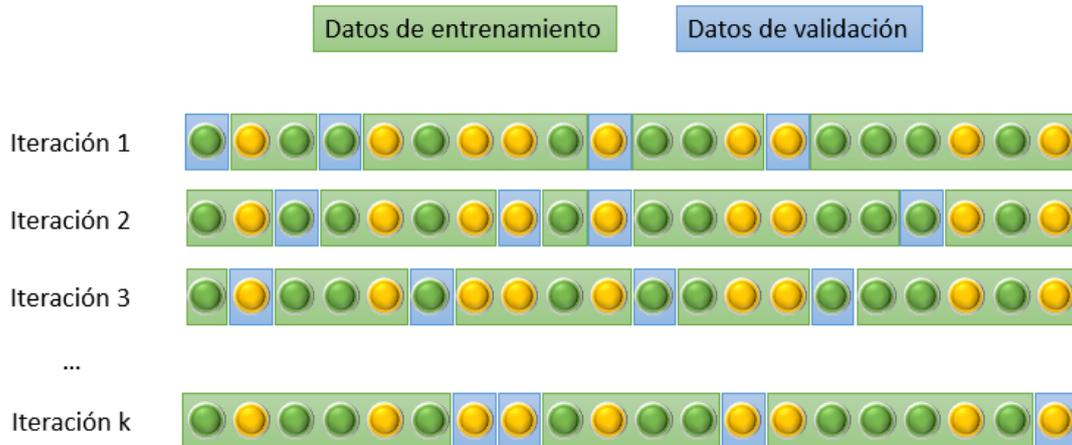
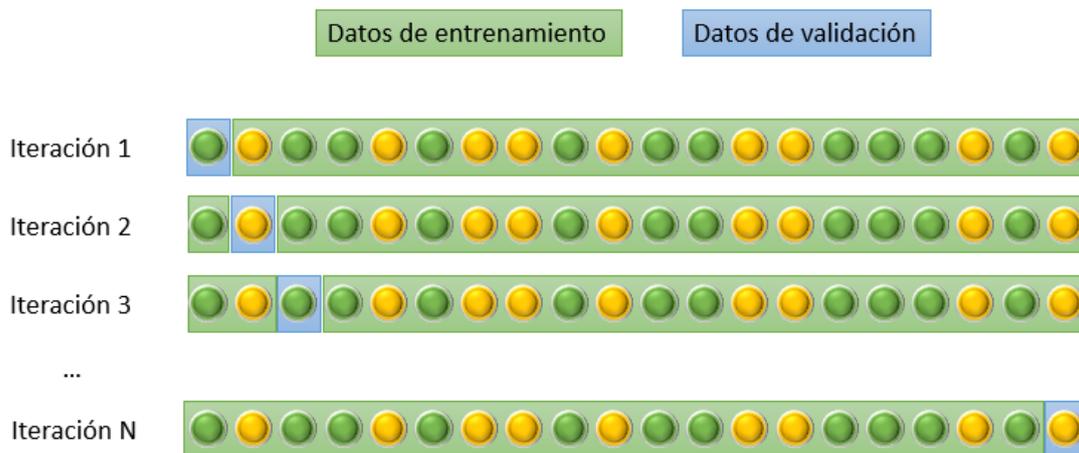


FIGURA 3.6: Funcionamiento de validación cruzada aleatoria

- Validación cruzada *Leave-One-Out* (LOO):** caso especial de la validación *K-fold*, donde  $K$  es igual al número de muestras en el conjunto de datos. Esto significa que en cada iteración, se utiliza una sola muestra como conjunto de validación y el resto como conjunto de entrenamiento. Este enfoque es útil para pequeños conjuntos de datos, aunque puede ser costoso computacionalmente para grandes conjuntos. En la figura 3.7 se muestra un ejemplo de validación cruzada *Leave-One-Out* [57]:

FIGURA 3.7: Funcionamiento de validación cruzada *Leave-One-Out*

- Validación cruzada *K-fold* estratificada:** método similar a la validación *K-fold*, pero con la diferencia de que garantiza que la proporción de clases en cada *fold* sea similar a la del conjunto original. Este método es especialmente útil cuando los datos están desbalanceados, ya que asegura que cada *fold* tenga una representación adecuada de todas las clases.

El uso de estos métodos ayuda a mejorar la generalización del modelo y a garantizar que la red neuronal no se sobreajuste a los datos de entrenamiento. Cada uno de ellos tiene sus

aplicaciones dependiendo del tipo de problema y los datos disponibles. La elección del método adecuado puede tener un impacto significativo en la evaluación del modelo y su capacidad de generalización.

#### 3.1.2.4 APRENDIZAJE DE REDES NEURONALES

El aprendizaje de redes neuronales es el proceso mediante el cual una red ajusta sus pesos y parámetros internos para mejorar su desempeño en una tarea específica, como clasificación, predicción o reconocimiento de patrones. La red aprende de los datos que se le proporcionan, ajustando gradualmente sus pesos para reducir la diferencia entre sus predicciones y los resultados esperados. Esto permite que las redes generalicen a datos no vistos, lo que las hace útiles en múltiples aplicaciones de Inteligencia Artificial.

Dependiendo del tipo de datos disponibles y del objetivo de la tarea, el aprendizaje de redes neuronales puede clasificarse en distintos paradigmas [42]:

- **Aprendizaje Supervisado:** la red se entrena con un conjunto de datos etiquetado, donde cada entrada tiene una salida conocida. El objetivo es aprender una función que asocie entradas con salidas correctas. Se utiliza en tareas como clasificación de imágenes, predicción de valores numéricos o reconocimiento de voz.
- **Aprendizaje No Supervisado:** la red trabaja con datos sin etiquetas y busca patrones o estructuras ocultas en ellos. Ejemplos de técnicas incluyen clustering, reducción de dimensionalidad y generación de representaciones compactas de los datos.
- **Aprendizaje por Refuerzo (Reinforcement Learning, RL):** la red aprende a través de la interacción con un entorno. No requiere datos etiquetados; en su lugar, recibe recompensas o penalizaciones según las acciones que realiza, aprendiendo a maximizar la recompensa acumulada. Este paradigma se emplea en robótica, juegos y sistemas de toma de decisiones autónoma.

Cada paradigma tiene aplicaciones y métodos específicos, y la elección de uno u otro depende de la naturaleza de la tarea y de la disponibilidad de datos etiquetados.

## 3.2 REDES PROFUNDAS

---

Como hemos visto, las **redes profundas** son la base de lo que se conoce como *Deep Learning*. Las redes profundas son un conjunto de modelos computacionales inspirados en el funcionamiento del cerebro humano, que han transformado la forma en que abordamos problemas complejos en diversas áreas de la Inteligencia Artificial. Estas redes están formadas por múltiples capas de neuronas artificiales, lo que les permite aprender representaciones de alto nivel a partir de grandes cantidades de datos. Cada capa sucesiva refina la información que recibe, permitiendo que el modelo capture patrones y características cada vez más abstractas.

En este apartado se presentarán los principales modelos de redes neuronales profundas, describiendo sus características, estructura y aplicaciones más relevantes.

### 3.2.1 REDES NEURONALES *feed-forward*

Las **redes neuronales *feed-forward*** (*feed-forward neural network*, FNN), son un tipo fundamental de red neuronal artificial en las que la información fluye en una única dirección. Esta va desde las entradas hacia las salidas, pasando por una o varias capas ocultas [43]. Además, este tipo de redes no presentan conexiones cíclicas. La estructura de una FNN puede observarse en la figura 3.8 [58].

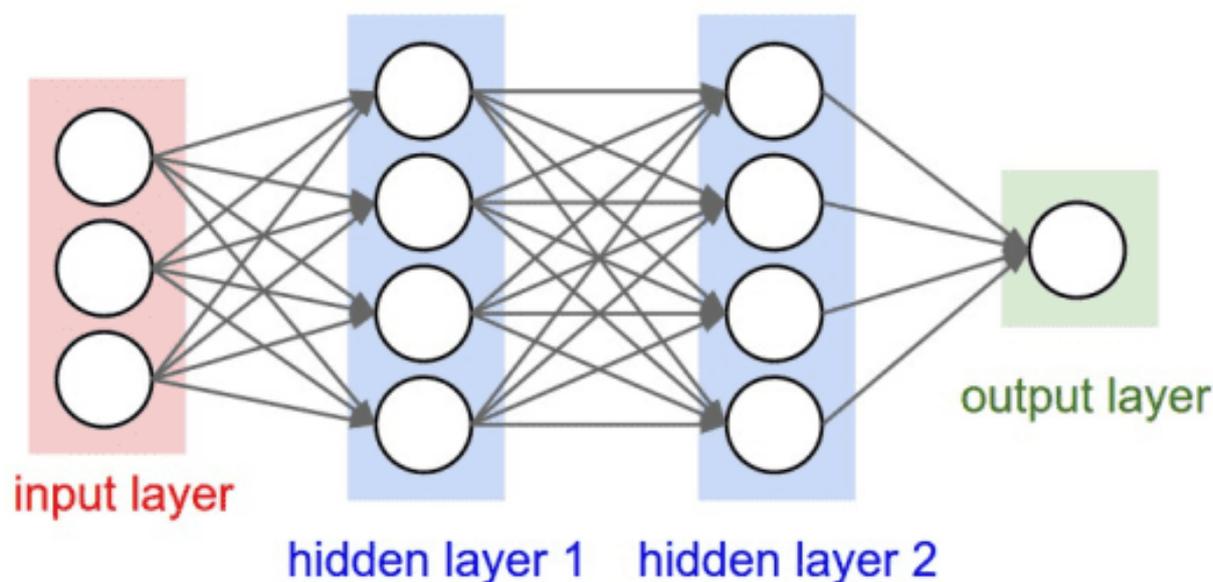


FIGURA 3.8: Estructura de una FNN [58]

Como puede verse, las redes *feed-forward* están compuestas por capas interconectadas, siendo estas: una capa de entrada, una o varias capas ocultas y una capa de salida. Cada neurona en una capa está conectada a todas las neuronas de la siguiente capa mediante pesos.

El entrenamiento de una FNN se realiza utilizando el algoritmo de *backpropagation* o retropropagación, que ajusta los pesos de las conexiones calculando el gradiente del error mediante la regla de la cadena. Este proceso consta de dos fases principales: una propagación hacia adelante, donde se calcula la salida de la red, y una propagación hacia atrás, donde el error se retropropaga para actualizar los pesos. Este enfoque iterativo permite a la red aprender con alta precisión [59].

Las FNN destacan por su capacidad para aproximar funciones complejas, su flexibilidad, escalabilidad y simpleza. Sin embargo, requieren datos etiquetados y pueden sobreajustarse si son demasiado complejas. Sus aplicaciones incluyen clasificación de imágenes y texto, regresión y detección de patrones.

### 3.2.2 REDES NEURONALES RECURRENTE

Las **redes neuronales recurrentes** (*recurrent neural network*, RNN) son un tipo de red neuronal diseñada para procesar datos secuenciales, como series temporales o texto. A diferencia de las redes *feed-forward* (FNN), las RNN poseen conexiones cíclicas que permiten que la información de pasos previos influya en el procesamiento actual. Esta característica las hace especialmente útiles para modelar dependencias temporales o contextuales en los datos [43].

La estructura de una RNN incluye una capa de entrada, una o más capas ocultas recurrentes y una capa de salida. En las capas recurrentes, cada neurona no solo recibe la entrada actual, sino también una retroalimentación de su estado previo, lo que crea un estado de memoria en la red. Esta memoria es clave para capturar patrones en datos secuenciales y puede extenderse en el tiempo dependiendo de la arquitectura de la red. En la figura 3.9 se muestra la estructura de una RNN [60]:

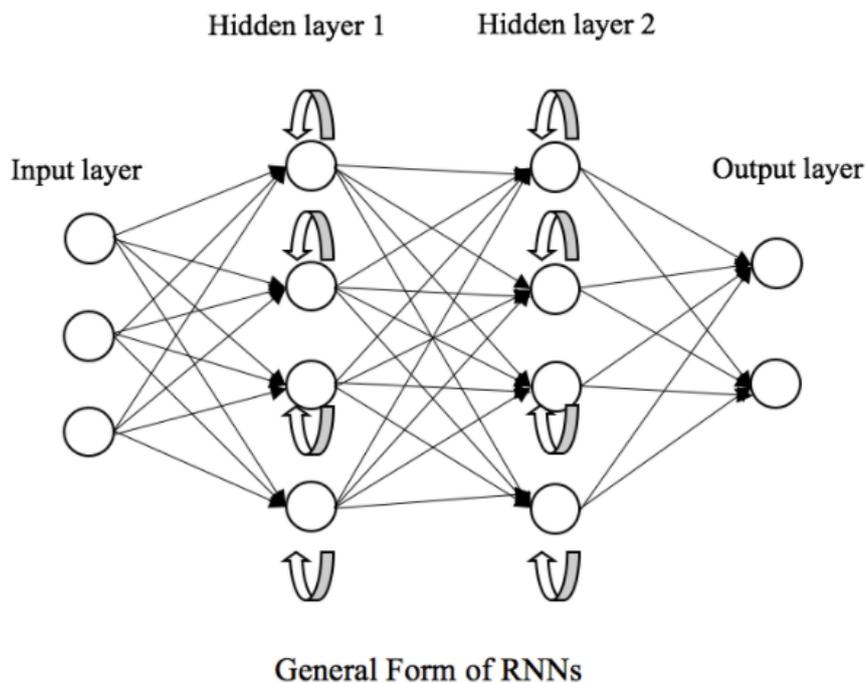


FIGURA 3.9: Estructura de una RNN [60]

El entrenamiento de las RNN también utiliza el algoritmo de *backpropagation*, adaptado para manejar secuencias en un proceso llamado *backpropagation through time* (BPTT). Este método ajusta los pesos considerando la influencia de los estados anteriores en el error actual [61].

Las RNN destacan por su capacidad para capturar relaciones temporales y contextuales. Entre sus aplicaciones más comunes se encuentran el reconocimiento de voz, la traducción automática, la generación de texto y el análisis de series temporales en finanzas y climatología. Su flexibilidad las convierte en una herramienta poderosa para abordar problemas secuenciales complejos.

### 3.2.3 REDES NEURONALES CONVOLUCIONALES

Las **redes neuronales convolucionales** (*Convolutional Neural Network, CNN*) son un tipo de red neuronal inspiradas en el sistema visual humano. Son especialmente eficaces en el procesamiento y análisis de datos estructurados en forma de matrices, esto es, imágenes y vídeos. Al igual que el ojo humano, las CNN destacan por su capacidad para identificar patrones locales y relaciones espaciales en los datos, basándose en la operación de convolución. Por lo tanto, las redes neuronales convolucionales son ampliamente utilizadas en tareas de detección y clasificación de imágenes [43].

En la figura 3.10, se muestra la estructura de una CNN [62]:

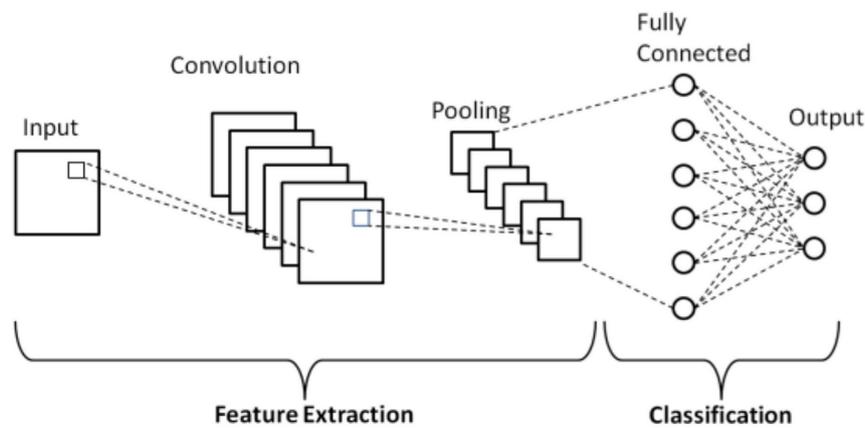


FIGURA 3.10: Estructura de una CNN [62]

Como puede observarse, las redes convolucionales están formadas por una serie de capas características [63]:

- Capa convolucional:** sirve para detectar características y patrones visuales en los datos de entrada. Utiliza filtros (*kernels*) que se aplican sobre pequeñas regiones de la imagen mediante una operación de convolución, que consiste en multiplicar y sumar los valores del filtro con los valores correspondientes de la imagen. Cada filtro extrae características específicas como bordes, texturas o formas. Como resultado, se generan mapas de características que indican qué patrones están presentes en diversas áreas de la imagen, permitiendo que la red identifique elementos relevantes para la tarea de clasificación o detección. En la figura 3.11, se representa el funcionamiento de la capa convolucional [64]:

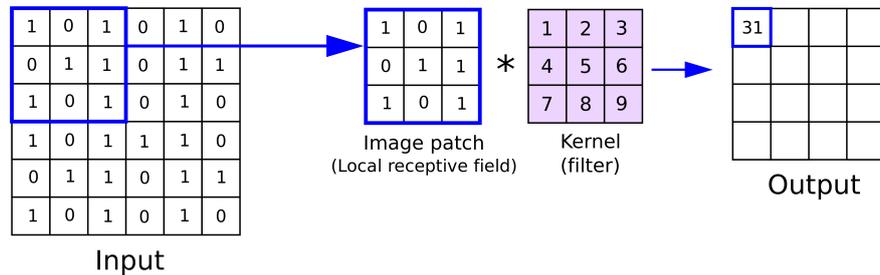
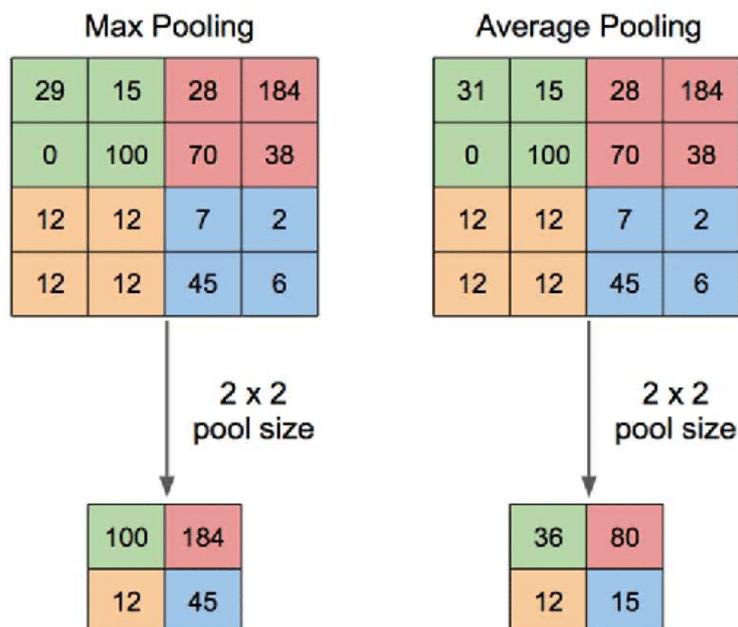


FIGURA 3.11: Funcionamiento de capa convolucional [64]

- Capa de *pooling*:** tiene como objetivo reducir la dimensionalidad de las características extraídas, mejorando la eficiencia computacional y ayudando a evitar el sobreajuste. Esta capa selecciona los valores más significativos dentro de pequeñas regiones de la imagen. Existen dos métodos principales para realizar esta operación, el *max pooling* en el que se toma el valor máximo de cada ventana, y el *average pooling* que calcula el valor promedio. Este proceso permite mantener las características más importantes y reduce la resolución de la imagen, lo que hace que la red sea más robusta y menos sensible a pequeños cambios en la imagen. En la figura 3.12, se muestra el funcionamiento de la capa *pooling*, representando tanto el método de *max pooling* como el de *average pooling* [65]:

FIGURA 3.12: Funcionamiento de capa *pooling* con los métodos *max pooling* y *average pooling* [65]

- Capa totalmente conectada o *fully connected*:** capa en la que cada neurona está conectada a todas las neuronas de la capa anterior, permitiendo que la red aprenda combinaciones

complejas de características extraídas previamente. Su función es transformar las activaciones de las capas anteriores en una representación más abstracta, útil para tareas como clasificación.

El entrenamiento de las CNN se realiza mediante el algoritmo de retropropagación y descenso de gradiente, ajustando los filtros y pesos para minimizar el error en las predicciones. Este proceso permite que las CNN aprendan eficientemente de grandes conjuntos de datos. Son ampliamente utilizadas en tareas de visión por computadora, como clasificación, detección de objetos y segmentación, así como en áreas como el análisis de imágenes médicas y la visión de vehículos autónomos. Su capacidad para aprender jerarquías de características las convierte en herramientas clave en el aprendizaje profundo.

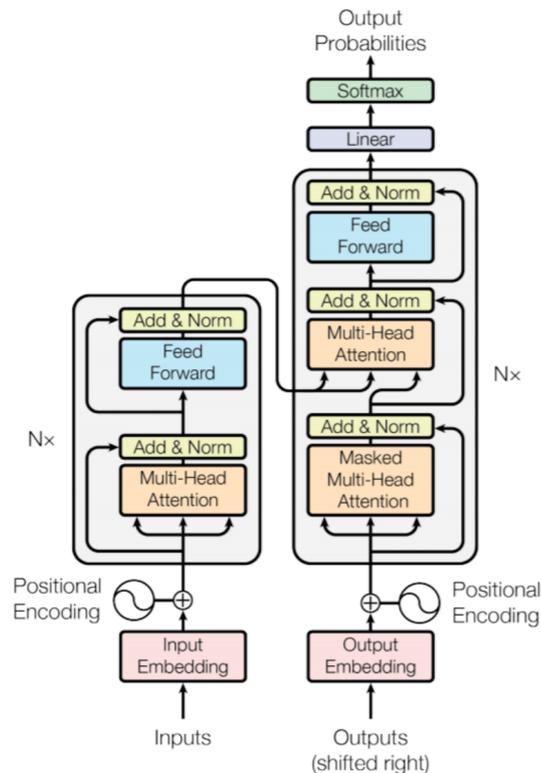
### 3.2.4 TRANSFORMERS

---

Los *Transformers* o transformadores son una arquitectura de aprendizaje profundo diseñada para abordar tareas que involucran secuencias de datos. Inicialmente desarrolladas para el procesamiento de lenguaje natural (NLP), como la traducción automática o la generación de texto, su capacidad de manejar dependencias a largo plazo sin requerir procesamiento secuencial las ha hecho muy eficaces y populares. A diferencia de las redes recurrentes, los *Transformers* no dependen de la secuencialidad para procesar la información, lo que les permite ser más eficientes en el manejo de largas secuencias y paralelizar mejor el entrenamiento. Una de las características fundamentales de los *Transformers* es cómo manejan la información, ya sea texto o imagen, esta se convierte en tokens. En NLP, las palabras o subpalabras se tokenizan para representar la entrada de manera más estructurada y que pueda ser procesada de manera eficiente [66].

La estructura de un *Transformer* se basa en un modelo codificador-decodificador (*encoder-decoder*). El codificador recibe la secuencia de entrada, mientras que el decodificador genera la salida. La clave de su funcionamiento es el **mecanismo de atención**, que permite que cada palabra o token en una secuencia se relacione con otros a través de un cálculo ponderado. Este proceso permite capturar dependencias a largo plazo de manera más eficiente que en las redes recurrentes.

En la figura 3.13 se muestra la estructura habitual de los *Transformers*, siendo el bloque de la izquierda el codificador y el de la derecha el decodificador [67]:

FIGURA 3.13: Estructura básica de *Transformers* [67]

A continuación se realiza una explicación de las capas presentes en un *Transformer* [68]:

- ***Input Embedding*** (incrustación de entrada): convierte las palabras o tokens de entrada en representaciones vectoriales que capturan su significado
- ***Positional Encoding*** (codificación posicional): agrega información sobre la posición de las palabras en la secuencia.
- ***Multi-Head Attention*** (atención multi-cabeza): permite que cada token se enfoque en otros tokens de la secuencia para capturar relaciones entre ellos. En el codificador, se trata de auto-atención, y en el decodificador, se emplea atención cruzada.
- ***Masked Multi-Head Attention*** (atención multi-cabeza enmascarada): versión enmascarada de la atención, que asegura que el modelo solo pueda "ver" los tokens previos y no los futuros, evitando que el modelo pueda mirar la respuesta completa en tareas como la generación de texto.
- ***Feed Forward***: consiste en una red FNN que realiza transformaciones no lineales a las representaciones de los tokens, aumentando la capacidad del modelo para aprender representaciones complejas.
- ***Add & Norm*** (suma y normalización): suma la entrada con la salida y aplica normalización para estabilizar el entrenamiento.

- **Output Embedding** (Incrustación de salida): convierte las representaciones generadas en el decodificador en una forma adecuada para la predicción, por ejemplo, probabilidades de tokens.
- **Linear & Softmax**: convierte las representaciones en un espacio de salida y luego aplica softmax para generar las probabilidades de los siguientes tokens en la secuencia.

El entrenamiento de los *Transformers* se divide en dos fases: preentrenamiento y *fine-tuning*. Durante el preentrenamiento, el modelo se entrena de manera autosupervisada en grandes volúmenes de datos no etiquetados, utilizando tareas como la predicción de palabras faltantes o la generación de texto. Posteriormente, en la fase de *fine-tuning*, el modelo se ajusta mediante un pequeño conjunto de datos etiquetados.

Las aplicaciones más relevantes de los *Transformers* incluyen la traducción automática, la generación de texto, la clasificación de texto y el análisis de sentimientos. Modelos como BERT (*Bidirectional Encoder Representations from Transformers*), GPT (*Generative Pre-trained Transformer*) o T5 (*Text-to-Text Transfer Transformer*) han demostrado un rendimiento sobresaliente en tareas de NLP, siendo utilizados en sistemas de recomendación, chatbots, motores de búsqueda y más [66].

#### 3.2.4.1 Vision transformer

Los *Vision Transformers* (ViT) o transformadores de visión son una variante de los *Transformers* diseñados para tareas de visión por computadora, como clasificación de imágenes, segmentación, etc. A diferencia de las CNN tradicionales que procesan las imágenes como una matriz de píxeles, los ViT dividen una imagen en parches no superpuestos y los tratan como secuencias, similar al procesamiento de texto en los *Transformers*. Los ViT son especialmente útiles en aplicaciones como la clasificación de imágenes, la segmentación y la detección de objetos [69].

En la figura 3.14 se puede observar el funcionamiento de un *Vision Transformer* [66]:

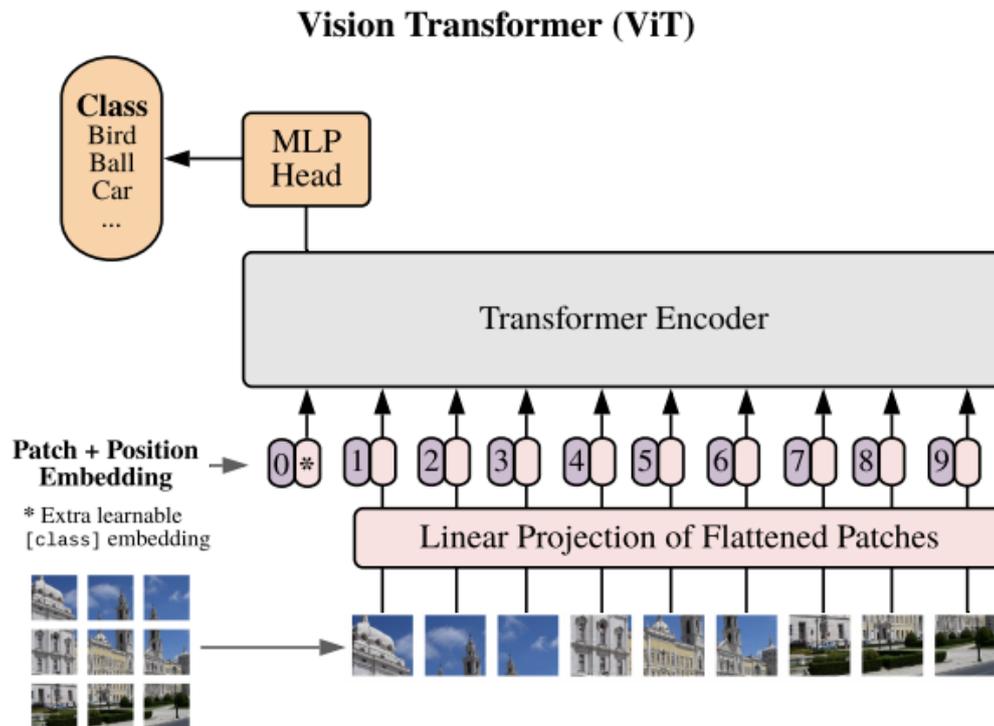


FIGURA 3.14: Funcionamiento de un *Vision Transformer* [66]

En su funcionamiento, los ViT dividen la imagen en pequeños parches no superpuestos, que son tratados como tokens. Cada parche se transforma en un vector de características mediante un proceso de embedding. Estos vectores se procesan luego a través de la arquitectura de *Transformers* estándar, donde se capturan relaciones espaciales entre los parches, permitiendo a la red comprender el contexto global de la imagen.

La principal diferencia con los *Transformers* tradicionales radica en que estos últimos procesan datos unidimensionales como texto, mientras que los ViT manejan imágenes, capturando relaciones espaciales en 2D entre los parches.

Los ViT se diferencian de las CNN en varios aspectos clave. Los ViT tienen menor eficiencia de datos, ya que requieren más datos para entrenarse de manera efectiva, aunque tienen mayor capacidad al poder manejar modelos más grandes. A diferencia de las CNN, que aplican filtros locales a toda la imagen y son más estables y eficientes en datos, los ViT utilizan autoatención, lo que les permite capturar relaciones globales entre parches y ser más robustos a distorsiones en las imágenes. Sin embargo, los ViT son más sensibles a la elección de hiperparámetros y requieren técnicas adicionales para mejorar su estabilidad [70].

---

### 3.3 INTELIGENCIA ARTIFICIAL EXPLICABLE

---

La Inteligencia Artificial explicable o XAI (*Explainable Artificial Intelligence*) es un campo dentro de la Inteligencia Artificial que busca hacer comprensibles y transparentes los modelos y algoritmos utilizados en sistemas de aprendizaje automático. Esto se logra desarrollando técnicas y herramientas que permiten a los usuarios comprender, interpretar y confiar en las decisiones que toma un sistema de IA [71].

La XAI busca abordar el problema de que las redes neuronales profundas actúan como cajas negras. Estos modelos generan resultados efectivos pero difíciles de interpretar, ya que no es intuitivo comprender los procesos internos que conducen a sus decisiones. El objetivo principal de la XAI es hacer transparentes estas decisiones, proporcionando herramientas que expliquen las razones detrás de las decisiones de las redes.

Gracias a estas explicaciones, la XAI mejora la confianza del usuario en los sistemas y facilita la detección de errores o sesgos en los modelos. Al mismo tiempo, contribuye a una mejor comprensión del funcionamiento de la Inteligencia Artificial, permitiendo integrar su uso de manera más responsable y eficaz.

---

#### 3.3.1 MÉTODOS DE XAI

---

En este apartado se van a explorar los diferentes métodos utilizados en la elaboración del presente trabajo a fin de proporcionar una visión más clara de los modelos empleados. Estos métodos consistirán en la creación de mapas de activación los cuales permiten ilustrar de manera visual las zonas más importantes de una imagen.

##### 3.3.1.1 MAPAS DE OCLUSIÓN

Los mapas de oclusión son una técnica utilizada en el campo de la interpretación de modelos de aprendizaje automático, especialmente en redes neuronales profundas, para visualizar y comprender cómo las características de entrada afectan a las decisiones de un modelo. Su principal objetivo es proporcionar información sobre las características más relevantes de la entrada (habitualmente una imagen) que influyen de manera significativa en la predicción del modelo.

Los mapas de oclusión se basan en la idea de ocultar partes de la imagen y observar cómo cambia la salida del modelo. Este proceso se repite de manera sistemática, donde se ocultan pequeñas regiones de la imagen de entrada, y se mide la disminución en el rendimiento del modelo. Las áreas que, al ser ocultadas, provocan una mayor disminución en el rendimiento son consideradas como las más importantes para la toma de decisiones del modelo [72, 73].

Esta técnica, permite observar cómo cada capa de la red interpreta la entrada al identificar las zonas más importantes para su procesamiento. Al ocultar diferentes regiones, se puede entender qué áreas de la imagen influyen más en las activaciones de cada capa y, en consecuencia, en la decisión final del modelo.

El proceso básico para la generación de los mapas de oclusión es el siguiente:

- Se selecciona la máscara y se oculta una sección de la imagen con la máscara.
- La imagen con la máscara es empleada como imagen de entrada en el modelo.
- Se evalúa el rendimiento del modelo con la imagen modificada. La diferencia en el rendimiento con respecto a la imagen original indica la importancia de la región oculta.
- Se repite este proceso moviendo la máscara por diferentes regiones de la imagen, generando un mapa visual que muestra qué áreas de la imagen son más relevantes para la predicción del modelo.

En la figura 3.15 se muestra de manera visual el proceso de creación de mapas de oclusión.

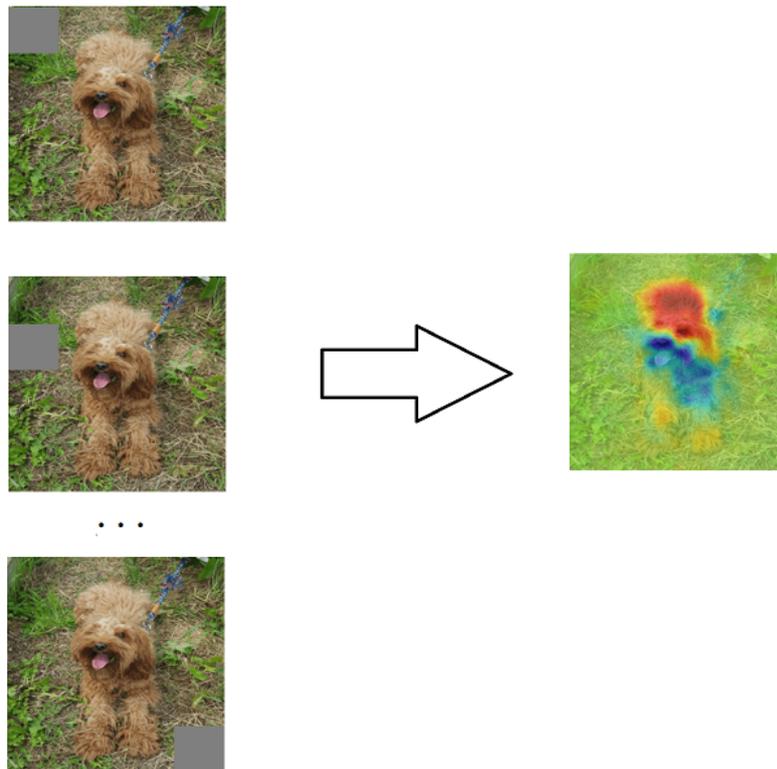


FIGURA 3.15: Proceso de generación de mapas de oclusión

Por lo tanto, los mapas de oclusión son útiles para interpretar los modelos de clasificación de imágenes y entender cómo los diferentes elementos visuales contribuyen a la toma de decisiones del modelo. Al resaltar las regiones de la imagen que tienen un mayor impacto en la predicción, se puede obtener una visión más clara de cómo el modelo procesa la información visual.

### 3.3.1.2 GRAD-CAM

Grad-CAM (*Gradient-weighted Class Activation Mapping*) es una técnica de visualización utilizada para interpretar y entender las decisiones de modelos de aprendizaje profundo, especialmente en redes neuronales convolucionales (CNN) aplicadas a tareas de clasificación de imágenes. A través de Grad-CAM, se pueden identificar las regiones de una imagen que más contribuyen a la predicción realizada por el modelo, proporcionando una explicación visual y fácil de interpretar de sus mecanismos internos.

Grad-CAM utiliza los gradientes de la clase objetivo con respecto a las activaciones de las últimas capas convolucionales del modelo para generar un mapa de activación. Este mapa destaca las áreas de la imagen que influyen de manera más significativa en la predicción del modelo, lo que permite entender qué características visuales son consideradas más importantes para la clasificación [74, 75].

El proceso básico de Grad-CAM consta de las siguientes etapas:

- **Forward Propagation:** se propaga la entrada a través de la red para obtener las predicciones y los mapas de características ( $A^k$ ) de la capa convolucional deseada (normalmente la última).
- **Calcular Gradientes:** se obtienen los gradientes de la puntuación de la clase objetivo ( $y^c$ ) con respecto a cada píxel de los mapas de características ( $A^k$ ).

$$\frac{\partial A_{i,j}^k}{\partial y^c} \quad (3.3)$$

Los gradientes indican cuán sensible es ( $y^c$ ) a las activaciones en ( $A^k$ ).

- **Calcular los pesos:** se promedian los gradientes de cada mapa de características para obtener un peso ( $w_k$ ) que mide su importancia para la clase objetivo.

$$w_k = \frac{1}{Z} \sum_{i,j} \frac{\partial A_{i,j}^k}{\partial y^c} \quad (3.4)$$

- **Generar mapa de activación:** se combinan los mapas de características usando los pesos:

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left( \sum_k w_k A^k \right) \quad (3.5)$$

La función ReLU asegura que solo se conserven las contribuciones positivas.

- **Ajustar y superponer:** se escala el mapa de activación al tamaño de la imagen y se superpone para destacar las regiones más relevantes.

En la figura 3.16 se puede observar el proceso al utilizar Grad-CAM [76]:

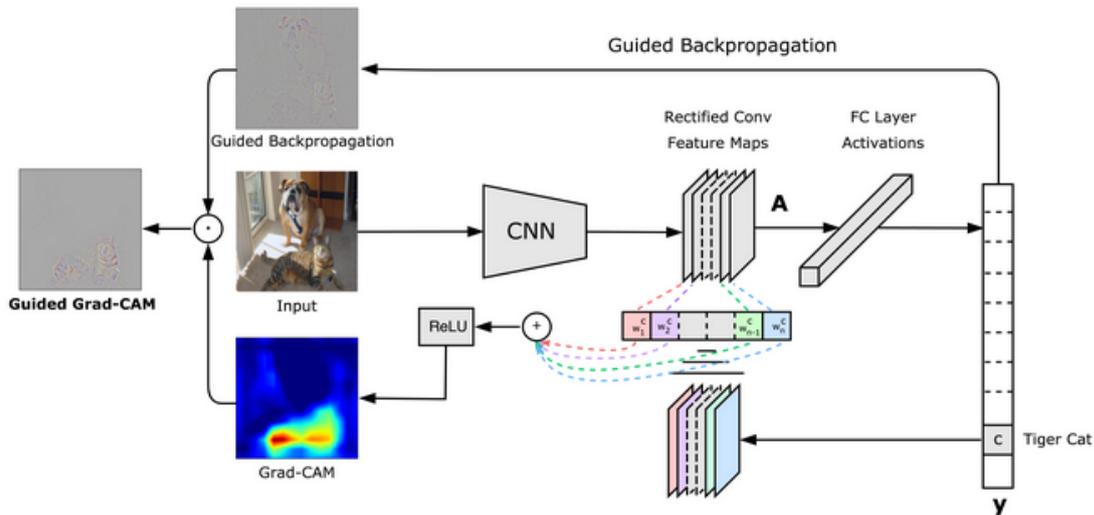


FIGURA 3.16: Proceso de Grad-CAM [76]

Grad-CAM es una técnica poderosa que permite comprender cómo las redes neuronales convolucionales realizan sus predicciones en el contexto de clasificación de imágenes. Al proporcionar una visualización clara de las regiones importantes para la decisión del modelo, ayuda a mejorar la interpretabilidad de los modelos de aprendizaje profundo, ofreciendo explicaciones fáciles de entender para los usuarios y desarrolladores. Además, al ser una técnica no invasiva, se puede aplicar a modelos preentrenados sin necesidad de modificar su arquitectura.

### 3.3.1.3 MECANISMOS DE ATENCIÓN

Los mecanismos de atención en Inteligencia Artificial son técnicas inspiradas en los procesos cognitivos humanos que permiten a una red neuronal centrarse en las partes más relevantes de una entrada. En lugar de procesar toda la información de manera uniforme, la atención asigna pesos a diferentes partes de la entrada, permitiendo que el modelo se enfoque en áreas específicas que contienen la información más importante para realizar una tarea [77].

En el contexto de redes neuronales que trabajan con imágenes, los mecanismos de atención asignan pesos a diferentes áreas de la imagen en función de su relevancia para la tarea de clasificación o detección de objetos. Este proceso permite al modelo enfocarse en las partes más importantes de la imagen, en lugar de procesar toda la información de manera uniforme. Los pesos se calculan a través de una serie de operaciones, como la comparación entre características de la imagen y representaciones internas del modelo, lo que guía al sistema a priorizar ciertas regiones.

Los mapas de atención resultantes permiten visualizar los pesos asignados a cada región de la imagen, destacando aquellas que más influyen en la decisión del modelo. Estos mapas se generan combinando de forma ponderada las representaciones intermedias obtenidas por la

red, utilizando coeficientes de atención aprendidos y se integran mapas de características de distintas capas asignando mayor peso a las regiones más activas o relevantes para la predicción. Durante el entrenamiento, estos pesos de atención se ajustan iterativamente, optimizando la capacidad del modelo para centrarse en las zonas más importantes de la imagen y mejorar así la precisión de sus predicciones [78].

Este enfoque se encuentra estrechamente relacionado con la arquitectura de los *Transformers*, donde la atención es el componente central. En concreto, los *Transformers* utilizan atención multi-cabeza (*multi-head attention*) para aprender relaciones entre diferentes partes de la entrada, ya sean palabras en texto o regiones de una imagen. Cada cabeza de atención puede capturar patrones distintos, lo que permite obtener mapas de atención ricos y variados. En el contexto visual, como en *Vision Transformers* (ViT), esta atención se aplica sobre parches de imagen, permitiendo al modelo integrar información contextual de forma global. Así, los mecanismos de atención no solo mejoran la precisión del modelo al enfocar el procesamiento en regiones clave, sino que también ofrecen explicaciones visuales claras de cómo se produce la toma de decisiones [79].

En resumen, los mecanismos de atención mejoran la precisión de las redes neuronales al enfocarse en las partes más importantes de las imágenes. Ofrecen una mayor interpretabilidad al generar mapas que permiten visualizar esas áreas clave, proporcionando una comprensión más profunda de cómo los modelos procesan la información visual.

## Capítulo 4

# METODOLOGÍA

El objetivo principal de este trabajo consiste en mejorar el sistema propuesto en [5], con la ayuda de nuevas herramientas avanzadas de inteligencia artificial explicable (XAI). El propósito de este era la identificación y diferenciación del subtipo hiperactivo de TDAH, mediante un sistema basado en *Deep Learning*. Para ello, se utilizarán datos de actimetría como base para diseñar y entrenar dos modelos de redes neuronales. A estos modelos se les aplicarán métodos de interpretabilidad de XAI, permitiendo no solo mejorar la precisión en la clasificación de los subtipos, sino también entender los patrones subyacentes que guían las predicciones del modelo.

A lo largo de este capítulo, se explican los métodos y procedimientos aplicados durante el desarrollo del trabajo, concretamente:

- Procesamiento y gestión de los datos de entrada, desde señales de actigrafía hasta su conversión en espectrogramas.
- Explicación de las redes neuronales utilizadas y de su proceso de entrenamiento
- Interpretación de los resultados ofrecidos por las redes neuronales mediante herramientas de XAI como el cálculo de mapas de activación y características espectrales.

En la figura 4.1 se muestra el flujo de trabajo seguido que a lo largo de este capítulo se explicará en detalle.

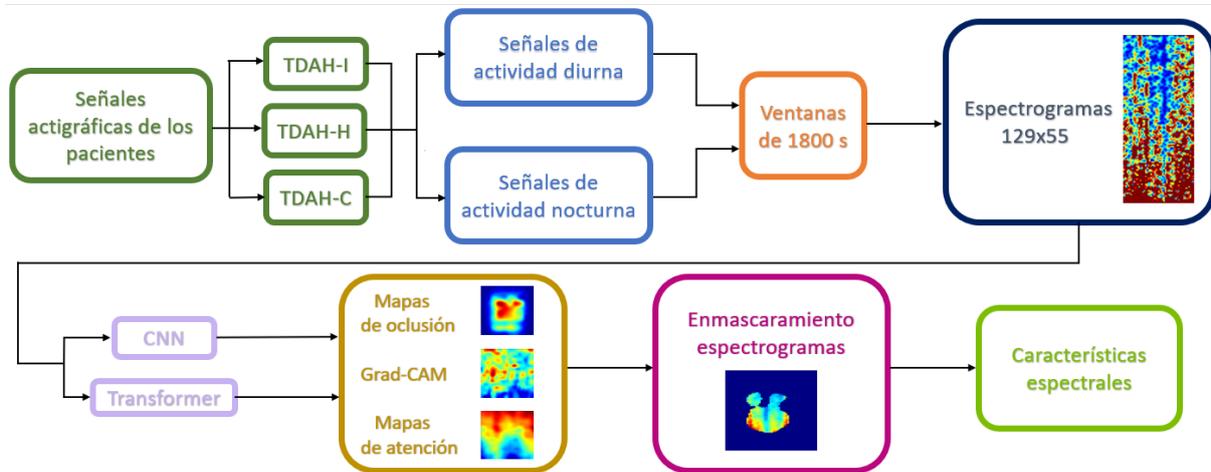


FIGURA 4.1: Flujo de trabajo de la sección

En comparación al trabajo realizado en [5], se ha ampliado el estudio, observándose las siguientes diferencias:

- Análisis en un nuevo modelo de red neuronal: *Transformer*.
- Creación de 3 tipos de mapas de activación (oclusión, Grad-CAM y atención), en lugar de solamente mapas de oclusión.
- Análisis de los mapas en las 3 capas convolucionales de la CNN en vez de únicamente en la tercera.
- Enmascaramiento de los espectrogramas a fin de resaltar las zonas más importantes de estos.
- Cálculo de características espectrales en lugar de parámetros estadísticos de las gaussianas.

## 4.1 PROCESAMIENTO DE LOS DATOS

En este apartado se detalla el procesamiento de los datos de entrada utilizados en el estudio. Se explica cómo las señales actigráficas 1D se convierten en imágenes (espectrogramas). Posteriormente, estas imágenes se organizan y dividen según los diferentes tipos estudiados, facilitando su clasificación y análisis comparativo.

### 4.1.1 SEÑALES DE ACTIMETRÍA

Tanto en este trabajo como en trabajos anteriores [3, 5] se parte de señales de actimetría o actigráficas como datos de entrada. Las señales actigráficas utilizadas en este trabajo son de

tipo unidimensional (1D) y se obtienen mediante un dispositivo actígrafo, el *ActiGraph GT3x*, colocado en la muñeca de los pacientes. Este dispositivo registra la aceleración en los tres ejes cartesianos durante un día completo, calculando la media de dichas aceleraciones. El actígrafo opera a una frecuencia de muestreo de 1 Hz, lo que equivale a registrar una muestra por segundo, generando un total de 86400 muestras diarias por paciente.

Para el análisis, las señales se segmentan en dos subconjuntos: actividad diurna y actividad nocturna. La delimitación de estos subconjuntos se realiza en función de los diferentes periodos de actividad, considerando que los pacientes muestran comportamientos distintos según el momento del día.

Posteriormente, cada subconjunto se divide en ventanas de 1800 segundos (30 minutos), con el objetivo de capturar la actividad asociada a tareas o comportamientos específicos durante intervalos definidos. Estudios previos [3, 4] exploraron la segmentación de las señales en ventanas de 60, 300 y 1800 segundos, representando actividades de corta, media y larga duración, respectivamente. Los resultados indicaron que las ventanas de 1800 segundos ofrecen un mejor rendimiento general. Por este motivo, en este trabajo se ha optado por utilizar únicamente ventanas de este tamaño.

#### 4.1.2 CREACIÓN DE ESPECTROGRAMAS

En el presente trabajo, se utiliza una red neuronal convolucional (CNN) y un *vision Transformer*. Como se vio en el apartado 3.2 estas redes están diseñadas para el análisis de imágenes. Por lo tanto, es necesario convertir las señales 1D de actimetría en imágenes que puedan ser analizadas por las redes. Para ello, las señales 1D son transformadas en espectrogramas.

Un **espectrograma** es una representación visual que muestra cómo la variación de la energía de las diferentes frecuencias de una señal en función del tiempo. Esta representación tridimensional utiliza el eje horizontal para indicar el tiempo, el eje vertical para las frecuencias, y los colores para representar la potencia de la señal. En nuestro caso, las potencias más bajas se muestran en tonos azules, mientras que las potencias más altas se destacan con colores cálidos.

Matemáticamente, un espectrograma se define como el cuadrado de la Transformada de Fourier de Tiempo Corto (*Short-Time Fourier Transform*, STFT) de la señal:

$$\mathbf{spectrogram}(t, \omega) = |STFT(x[n])|^2 \quad (4.1)$$

La STFT consiste en dividir la señal en segmentos cortos y aplicar la Transformada de Fourier a cada segmento, lo que permite obtener su representación en el dominio de la frecuencia. La fórmula discreta de la STFT es:

$$\mathbf{STFT}(x[n]) = X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n]w[n - m]e^{-j\omega n} \quad (4.2)$$

Donde  $x[n]$  es la señal discreta, y  $w[n - m]$  es la función de ventana que segmenta la señal.

Al crear espectrogramas en Matlab, es crucial seleccionar correctamente parámetros como el tamaño de la ventana, la frecuencia de muestreo, el tamaño del solapamiento (overlap) y el tamaño de la Transformada Discreta de Fourier (DFT). Un mayor solapamiento mejora la resolución temporal, mientras que un tamaño de DFT más grande incrementa la resolución frecuencial.

En este trabajo, se han empleado ventanas de 1800 segundos con una frecuencia de muestreo de 1 Hz. Además, se utilizó un solapamiento de 150 muestras y un tamaño de DFT de 256, lo que resultó en espectrogramas de 129x55 píxeles. Esto significa que cada espectrograma representa 129 frecuencias distintas y 55 instantes de tiempo. En la figura 4.2 se presenta un ejemplo de uno de estos espectrogramas.

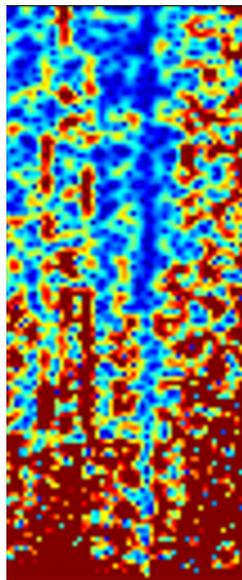


FIGURA 4.2: Ejemplo de uno de los espectrogramas obtenidos de 129x55 píxeles (imagen ampliada)

En resumen, los espectrogramas son herramientas valiosas para analizar cómo evolucionan las características espectrales de una señal en el tiempo. Son ampliamente utilizados en el análisis de señales unidimensionales para identificar patrones, detectar eventos y extraer información relevante sobre el contenido espectral. En este proyecto, estos espectrogramas serán empleados como entrada para nuestra redes neuronales.

#### 4.1.3 SECTORIZACIÓN DE PACIENTES

---

Para terminar la explicación del procesamiento de los datos, falta hablar sobre la sectorización de pacientes. Puesto que el objetivo del trabajo es hallar diferencias entre los diferentes subtipos de TDAH, es necesario dividir los datos previamente en función de su tipo.

Como datos de partida, se tienen en total 118 sujetos diagnosticados con TDAH, distribuidos de la siguiente manera:

- 40 diagnosticados con TDAH de tipo inatento (TDAH-I)

- 62 diagnosticados con TDAH de tipo combinado (TDAH-C)
- 16 diagnosticados con TDAH de tipo hiperactivo (TDAH-H)

El diagnóstico tanto del trastorno como de su subtipo se realizó siguiendo los criterios establecidos en el Diagnostic and Statistical Manual of Mental Disorders (DSM-V) [1].

Es importante destacar que el número de pacientes con diagnóstico de TDAH-H es significativamente menor, lo que representa un desafío particular para este estudio. En términos de género, 93 de los pacientes son hombres y 25 son mujeres, evidenciando una mayor prevalencia masculina en la muestra. En cuanto a la edad, los sujetos tienen entre 6 y 15 años, lo que permite analizar datos de diferentes etapas del desarrollo, desde la niñez hasta la adolescencia. Sin embargo, para este estudio el análisis no se centrará en el género ni en la edad de los sujetos.

Para simplificar la nomenclatura, a partir de ahora, se referirá a los pacientes con TDAH-I como Tipo1, a los pacientes con TDAH-C como Tipo2 y a los pacientes con TDAH-H como Tipo3.

En la tabla 4.1 se presenta una representación gráfica del número y porcentaje de pacientes correspondientes a cada tipo.

Tipo de TDAH	Grupo	Número	Porcentaje
TDAH-I	Tipo1	40	33.9 %
TDAH-C	Tipo2	62	52.5 %
TDAH-H	Tipo3	16	13.6 %

TABLA 4.1: Sujetos de estudio de cada grupo

## 4.2 APLICACIÓN DE REDES NEURONALES

En este apartado, se aborda la aplicación de redes neuronales en el proyecto, explicando detalladamente el diseño y arquitectura de la red neuronal convolucional (CNN) y del *vision Transformer* (ViT) utilizados. Además, se describe el proceso de entrenamiento implementado para ambas redes y los resultados del mismo.

### 4.2.1 DISEÑO DE LA CNN

En este apartado se detalla la estructura de la red neuronal convolucional (CNN) utilizada, describiendo las capas que la componen, su configuración y el propósito de cada una en el procesamiento de las imágenes.

La red convolucional está compuesta por las siguientes capas [3, 5]:

- **Capas convolucionales:** tres capas convolucionales con 32, 64 y 128 filtros, cada uno de tamaño 10x10. Después de cada capa convolucional, se añade una capa de normalización *Batch* y una capa ReLU, lo que optimiza el rendimiento y la velocidad de la red.

- **Capas de *pooling*:** se implementan dos capas de *max pooling* con un tamaño de  $2 \times 1$ , lo que reduce la altura de las imágenes a la mitad en cada una de ellas. Posteriormente, se aplica una capa de *Dropout* para evitar el *overfitting*, desactivando aleatoriamente una fracción de neuronas.
- **Capa *fully connected*** (completamente conectada): una capa *fully connected* con dos neuronas, una para cada clase, que realiza la clasificación final.
- **Capa *Softmax*:** una capa *Softmax* que toma la decisión de clasificación, asignando probabilidades a cada clase.

Este diseño de red convolucional permite lograr buenos resultados en la tarea de clasificación, manteniendo una duración razonable en los entrenamientos. A medida que las imágenes disminuyen de tamaño, se incrementa el número de filtros para capturar características más detalladas, sin aumentar significativamente la carga computacional.

En la figura 4.3 se puede ver de manera esquemática la estructura de la CNN.

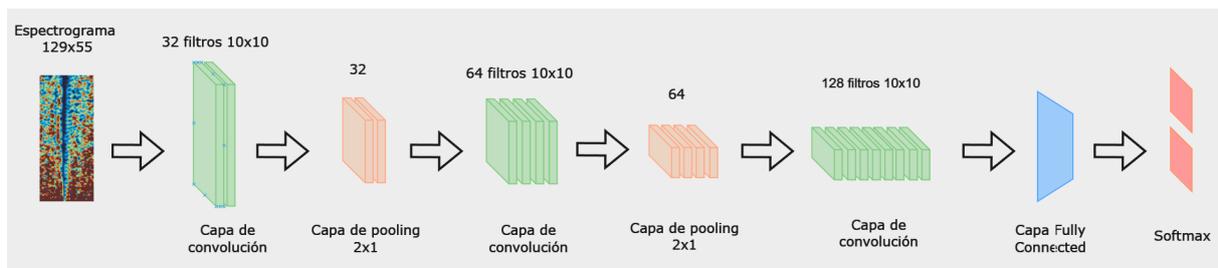


FIGURA 4.3: Arquitectura de la CNN

#### 4.2.2 DISEÑO DEL *Transformer*

La arquitectura del *Transformer* implementada en este trabajo, inspirada en el modelo *vision Transformer* (ViT), se detalla a continuación.

- **Entrada y partición:** Los espectrogramas ( $129 \times 55$  píxeles) son divididos en parches de  $25 \times 25$ .
- ***Embedding*:** cada parche se convierte en un embedding de dimensión 128, añadiendo el código de posición.
- **Bloques *Transformer*:** se tienen 2 bloques *Transformer*, cada uno compuesto por:
  - 1 capa de atención multi-cabeza con 4 cabezas y dimensión de 128.
  - 1 capa de normalización (*Layer Normalization*).
  - 1 capa de FNN con 256 unidades en la capa oculta y activación ReLU.

Cuenta con una tasa de *dropout* de 0.2 (20%) en las capas de atención y en la FNN para evitar el *overfitting*.

- **Clasificación:** Los parches procesados se aplanan (*Flatten*) y pasan por una capa densa intermedia de 128 unidades con activación ReLU y *dropout*. Finalmente, se utiliza una capa Softmax con 2 neuronas de salida, correspondientes a cada una de las clases.
- **Optimización:** se emplea el optimizador Adam con una tasa de aprendizaje inicial de  $10^{-3}$  y *categorical crossentropy* como función de pérdida.

Asimismo, se probaron con otras combinaciones de parámetros para la configuración del *Transformer*, pero estos resultaron los más adecuados para este tamaño de imagen. Estos parámetros, ofrecen un equilibrio entre unos buenos resultados y tiempos de ejecución razonables.

### 4.2.3 ENTRENAMIENTO DE LAS REDES

Al igual que en [5], el objetivo es diferenciar y caracterizar el TDAH de tipo hiperactivo (TDAH-H) que, como se ha visto, se trata del tipo menos diagnosticado. Por lo tanto, se han realizado dos grupos de entrenamiento diferentes. En el primero, se ha llevado a cabo una comparativa entre TDAH-I y TDAH-H (*Tipo1 vs Tipo3*), mientras que en el segundo se ha comparado entre TDAH-C y TDAH-H (*Tipo2 vs Tipo3*).

Para estos entrenamientos, se utiliza validación cruzada con 10  *folds*, distribuyendo los datos en un 70 % para el conjunto de entrenamiento y un 30 % para el conjunto de prueba. De esta manera, se generan 10  *folds* con sujetos Tipo1 y Tipo3, y otros 10  *folds* con sujetos Tipo2 y Tipo3.

En la tabla 4.2, se representa el número de muestras (espectrogramas) de cada uno de los tipos, para periodo diurno y nocturno.

Grupo	Nº espectrogramas Día	Nº espectrogramas Noche
Tipo1	3343	1288
Tipo2	5155	2120
Tipo3	1339	557

TABLA 4.2: Número de espectrogramas disponibles para cada tipo

Como ya se vio en [5], el número de espectrogramas de Tipo3 es significativamente menor comparado con los otros dos tipos. Esto supone un reto a la hora de realizar un buen entrenamiento ya que el número de muestras no está balanceado. Por ello, en [5] se estudió la idea de realizar un entrenamiento balanceado las muestras, esto es, que hubiese un 50 % de muestras de cada tipo. Sin embargo, se comprobó que se obtenían peores resultados balanceando las muestras ya que así, se tenían muy pocas muestras no solo del Tipo3 sino de ambos. Es por ello que en el presente trabajo, se entrenará la red mediante validación cruzada con 10  *folds* no balanceados. En la tabla 4.3, se muestra la distribución de los  *folds* utilizados:

Red	Grupo	Porcentaje	Entrenamiento	Test
<b>Tipo1 vs Tipo3</b>	Tipo1	71.4 %	28	12
	Tipo3	28.6 %	12	4
<b>Tipo2 vs Tipo3</b>	Tipo2	79.5 %	44	18
	Tipo3	20.5 %	12	4

TABLA 4.3: Distribución de los *fold*s no balanceados

Para evaluar el rendimiento del entrenamiento se va a emplear la precisión (*accuracy*) del conjunto de test que consiste en la proporción de predicciones correctas realizadas por un modelo sobre los datos de test. Esta métrica mide qué tan bien generaliza el modelo datos no vistos después del entrenamiento. Matemáticamente, la precisión del conjunto de prueba se calcula como:

$$\text{Precisión} = \frac{\text{N}^\circ \text{ de muestras correctamente clasificadas}}{\text{N}^\circ \text{ total de muestras de test}} \quad (4.3)$$

No obstante, no se realiza un diagnóstico individual por paciente. En caso de necesitar un diagnóstico personalizado, sería necesario implementar clasificadores que evalúen al paciente en su totalidad. En este enfoque, la red etiqueta cada espectrograma de un paciente como perteneciente a uno u otro tipo.

A continuación, en la tabla 4.4 se muestran los resultados del entrenamiento de las dos comparativas entre tipos para las dos redes explicadas en esta sección.

Red	Periodo	CNN	<i>Transformer</i>
<b>Tipo1 vs Tipo3</b>	Día	62,8 %	75.31 %
	Noche	59.7 %	72.35 %
<b>Tipo2 vs Tipo3</b>	Día	69,9 %	81.83 %
	Noche	69.6 %	80.92 %

TABLA 4.4: Media de las precisiones del test de las redes entrenadas con 10 *fold*s

En general, se observa que el modelo *Transformer* presenta mejores valores de precisión en comparación con el modelo CNN para todas las combinaciones de subtipos y periodos de actividad, lo que sugiere que el *Transformer* es más adecuado para las tareas de clasificación de los subtipos de TDAH en este conjunto de datos. Por otro lado, para el periodo diurno, se obtiene una leve mejora de los valores en comparación con el periodo nocturno.

Si bien los resultados obtenidos no son desfavorables, particularmente en el caso del modelo *Transformer*, tampoco pueden considerarse sobresalientes ni especialmente destacables. Por lo tanto, con el objetivo de mejorar los resultados y obtener conclusiones más precisas para diferenciar los subtipos de TDAH, se van a aplicar herramientas de XAI que permitirán interpretar mejor el comportamiento del modelo. Estas herramientas ayudarán a identificar las características más relevantes de los espectrogramas que influyen en las decisiones del modelo, y de esta forma, optimizar la clasificación de los subtipos de TDAH, mejorando su capacidad predictiva y su comprensión.

## 4.3 INTERPRETACIÓN MEDIANTE TÉCNICAS XAI

---

Las técnicas de Inteligencia Artificial Explicable permiten desentrañar cómo los modelos toman decisiones, proporcionando transparencia y confianza en sus resultados. En este apartado se van a abordar las diferentes técnicas de interpretación utilizadas a fin de obtener unos resultados más concluyentes.

### 4.3.1 CREACIÓN DE MAPAS

---

Una vez entrenados los modelos, a partir de ellos y de los espectrogramas, se crean los mapas explicados en el apartado 3.3.1:

- Mapas de oclusión
- Mapas Grad-CAM
- Mapas de atención

Tanto para la red neuronal convolucional como para el *Transformer*, se han creado los 3 tipos de mapas (oclusión, Grad-CAM y atención). Para el caso de la CNN, se han obtenido los mapas en cada una de las 3 capas convolucionales que tiene la red. En cuanto al *ViT*, los mapas de atención se han calculado en las dos capas de atención del *Transformer*, mientras que los mapas de oclusión y Grad-CAM son calculados para el modelo de manera global ya que son calculados a posteriori.

Para la creación de los mapas, se seleccionaron únicamente los espectrogramas con los porcentajes de acierto más altos, de manera que los mapas reflejen la información más representativa y fiable posible. La elección del umbral se realizó teniendo en cuenta que se dispusiera de un número suficiente de muestras para cada porcentaje. En el caso de los mapas obtenidos a partir de la CNN, se incluyeron los espectrogramas cuya clasificación alcanzó un acierto igual o superior al 80%. Por otro lado, para los mapas generados a partir del *Transformer*, se seleccionaron aquellos espectrogramas con un acierto igual o superior al 90%, ya que este modelo presenta un rendimiento superior en términos de precisión y permite establecer un umbral más alto sin comprometer la cantidad de muestras disponibles (véase tabla 4.4).

A continuación, en las figuras 4.4, 4.5 y 4.6 se muestra un ejemplo de cada tipo de mapa creados a partir de la CNN correspondientes a un mismo paciente.

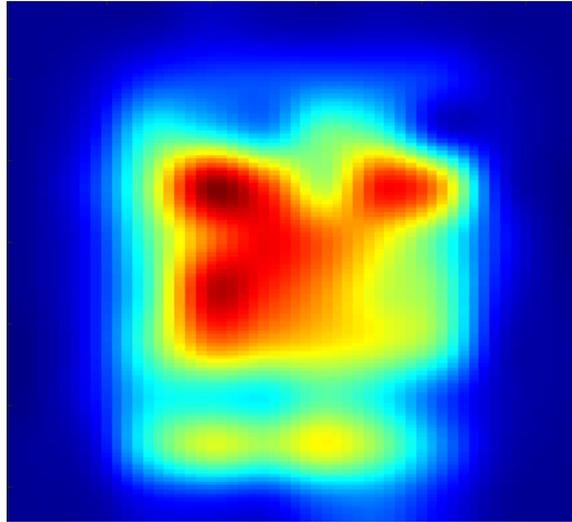


FIGURA 4.4: Ejemplo de mapa de oclusión de la CNN

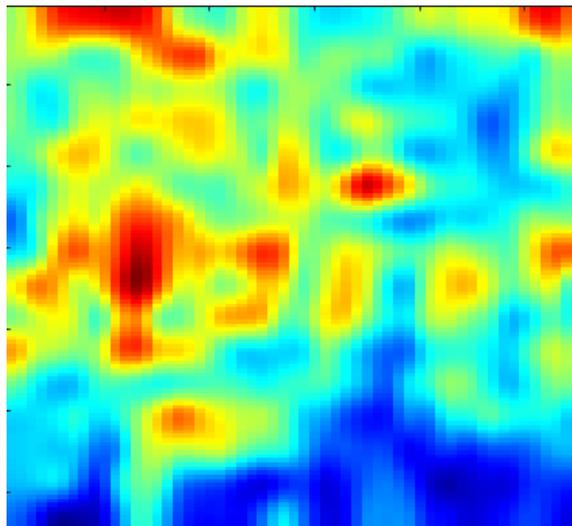


FIGURA 4.5: Ejemplo de mapa Grad-CAM de la CNN

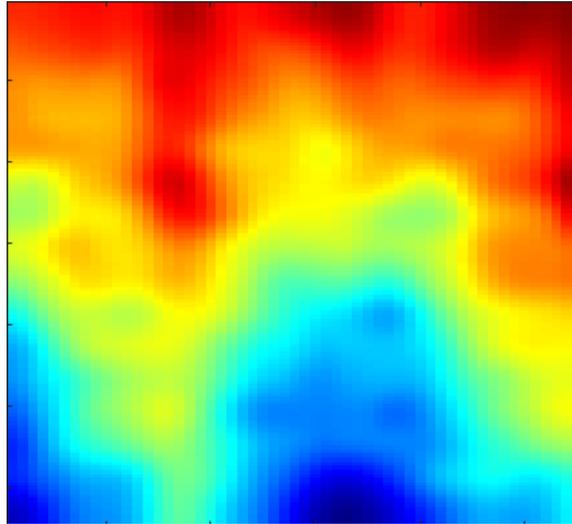
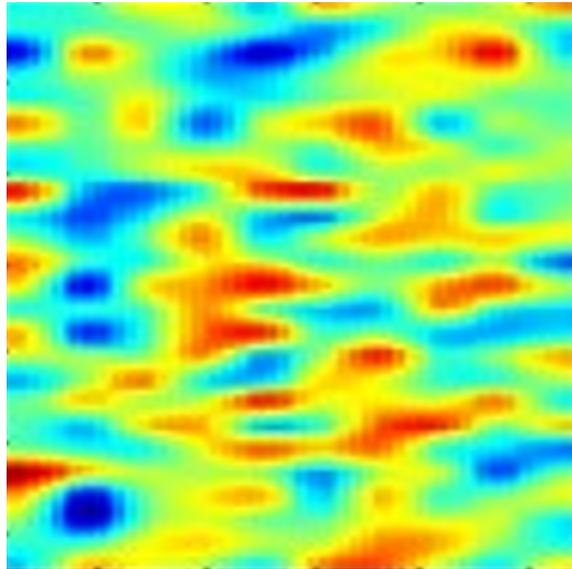


FIGURA 4.6: Ejemplo de mapa de atención de la CNN

A continuación, en las figuras 4.7, 4.8 y 4.9 se muestra un ejemplo de cada tipo de mapa creados a partir del *Transformer* (ViT) correspondientes a un mismo paciente.

FIGURA 4.7: Ejemplo de mapa de oclusión del *Transformer*

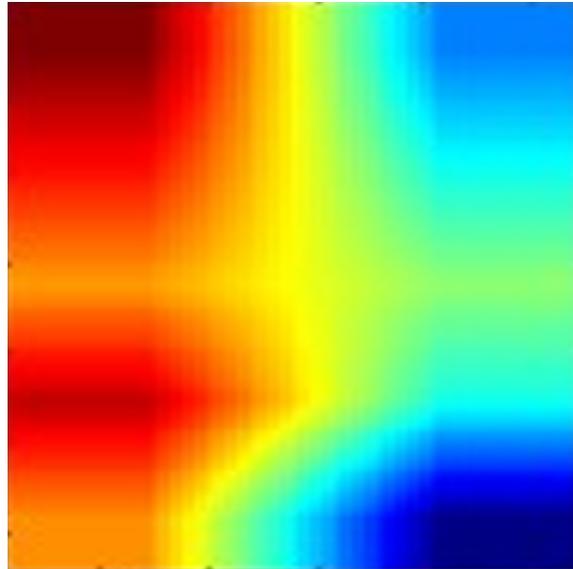


FIGURA 4.8: Ejemplo de mapa Grad-CAM del *Transformer*

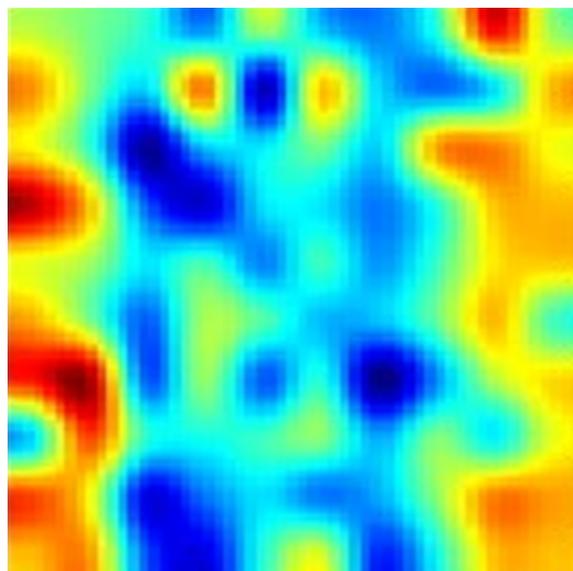


FIGURA 4.9: Ejemplo de mapa de atención del *Transformer*

### 4.3.2 ENMASCARAMIENTO DE ESPECTROGRAMAS

Tras crear los diferentes mapas para cada modelo, se lleva a cabo un proceso de **enmascaramiento** de estos mapas. El enmascaramiento consiste en aplicar una máscara que oculta partes del espectrograma, dejando visibles únicamente aquellas regiones que son de mayor interés o relevancia según ciertos criterios, como la intensidad de la señal. Este proceso se utiliza para reducir la complejidad de la visualización y focalizar el análisis en las zonas más importantes, eliminando información que puede ser redundante o que no contribuye de manera significativa al estudio.

El proceso de enmascaramiento aplicado se basa en la generación de una máscara a partir de los mapas de activación generados. Este tipo de mapas se obtienen identificando las zonas donde la red o el modelo utilizado para analizar los espectrogramas muestra una mayor activación, es decir, las regiones que tienen mayor influencia en la toma de decisiones. Estas zonas activas son las que se destacan como relevantes dentro del espectro.

Una vez que se genera el mapa, se aplica un umbral para crear la máscara, que representa gráficamente las áreas de interés. Este umbral puede ajustarse según el porcentaje de enmascaramiento e indican el porcentaje de la imagen total que queda cubierto, limitando la visualización a las zonas más relevantes del mapa de ocusión. En concreto, para nuestros mapas, se han utilizado porcentajes de enmascaramiento de 60 %, 70 % y 80 %. Cuanto mayor sea este valor, la máscara será más restrictiva, esto es, será más pequeña; mientras que si es más bajo, la máscara será más grande, se cogerá más parte del mapa. El porcentaje de enmascaramiento más conveniente dependerá del tipo de mapa utilizado. A continuación, en la figura 4.10 se muestra un ejemplo de máscara para los diferentes porcentajes de enmascaramiento para un mismo mapa.

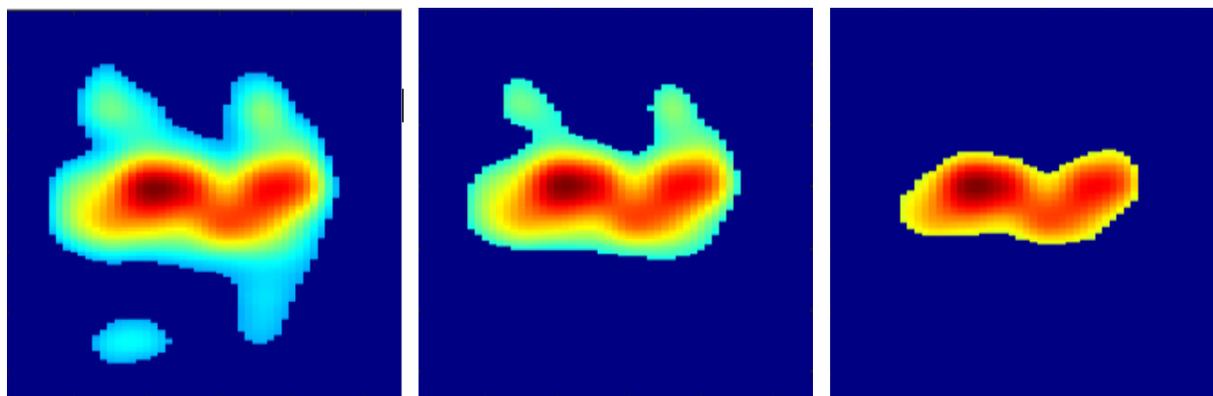


FIGURA 4.10: Máscara para diferentes porcentajes de enmascaramiento. De izquierda a derecha: 60 %, 70 % y 80 %

Por último, la máscara generada se coloca sobre el espectrograma original, creando lo que denominaremos **espectrograma enmascarado**. Este espectrograma enmascarado muestra sus zonas de mayor relevancia, mientras oculta el resto de la información menos significativa. El objetivo de este proceso es mejorar la claridad y precisión del análisis, al dirigir la atención hacia las partes del espectrograma que tienen un mayor impacto o relevancia para el estudio.

En las figuras 4.11, a 4.16 se muestra un ejemplo del proceso de enmascaramiento para cada tipo

de mapa obtenido. En orden, se muestra: el mapa original, la máscara creada a partir del mapa, el espectrograma original y el espectrograma enmascarado.

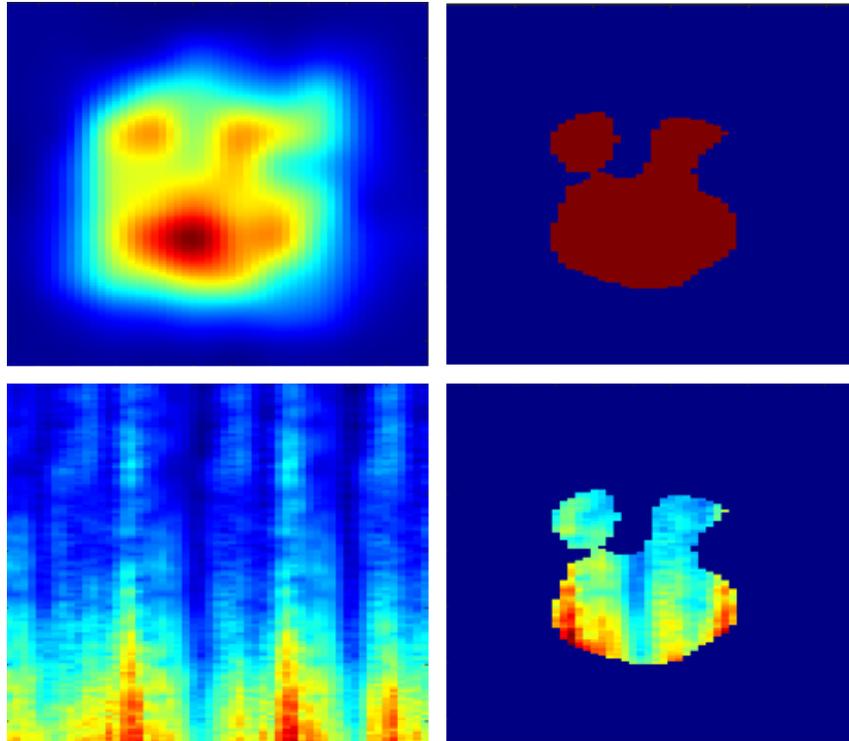


FIGURA 4.11: Proceso de enmascaramiento a partir de un mapa de oclusión para la CNN

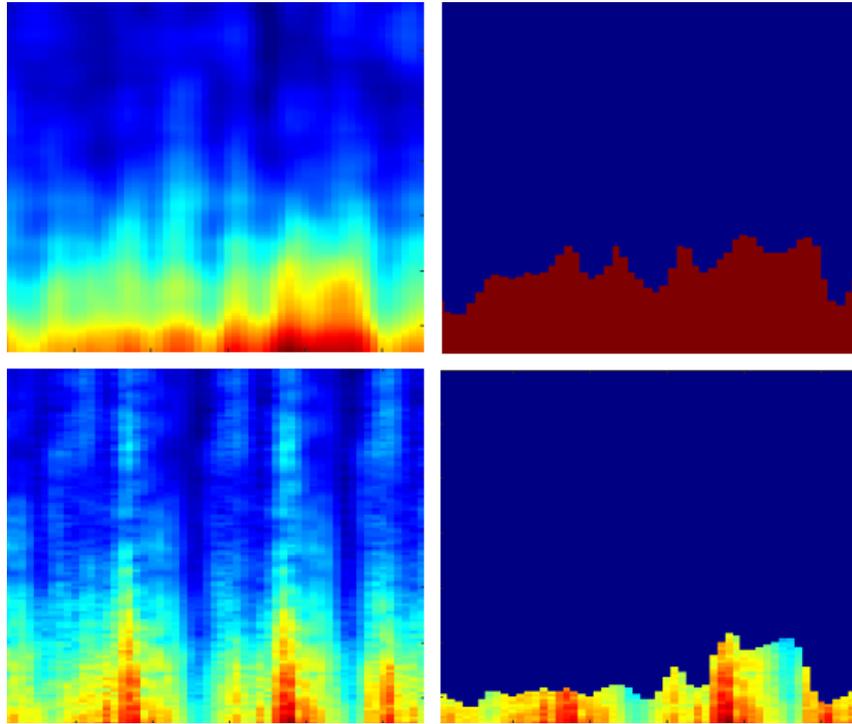


FIGURA 4.12: Proceso de enmascaramiento a partir de un mapa Grad-CAM para la CNN

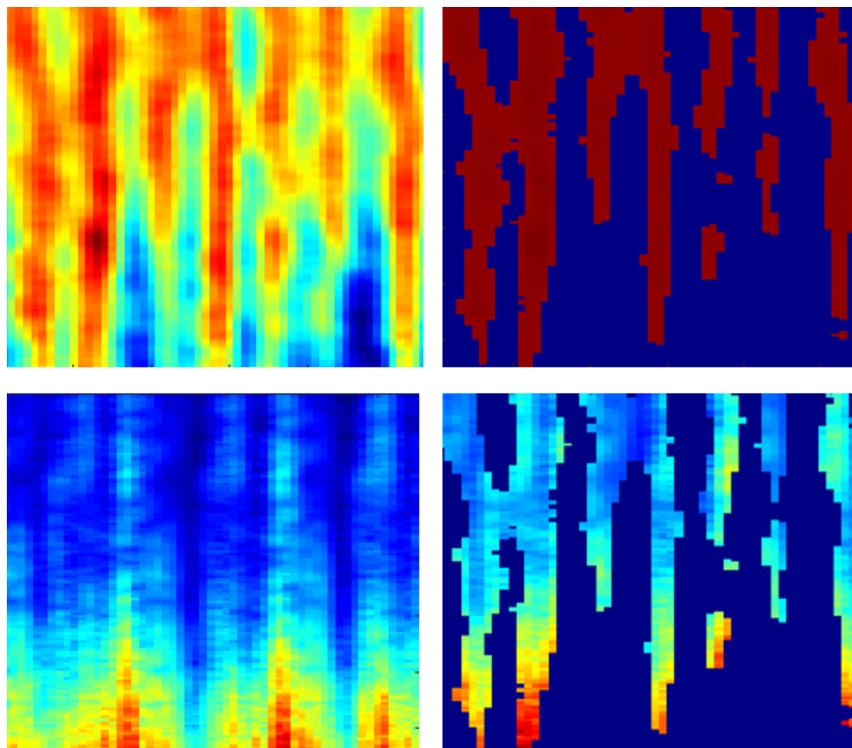


FIGURA 4.13: Proceso de enmascaramiento a partir de un mapa de activación para la CNN

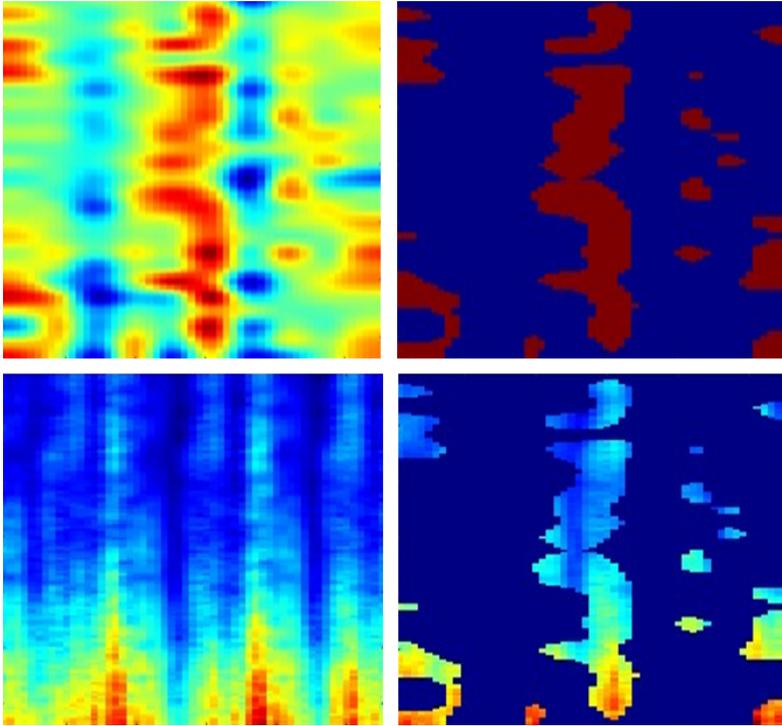


FIGURA 4.14: Proceso de enmascaramiento a partir de un mapa de oclusión para el *Transformer*

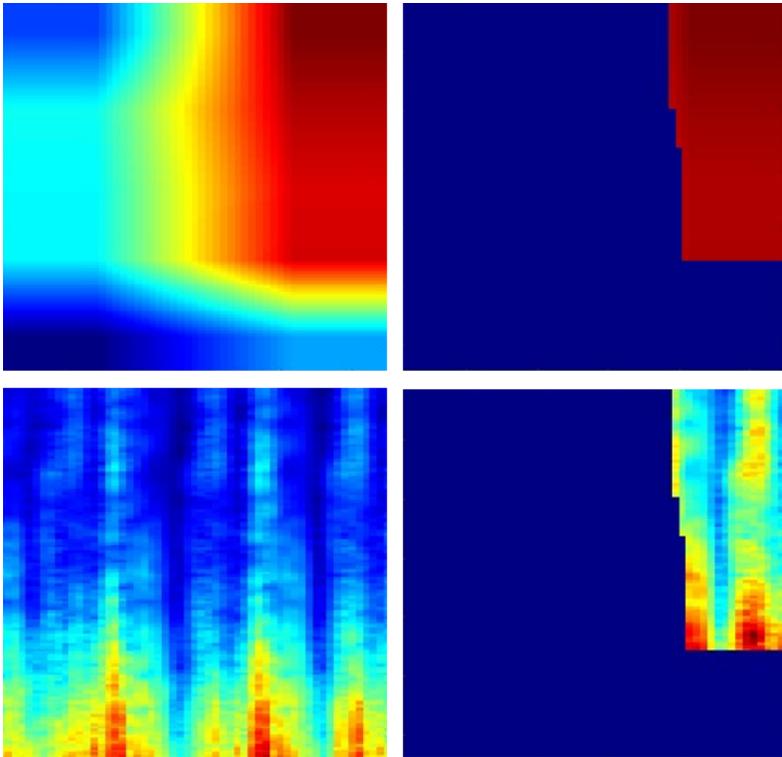


FIGURA 4.15: Proceso de enmascaramiento a partir de un mapa Grad-CAM para el *Transformer*

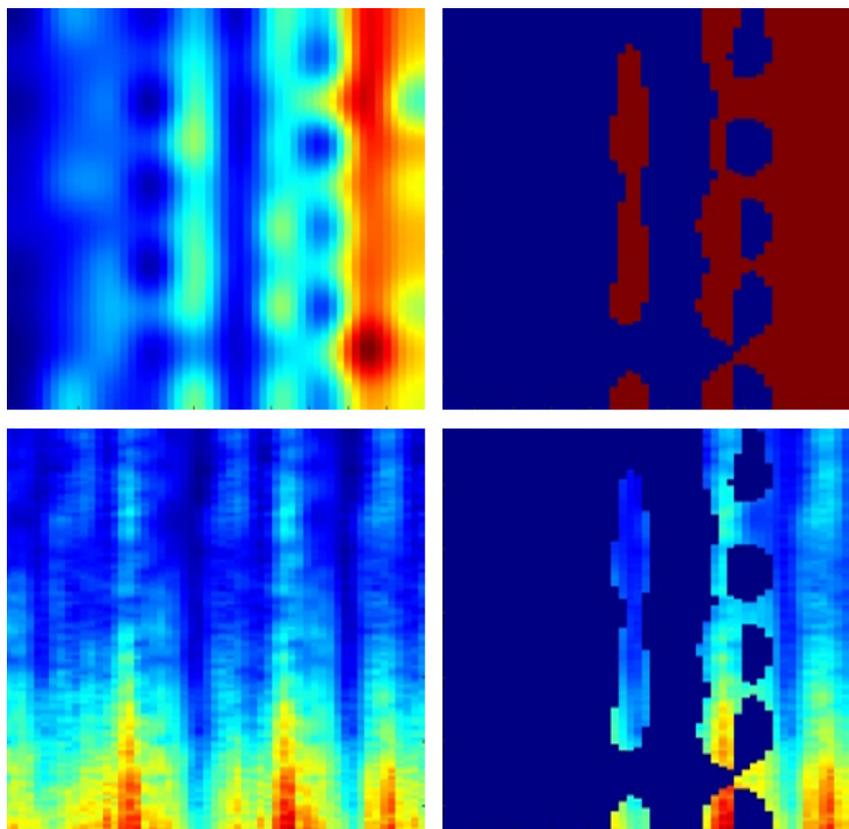


FIGURA 4.16: Proceso de enmascaramiento a partir de un mapa de atención para el *Transformer*

### 4.3.3 CÁLCULO DE CARACTERÍSTICAS ESPECTRALES

A partir de los espectrogramas enmascarados, se realiza el cálculo de las características espectrales. Además de los mecanismos de XAI previamente explicados y los mapas, el cálculo de características espectrales o frecuenciales se presenta como una herramienta de gran utilidad. El cálculo de estas características ofrece una forma de analizar y comprender mejor los conjuntos de datos a comparar, revelando aspectos que a simple vista pueden ser difíciles de identificar como patrones o particularidades en las variaciones de intensidad y frecuencia de los mapas. Al aplicar este enfoque, se puede obtener una comprensión más clara de las particularidades de los datos, mejorando la interpretabilidad y la transparencia del modelo en el contexto de la Inteligencia Artificial explicable (XAI).

Este proceso implica extraer métricas relacionadas con diferentes aspectos clave como la energía, frecuencia y dinámica temporal de las señales. Dichas características son esenciales para capturar patrones diferenciadores en las señales, facilitando la comparación entre los diferentes subtipos de TDAH.

Las características espectrales calculadas incluyen varias métricas que describen diferentes aspectos de la distribución de frecuencia en los mapas. Se han extraído diversas métricas espectrales a partir de los espectrogramas ponderados que permiten caracterizar la distribución de energía en frecuencia y su evolución temporal.

Para formalizar las expresiones utilizadas, se define:

- $\mathcal{S}[k, n]$  es el espectrograma ponderado unilateral y discreto
- $k = 0, \dots, K$  es el índice de frecuencia.
- $n = 0, \dots, N - 1$  es el índice temporal.
- $f[k]$  son las frecuencias discretas.

A continuación se presentan dichas características junto a una breve descripción [80, 81, 82]:

- **Relative power (RP) o potencia relativa:** proporción de energía alterna (AC) respecto al total.

$$RP = \frac{1}{N} \sum_{n=0}^{N-1} \left[ \frac{\sum_{k=1}^K \mathcal{S}[k, n]}{\sum_{k=0}^K \mathcal{S}[k, n]} \right] \quad (4.4)$$

- **Spectral bandwidth (SpecBand) o Ancho de banda espectral:** medida de dispersión del espectro respecto a su centroide ( $SpC[n]$ ).

$$SpecBand = \frac{1}{N} \sum_{n=0}^{N-1} \left[ \frac{\sum_{k=0}^K (f[k] - SpC[n])^2 \cdot \mathcal{S}[k, n]}{\sum_{k=0}^K \mathcal{S}[k, n]} \right] \quad (4.5)$$

- **Spectral crest factor (SpecCrestFac) o Factor de cresta espectral:** identifica la frecuencia dominante, comparando el pico espectral con la energía total.

$$SpecCrestFac = \frac{1}{N} \sum_{n=0}^{N-1} \left[ \frac{\max_k(\mathcal{S}[k, n])}{C \cdot \sum_{k=0}^K \mathcal{S}[k, n]} \right] \quad (4.6)$$

- **Spectral entropy (SpecEn) o entropía espectral:** entropía de la potencia relativa.

$$SpecEn = \frac{1}{N} \sum_{n=0}^{N-1} \left[ \sum_{k=0}^K RP[k, n] \cdot \log_2(RP[k, n]) \right] \quad (4.7)$$

donde  $RP[k, n]$  es la potencia relativa de la señal.

- **Spectral flatness (SpecFlat) o planicidad espectral:** indica cuán plana es la distribución espectral.

$$SpecFlat = \frac{1}{N} \sum_{n=0}^{N-1} \left[ \frac{\exp\left(\frac{1}{K+1} \sum_{k=0}^K \log(\mathcal{S}[k, n])\right)}{\frac{1}{K+1} \sum_{k=0}^K \mathcal{S}[k, n]} \right] \quad (4.8)$$

- **Spectral flux (SpecFlux) o flujo espectral:** cuantifica los cambios abruptos entre espectros de instantes consecutivos.

$$SpecFlux = \frac{1}{N-1} \sum_{n=1}^{N-1} \left[ \sum_{k=0}^K (\mathcal{S}[k, n] - \mathcal{S}[k, n-1]) \right] \quad (4.9)$$

- **Spectral Renyi entropy (SpecRenyiEn) o entropía de Renyi espectral:** mide la incertidumbre o aleatoriedad en la distribución espectral.

$$SpecRenyiEn = \frac{1}{N} \sum_{n=0}^{N-1} \left[ \frac{1}{1-q} \cdot \log \left( \sum_{k=0}^K \mathcal{S}[k, n] \right)^q \right] \quad (4.10)$$

- **Spectral roll-off (SpecRolloff) o punto de corte espectral:** define la frecuencia en la que se incluye el 85% de la energía total.

$$SpecRolloff = \frac{1}{N} \sum_{n=0}^{N-1} f[k_{85}[n]] \quad (4.11)$$

Estas características, al analizarse en conjunto con los mapas generados por las técnicas de XAI, proporcionan una perspectiva más rica y detallada del comportamiento del modelo. Este enfoque permite una interpretación más precisa y completa de las decisiones del modelo, mejorando la transparencia y la confianza en sus predicciones.

#### 4.3.4 ANÁLISIS ESTADÍSTICO: *p-valores*

Si bien los mecanismos de XAI proporcionan explicaciones sobre cómo se llega a una predicción, es fundamental complementar estos enfoques con herramientas estadísticas que permitan evaluar la relevancia e importancia de los resultados obtenidos. En este contexto, tras la extracción de las características espectrales, se realiza un análisis estadístico mediante el cálculo de los *p-valores*, lo cual permite determinar si las diferencias observadas entre los subtipos de TDAH son estadísticamente significativas. Este análisis refuerza la confiabilidad de las conclusiones extraídas y permite identificar las características más relevantes, facilitando así una evaluación objetiva y rigurosa de las particularidades que distinguen a cada subtipo.

El *p-valor* es el resultado de un test de hipótesis que se utiliza como herramienta estadística para evaluar la probabilidad de que los resultados observados sean consistentes con la hipótesis nula ( $H_0$ ). La hipótesis nula plantea que no existen diferencias significativas entre los parámetros comparados o que cualquier efecto observado se debe al azar. Un *p-valor* bajo indica que los resultados obtenidos son poco probables bajo la suposición de la hipótesis nula, lo que podría sugerir la presencia de una diferencia significativa. Generalmente, se considera un nivel de 0.05 como umbral para determinar si un *p-valor* es suficientemente pequeño como para rechazar la hipótesis nula.

El cálculo de los *p-valores* es una herramienta clave para determinar si existe una diferencia significativa entre dos grupos. Dependiendo de las características de los datos, se puede utilizar la *prueba t* de dos muestras o la prueba de *Wilcoxon* [83]:

- La *prueba t* de dos muestras es un test paramétrico que se usa para comparar las medias de dos grupos, bajo el supuesto de que los datos siguen una distribución normal y las varianzas son iguales. Calcula la diferencia entre las medias dividida por el error estándar, obteniendo un *t-estadístico*, cuyo *p-valor* se utiliza para determinar si hay una diferencia significativa entre los grupos. La fórmula del *t-estadístico* es [84]:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \quad (4.12)$$

donde  $\bar{x}$  y  $\bar{y}$  son las medias de los dos grupos,  $s_x^2$  y  $s_y^2$  son las varianzas y  $n$  y  $m$  son los tamaños de muestra.

- La prueba de *Wilcoxon*, también conocida como U de *Mann-Whitney*, es un test no paramétrico que compara las distribuciones de dos grupos, sin requerir normalidad ni igualdad de varianzas. Se basa en los rangos de los datos y calcula un estadístico U. El *p-valor* resultante indica la probabilidad de observar una diferencia en las distribuciones, asumiendo que no existe una diferencia real. La fórmula del estadístico U es [85]:

$$U = W - \frac{n_X(n_X + 1)}{2} \quad (4.13)$$

donde  $W$  es la suma de los rangos de uno de los grupos y  $n_x$  el tamaño de la muestra.

En el presente trabajo, los *p-valores* se obtienen automáticamente mediante un procedimiento que primero evalúa si los datos presentan una distribución aproximadamente normal. Esta comprobación se realiza mediante el test de bondad de ajuste  $\chi^2$ , implementado en MATLAB con la función `chi2gof`. Si ambos grupos se consideran normales, se aplica la prueba t; en caso contrario, se aplica la prueba de Wilcoxon. De esta forma, se asegura que la prueba estadística empleada sea coherente con la naturaleza de los datos, garantizando la validez de los resultados [86].

## Capítulo 5

# RESULTADOS Y DISCUSIÓN

En el presente capítulo se presentan los resultados obtenidos tras emplear la metodología descrita en el capítulo 4. El propósito es proporcionar resultados que sean fácilmente comprensibles y para ello, se realizarán las siguientes tareas:

- Presentación de los resultados mediante la exposición de los *p-valores* obtenidos para cada método.
- Análisis de los resultados con objeto de caracterizar los diferentes subtipos.
- Discusión final de los resultados obtenidos.

## 5.1 PRESENTACIÓN DE LOS RESULTADOS

---

En esta sección se van a presentar los resultados obtenidos tras aplicar el análisis visto en el apartado 4:

- Creación de los mapas, utilizando los 3 mecanismos vistos (mapas de oclusión, Grad-CAM y mecanismos de atención) para la CNN y el *Transformer*.
- Enmascaramiento de los espectrogramas para porcentajes de enmascaramiento de 60 %, 70 % y 80 % en función de los mapas.
- Cálculo de las características espectrales a partir de los espectrogramas enmascarados.
- Obtención de *p-valores* de las características espectrales.

Como se detalló en el apartado 4.3.4, un *p-valor* menor a 0.05 sugiere que el parámetro es significativo, lo que indica la existencia de diferencias relevantes entre los tipos analizados. Por lo tanto, los *p-valores* nos ayudarán a discernir qué método de interpretación es el más efectivo en nuestro caso, así como en qué capa convolucional debe ser aplicado. Además, se podrá ver el porcentaje de enmascaramiento óptimo.

En las secciones siguientes, se detallarán los resultados obtenidos para los 3 métodos de interpretación vistos y para los diferentes porcentajes de enmascaramiento (60 %, 70 % y 80 %). Como se explicó en el apartado 4.2.3, estos resultados se han calculado para las dos comparativas entre

subtipos, Tipo1 frente a Tipo3 y Tipo2 frente a Tipo3 y diferenciando entre actividad diurna y nocturna.

En cuanto al porcentaje de enmascaramiento, para cada método, únicamente se mostrará aquel para el cual se obtiene un mayor número de *p-valores* significativos (menores a 0.05) o bien, *p-valores* considerablemente pequeños. De esta manera, se sintetizan los resultados, mostrando aquellos de mayor interés para el estudio.

Los resultados serán divididos por modelo (CNN o *transformer*) así como por comparativa de tipos y periodo. Para la CNN, se mostrarán los 3 tipos de mapas obtenidos en las 3 capas convolucionales. Mientras que para el *Transformer*, como se explicó en el apartado 4.3.1, no se realizó un análisis por capas para la obtención de los mapas de oclusión y de atención como en la CNN, sino para el modelo de manera global por lo que se mostrarán en la misma tabla: mapas de oclusión, Grad-CAM y atención (capa 1 y capa 2).

### 5.1.1 RESULTADOS: COMPARATIVA *Tipo1 vs Tipo3*

A continuación, se van a representar los resultados obtenidos para la comparativa realizada entre Tipo1 y Tipo3.

#### 5.1.1.1 *Tipo1 vs Tipo3*: CNN CAPA CONVOLUCIONAL 3

En la tabla 5.1 se representan los *p-valores* de las características frecuenciales de cada uno de los métodos para *Tipo1 vs Tipo3*, en el periodo nocturno para la capa convolucional 3 de la CNN, destacando aquellos *p-valores* menores de 0.05.

Capa convolucional 3 CNN			
NOCHE: <i>Tipo1 vs Tipo3</i>			
Parámetros	Oclusión (80 %)	Grad-CAM (70 %)	Atención (80 %)
RP	0.5482	0.5890	0.5950
SpecBand	<b>8.8698E-07</b>	0.1525	0.1105
SpecCrestFac	<b>4.9704E-04</b>	0.7728	0.9566
SpecEn	<b>3.7462E-06</b>	0.7578	0.8419
SpecFlat	<b>4.0881E-06</b>	<b>0.0206</b>	<b>0.0342</b>
SpecFlux	<b>0.0028</b>	0.6765	0.2314
SpecRenyiEn	<b>4.0881E-06</b>	<b>0.0223</b>	<b>0.0081</b>
SpecRolloff	0.2826	0.6243	0.4637

TABLA 5.1: *p-valores* de las características frecuenciales para *Tipo1 vs Tipo3*, en el periodo nocturno para la capa convolucional 3 de la CNN

En la tabla 5.2 se representan los *p-valores* de las características frecuenciales de cada uno de los métodos para *Tipo1 vs Tipo3*, en el periodo diurno para la capa convolucional 3 de la CNN, destacando aquellos *p-valores* menores de 0.05.

Capa convolucional 3 CNN			
DÍA: Tipo1 vs Tipo3			
Parámetros	Oclusión (70 %)	Grad-CAM (80 %)	Atención (60 %)
RP	0.5821	0.6176	0.5098
SpecBand	0.4354	0.4571	0.4793
SpecCrestFac	<b>0.0338</b>	0.2930	0.9566
SpecEn	0.3840	0.3742	0.9855
SpecFlat	0.4907	0.3800	0.4705
SpecFlux	<b>9.0305E-04</b>	0.8990	0.9133
SpecRenyiEn	0.3742	0.3638	0.2846
SpecRolloff	0.1885	0.5621	0.1692

TABLA 5.2:  $p$ -valores de las características frecuenciales para Tipo1 vs Tipo3, en el periodo diurno para la capa convolucional 3 de la CNN

Se puede observar que los  $p$ -valores son notablemente más bajos durante la noche, especialmente con el método de oclusión, lo que indica una mayor capacidad para distinguir entre Tipo1 y Tipo3. En cambio, durante el día y con Grad-CAM o atención, los valores son más elevados, mostrando una menor sensibilidad.

#### 5.1.1.2 Tipo1 vs Tipo3: CNN CAPA CONVOLUCIONAL 2

En la tabla 5.3 se representan los  $p$ -valores de las características frecuenciales de cada uno de los métodos para Tipo1 vs Tipo3, en el periodo nocturno para la capa convolucional 2 de la CNN, destacando aquellos  $p$ -valores menores de 0.05.

Capa convolucional 2 CNN			
NOCHE: Tipo1 vs Tipo3			
Parámetros	Oclusión (80 %)	Grad-CAM (80 %)	Atención (70 %)
RP	0.5594	0.5315	0.6604
SpecBand	<b>0.0085</b>	<b>0.0295</b>	0.3864
SpecCrestFac	<b>4.8876E-04</b>	0.6833	0.1986
SpecEn	<b>0.0111</b>	0.9277	0.9133
SpecFlat	<b>0.0100</b>	<b>0.0401</b>	0.4682
SpecFlux	0.4182	0.3083	0.1681
SpecRenyiEn	<b>0.0117</b>	<b>0.0404</b>	0.4243
SpecRolloff	0.5451	0.4658	0.4850

TABLA 5.3:  $p$ -valores de las características frecuenciales para Tipo1 vs Tipo3, en el periodo nocturno para la capa convolucional 2 de la CNN

En la tabla 5.4 se representan los  $p$ -valores de las características frecuenciales de cada uno de los métodos para Tipo1 vs Tipo3, en el periodo diurno para la capa convolucional 2 de la CNN, destacando aquellos  $p$ -valores menores de 0.05.

Capa convolucional 2 CNN			
DÍA: Tipo1 vs Tipo3			
Parámetros	Oclusión (60 %)	Grad-CAM (60 %)	Atención (70 %)
RP	0.5364	0.5342	0.8707
SpecBand	<b>0.0029</b>	0.7958	<b>0.0038</b>
SpecCrestFac	<b>0.0422</b>	<b>0.0143</b>	<b>6.6948E-04</b>
SpecEn	<b>7.9255E-04</b>	0.0726	<b>3.2754E-05</b>
SpecFlat	<b>1.6159E-04</b>	<b>0.0100</b>	0.0501
SpecFlux	0.1065	0.7395	0.5618
SpecRenyiEn	<b>1.0387E-04</b>	0.0882	<b>0.0121</b>
SpecRolloff	0.7235	0.5988	<b>1.3744E-03</b>

TABLA 5.4: *p-valores* de las características frecuenciales para Tipo1 vs Tipo3, en el periodo diurno para la capa convolucional 2 de la CNN

En este caso, los *p-valores* más bajos se concentran en el periodo diurno, especialmente con los métodos de oclusión y atención, lo que indica una mayor capacidad de discriminación entre Tipo1 y Tipo3 durante el día. En cambio, por la noche, aunque oclusión sigue mostrando cierta sensibilidad, los valores en general son más altos, especialmente en Grad-CAM y atención, lo que sugiere un menor rendimiento en la comparativa.

#### 5.1.1.3 Tipo1 vs Tipo3: CNN CAPA CONVOLUCIONAL 1

En la tabla 5.5 se representan los *p-valores* de las características frecuenciales de cada uno de los métodos para Tipo1 vs Tipo3, en el periodo nocturno para la capa convolucional 1 de la CNN, destacando aquellos *p-valores* menores de 0.05.

Capa convolucional 1 CNN			
NOCHE: Tipo1 vs Tipo3			
Parámetros	Oclusión (60 %)	Grad-CAM (80 %)	Atención (60 %)
RP	0.5914	0.6218	0.7338
SpecBand	0.3274	0.0817	<b>1.7817E-05</b>
SpecCrestFac	<b>0.0175</b>	0.7717	0.9566
SpecEn	0.3232	0.1737	<b>0.0295</b>
SpecFlat	0.8703	0.1795	<b>8.9628E-05</b>
SpecFlux	0.9370	0.5770	<b>4.6382E-05</b>
SpecRenyiEn	0.6116	0.4793	<b>0.0065</b>
SpecRolloff	0.5644	0.5661	<b>1.0465E-05</b>

TABLA 5.5: *p-valores* de las características frecuenciales para Tipo1 vs Tipo3, en el periodo nocturno para la capa convolucional 1 de la CNN

En la tabla 5.6 se representan los *p-valores* de las características frecuenciales de cada uno de los métodos para Tipo1 vs Tipo3, en el periodo diurno para la capa convolucional 1 de la CNN, destacando aquellos *p-valores* menores de 0.05.

Capa convolucional 1 CNN			
DÍA: Tipo1 vs Tipo3			
Parámetros	Oclusión (70 %)	Grad-CAM (70 %)	Atención (60 %)
RP	0.5875	0.5272	0.4398
SpecBand	0.5375	<b>0.0336</b>	<b>9.0679E-04</b>
SpecCrestFac	0.7168	0.0545	<b>0.0071</b>
SpecEn	0.3665	0.3012	<b>0.0080</b>
SpecFlat	0.3695	0.2457	<b>0.0453</b>
SpecFlux	0.3526	0.0726	<b>3.8059E-07</b>
SpecRenyiEn	<b>0.0234</b>	<b>5.3179E-04</b>	<b>0.0065</b>
SpecRolloff	0.4354	0.0788	0.1199

TABLA 5.6: *p*-valores de las características frecuenciales para Tipo1 vs Tipo3, en el periodo diurno para la capa convolucional 1 de la CNN

En la capa 1 de la CNN, los *p*-valores más bajos se observan durante el día, especialmente con el método de atención, seguido de Grad-CAM. En el periodo nocturno también se aprecian valores reducidos, sobre todo con atención, mientras que oclusión y Grad-CAM presentan resultados más elevados en general.

#### 5.1.1.4 Tipo1 vs Tipo3: Transformer

En la tabla 5.7 se representan los *p*-valores de las características frecuenciales de cada uno de los métodos para Tipo1 vs Tipo3, en el periodo nocturno para el Transformer, destacando aquellos *p*-valores menores de 0.05.

Transformer				
NOCHE: Tipo1 vs Tipo3				
Parámetros	Oclusión (80 %)	Grad-CAM (80 %)	At. Capa 1 (60 %)	At. Capa2 (80 %)
RP	0.7895	0.9349	<b>0.0163</b>	0.7236
SpecBand	<b>0.0158</b>	0.3693	<b>0.0035</b>	0.1020
SpecCrestFac	<b>0.0372</b>	0.8066	<b>0.0443</b>	<b>0.0133</b>
SpecEn	0.5880	0.9494	0.0605	0.7647
SpecFlat	<b>0.0075</b>	0.2348	<b>0.0051</b>	<b>0.0109</b>
SpecFlux	0.6823	0.1885	0.9511	0.3025
SpecRenyiEn	<b>0.0073</b>	0.2805	<b>0.0287</b>	<b>0.0147</b>
SpecRolloff	<b>0.0027</b>	0.6764	<b>0.0005</b>	<b>0.0275</b>

TABLA 5.7: *p*-valores de las características frecuenciales para Tipo1 vs Tipo3, en el periodo nocturno para el Transformer

En la tabla 5.8 se representan los *p*-valores de las características frecuenciales de cada uno de los métodos para Tipo1 vs Tipo3, en el periodo diurno para el Transformer, destacando aquellos *p*-valores menores de 0.05.

<i>Transformer</i>				
<i>DÍA: Tipo1 vs Tipo3</i>				
Parámetros	Oclusión (80 %)	Grad-CAM (70 %)	At. Capa 1 (70 %)	At. Capa2 (60 %)
RP	0.1284	0.2805	0.1709	0.3693
SpecBand	0.3990	0.4947	<b>0.0008</b>	<b>0.0362</b>
SpecCrestFac	0.9205	0.6318	0.1731	0.1455
SpecEn	0.2618	0.3054	0.3644	0.1484
SpecFlat	<b>0.0133</b>	0.4996	0.1885	0.4503
SpecFlux	0.3535	0.6562	0.0943	0.9763
SpecRenyiEn	<b>0.0079</b>	0.4844	<b>0.0277</b>	0.2786
SpecRolloff	0.7583	0.2175	<b>0.0051</b>	<b>0.0168</b>

TABLA 5.8: *p-valores* de las características frecuenciales para *Tipo1 vs Tipo3*, en el periodo diurno para el *Transformer*

En este caso, los *p-valores* más bajos se concentran durante la noche, especialmente en las capas de atención del *Transformer*, donde. oclusión también aporta buenos resultados nocturnos. No obstante, si bien es cierto que se detectan bastantes características con *p-valores* inferiores a 0.05, sobre todo en el periodo nocturno, ninguno de ellos resulta especialmente bajo en comparación con otras comparativas realizadas, lo que sugiere una menor diferenciación entre los grupos en este caso.

### 5.1.2 RESULTADOS: COMPARATIVA *Tipo2 vs Tipo3*

A continuación, se van a representar los resultados obtenidos para la comparativa realizada entre *Tipo2* y *Tipo3*.

#### 5.1.2.1 *Tipo2 vs Tipo3*: CNN CAPA CONVOLUCIONAL 3

En la tabla 5.9 se representan los *p-valores* de las características frecuenciales de cada uno de los métodos para *Tipo2 vs Tipo3*, en el periodo nocturno para la capa convolucional 3 de la CNN, destacando aquellos *p-valores* menores de 0.05.

Capa convolucional 3 CNN			
NOCHE: Tipo2 vs Tipo3			
Parámetros	Oclusión (70 %)	Grad-CAM (70 %)	Atención (80 %)
RP	0.1118	0.2473	<b>0.0247</b>
SpecBand	<b>7.3486E-05</b>	0.1172	<b>1.6456E-04</b>
SpecCrestFac	<b>0.0225</b>	<b>0.0112</b>	<b>4.8035E-05</b>
SpecEn	<b>0.0018</b>	<b>0.0089</b>	<b>6.6145E-05</b>
SpecFlat	<b>0.0011</b>	0.1476	<b>2.5555E-04</b>
SpecFlux	<b>0.0042</b>	<b>0.0271</b>	<b>0.0328</b>
SpecRenyiEn	0.0664	0.2938	<b>3.9222E-04</b>
SpecRolloff	<b>0.0276</b>	<b>0.0081</b>	<b>0.0225</b>

TABLA 5.9: *p-valores* de las características frecuenciales para Tipo2 vs Tipo3, en el periodo nocturno para la capa convolucional 3 de la CNN

En la tabla 5.10 se representan los *p-valores* de las características frecuenciales de cada uno de los métodos para Tipo2 vs Tipo3, en el periodo diurno para la capa convolucional 3 de la CNN, destacando aquellos *p-valores* menores de 0.05.

Capa convolucional 3 CNN			
DÍA: Tipo2 vs Tipo3			
Parámetros	Oclusión (70 %)	Grad-CAM (70 %)	Atención (70 %)
RP	0.1645	0.3696	<b>1.5287E-06</b>
SpecBand	0.0988	0.0889	<b>0.0085</b>
SpecCrestFac	<b>0.0057</b>	0.7332	0.2276
SpecEn	0.1398	0.2276	<b>0.0016</b>
SpecFlat	0.1398	0.1306	<b>4.5549E-05</b>
SpecFlux	<b>1.8869E-05</b>	0.1225	0.12645
SpecRenyiEn	0.1549	0.0975	<b>1.7350E-08</b>
SpecRolloff	<b>0.0454</b>	0.1037	<b>5.7723E-06</b>

TABLA 5.10: *p-valores* de las características frecuenciales para Tipo2 vs Tipo3, en el periodo diurno para la capa convolucional 3 de la CNN

Los *p-valores* más bajos se concentran en el método de atención, tanto durante el día como la noche, donde se obtienen valores notablemente reducidos. En el periodo nocturno también hay numerosos valores pequeños, destacando especialmente los resultados con atención y oclusión, mientras que Grad-CAM continúa mostrando valores más altos en general.

#### 5.1.2.2 Tipo2 vs Tipo3: CNN CAPA CONVOLUCIONAL 2

En la tabla 5.11 se representan los *p-valores* de las características frecuenciales de cada uno de los métodos para Tipo2 vs Tipo3, en el periodo nocturno para la capa convolucional 2 de la CNN, destacando aquellos *p-valores* menores de 0.05.

Capa convolucional 2 CNN			
NOCHE: Tipo2 vs Tipo3			
Parámetros	Oclusión (60 %)	Grad-CAM (80 %)	Atención (60 %)
RP	<b>0.0099</b>	0.1754	0.6605
SpecBand	<b>0.0360</b>	0.5159	0.1044
SpecCrestFac	0.2901	0.8382	<b>0.0239</b>
SpecEn	0.6759	0.7059	0.2423
SpecFlat	<b>3.1271E-04</b>	0.0705	<b>0.0299</b>
SpecFlux	0.5701	<b>0.0232</b>	0.0953
SpecRenyiEn	0.8868	<b>0.0574</b>	0.1246
SpecRolloff	0.1155	0.0574	<b>1.6140E-08</b>

TABLA 5.11: *p-valores* de las características frecuenciales para Tipo2 vs Tipo3, en el periodo nocturno para la capa convolucional 2 de la CNN

En la tabla 5.12 se representan los *p-valores* de las características frecuenciales de cada uno de los métodos para Tipo2 vs Tipo3, en el periodo diurno para la capa convolucional 2 de la CNN, destacando aquellos *p-valores* menores de 0.05.

Capa convolucional 2 CNN			
DÍA: Tipo2 vs Tipo3			
Parámetros	Oclusión (60 %)	Grad-CAM (70 %)	Atención (60 %)
RP	<b>0.0023</b>	0.3223	<b>5.6838E-04</b>
SpecBand	0.1853	<b>2.1686E-06</b>	<b>2.6019E-11</b>
SpecCrestFac	0.7336	0.2061	<b>0.0034</b>
SpecEn	0.1359	<b>2.5482E-07</b>	<b>5.9427E-07</b>
SpecFlat	0.0762	0.3464	0.8114
SpecFlux	0.4752	0.5080	<b>9.1552E-06</b>
SpecRenyiEn	0.0513	<b>0.0328</b>	0.6657
SpecRolloff	0.4001	<b>0.0338</b>	0.2016

TABLA 5.12: *p-valores* de las características frecuenciales para Tipo2 vs Tipo3, en el periodo diurno para la capa convolucional 2 de la CNN

Se observan *p-valores* bajos tanto de noche como de día. Por la noche, atención presenta algunos valores considerablemente reducidos. Sin embargo, en el periodo diurno, atención y Grad-CAM muestran *p-valores* particularmente bajos en varias métricas, indicando que los resultados varían notablemente entre los métodos y los periodos.

### 5.1.2.3 Tipo2 vs Tipo3: CNN CAPA CONVOLUCIONAL 1

En la tabla 5.13 se representan los *p-valores* de las características frecuenciales de cada uno de los métodos para Tipo2 vs Tipo3, en el periodo nocturno para la capa convolucional 1 de la CNN, destacando aquellos *p-valores* menores de 0.05.

Capa convolucional 1 CNN			
NOCHE: Tipo2 vs Tipo3			
Parámetros	Oclusión (80 %)	Grad-CAM (80 %)	Atención (80 %)
RP	<b>0.0049</b>	0.1834	0.2663
SpecBand	<b>0.0122</b>	<b>0.0203</b>	<b>1.7661E-04</b>
SpecCrestFac	<b>5.3998E-05</b>	0.4541	<b>5.0648E-05</b>
SpecEn	0.1917	0.2679	<b>3.4847E-05</b>
SpecFlat	<b>0.0131</b>	<b>0.0263</b>	<b>6.2437E-05</b>
SpecFlux	0.3028	0.1390	0.1500
SpecRenyiEn	<b>0.0172</b>	<b>0.0040</b>	-
SpecRolloff	<b>5.6765E-04</b>	<b>1.9996E-04</b>	0.0663

TABLA 5.13:  $p$ -valores de las características frecuenciales para Tipo2 vs Tipo3, en el periodo nocturno para la capa convolucional 1 de la CNN

En la tabla 5.14 se representan los  $p$ -valores de las características frecuenciales de cada uno de los métodos para Tipo2 vs Tipo3, en el periodo diurno para la capa convolucional 1 de la CNN, destacando aquellos  $p$ -valores menores de 0.05.

Capa convolucional 1 CNN			
DÍA: Tipo2 vs Tipo3			
Parámetros	Oclusión (60 %)	Grad-CAM (80 %)	Atención (70 %)
RP	0.3966	<b>8.5346E-05</b>	0.2120
SpecBand	0.5734	<b>1.9302E-08</b>	0.4248
SpecCrestFac	0.6605	<b>0.0010</b>	0.5567
SpecEn	0.5001	<b>1.9269E-13</b>	<b>0.0317</b>
SpecFlat	0.0551	<b>0.0039</b>	0.9655
SpecFlux	0.8479	<b>4.3186E-05</b>	<b>1.7737E-05</b>
SpecRenyiEn	<b>0.0004</b>	<b>0.0013</b>	-
SpecRolloff	0.8141	0.3192	0.5989

TABLA 5.14:  $p$ -valores de las características frecuenciales para Tipo2 vs Tipo3, en el periodo diurno para la capa convolucional 1 de la CNN

Durante el periodo nocturno, los tres métodos reflejan varios valores significativos, especialmente atención, que destaca con una mayor cantidad de  $p$ -valores bajos. oclusión y Grad-CAM también muestran resultados relevantes, aunque en menor medida. En el periodo diurno, Grad-CAM presenta una concentración notable de valores significativos, mientras que atención mantiene algunos, y oclusión muestra menos relevancia.

#### 5.1.2.4 Tipo2 vs Tipo3: Transformer

En la tabla 5.15 se representan los  $p$ -valores de las características frecuenciales de cada uno de los métodos para Tipo2 vs Tipo3, en el periodo nocturno para el Transformer, destacando aquellos  $p$ -valores menores de 0.05.

<i>Transformer</i>				
<i>NOCHE: Tipo2 vs Tipo3</i>				
Parámetros	Oclusión (60 %)	Grad-CAM (60 %)	At. Capa 1 (70 %)	At. Capa2 (80 %)
RP	<b>1.0170E-09</b>	0.7902	0.1494	0.2047
SpecBand	0.6420	0.1794	<b>0.0095</b>	<b>0.0328</b>
SpecCrestFac	<b>5.1599E-04</b>	0.7329	0.0809	0.1754
SpecEn	<b>4.7749E-19</b>	0.7712	0.3242	0.1960
SpecFlat	0.1674	0.1118	<b>0.0394</b>	0.3376
SpecFlux	0.2080	0.9552	0.2914	0.7512
SpecRenyiEn	-	0.3314	0.0821	0.0626
SpecRolloff	<b>3.3800E-04</b>	0.0727	<b>0.0343</b>	0.1342

TABLA 5.15: *p-valores* de las características frecuenciales para *Tipo2 vs Tipo3*, en el periodo nocturno para el *Transformer*

En la tabla 5.16 se representan los *p-valores* de las características frecuenciales de cada uno de los métodos para *Tipo2 vs Tipo3*, en el periodo diurno para el *Transformer*, destacando aquellos *p-valores* menores de 0.05.

<i>Transformer</i>				
<i>DÍA: Tipo2 vs Tipo3</i>				
Parámetros	Oclusión (80 %)	Grad-CAM (60 %)	At. Capa 1 (70 %)	At. Capa2 (60 %)
RP	<b>3.0048E-10</b>	0.5819	0.8576	0.8673
SpecBand	<b>1.9103E-21</b>	0.1897	0.4922	<b>0.0383</b>
SpecCrestFac	<b>5.9300E-05</b>	0.5476	0.7807	0.9666
SpecEn	<b>4.2393E-09</b>	0.5734	0.2844	0.9753
SpecFlat	<b>7.4624E-04</b>	0.1523	0.7796	0.2040
SpecFlux	<b>6.4879E-05</b>	0.2817	0.2703	0.7217
SpecRenyiEn	<b>0.0018</b>	0.6381	0.5834	<b>0.0470</b>
SpecRolloff	<b>1.2518E-27</b>	0.8523	0.4071	<b>0.0262</b>

TABLA 5.16: *p-valores* de las características frecuenciales para *Tipo2 vs Tipo3*, en el periodo diurno para el *Transformer*

En ambos periodos, oclusión destaca al mostrar varios *p-valores* extremadamente bajos, lo que refuerza su papel como método más efectivo en esta comparación. Por otro lado, los métodos restantes, presentan ciertos valores relevantes, pero en general sus resultados son más moderados en comparación con oclusión.

## 5.2 ANÁLISIS DE LOS RESULTADOS

En este apartado se van a analizar los resultados obtenidos tras la extracción y comparación de características espectrales entre los distintos subtipos de TDAH. Como bien se ha podido observar en las tablas previas, el volumen de datos generados es considerable, lo que puede

dificultar una interpretación clara y directa. Con el objetivo de visualizar mejor estas diferencias y facilitar la interpretación de los resultados, a continuación se van a mostrar las características espectrales que han obtenido los mejores  $p$ -valores, para cada comparativa de subtipos.

En la tablas 5.17 y 5.18 se muestran las 10 características espectrales con  $p$ -valor más bajo para *Tipo1 vs Tipo3* y *Tipo2 vs Tipo3* respectivamente:

Características con mejores $p$ -valores					
<i>Tipo1 vs Tipo3</i>					
Característica	Método	Red/Capa	Porcentaje	Periodo	$p$ -valor
SpecFlux	Atención	CNN Capa 1	60 %	Día	3.8059E-07
SpecBand	Oclusión	CNN Capa 3	80 %	Noche	8.8698E-07
SpecEn	Oclusión	CNN Capa 3	80 %	Noche	3.7462E-06
SpecFlat	Oclusión	CNN Capa 3	80 %	Noche	4.0881E-06
SpecRenyiEn	Oclusión	CNN Capa 3	80 %	Noche	4.0881E-06
SpecRolloff	Atención	CNN Capa 1	60 %	Noche	1.0465E-05
SpecBand	Atención	CNN Capa 1	60 %	Noche	1.7817E-05
SpecEn	Atención	CNN Capa 2	70 %	Día	3.2754E-05
SpecFlux	Atención	CNN Capa 1	60 %	Noche	4.6382E-05
SpecRolloff	Atención Capa 2	<i>Transformer</i>	80 %	Noche	5.2753E-04

TABLA 5.17: 10 características con mejores  $p$ -valores para *Tipo1 vs Tipo3*

Características con mejores $p$ -valores					
<i>Tipo2 vs Tipo3</i>					
Característica	Método	Red/Capa	Porcentaje	Periodo	$p$ -valor
SpecRolloff	Oclusión	<i>Transformer</i>	80 %	Día	1.2518E-27
SpecBand	Oclusión	<i>Transformer</i>	80 %	Día	1.9103E-21
SpecEn	Oclusión	<i>Transformer</i>	60 %	Noche	4.7749E-19
SpecEn	Grad-CAM	CNN Capa 1	80 %	Día	1.9269E-13
SpecBand	Atención	CNN Capa 2	60 %	Día	2.6019E-11
RP_feat	Oclusión	<i>Transformer</i>	80 %	Día	3.0048E-10
SpecEn	Oclusión	<i>Transformer</i>	80 %	Día	1.0170E-09
SpecRolloff	Atención	CNN Capa 2	60 %	Noche	1.6140E-08
SpecRenyiEn	Atención	CNN Capa 3	70 %	Día	1.7350E-08
SpecBand	Grad-CAM	CNN Capa 1	80 %	Día	1.9302E-08

TABLA 5.18: 10 características con mejores  $p$ -valores para *Tipo2 vs Tipo3*

En la comparación entre Tipo1 y Tipo3 destaca con claridad la CNN entre los mejores  $p$ -valores, especialmente en las capas 1 y 3, donde tanto los métodos de atención como los de oclusión en periodo nocturno aparecen de forma más recurrente. Esto pone de manifiesto que, para diferenciar el subtipo1 del 3, las señales captadas en las fases tempranas y profundas de la CNN durante la noche resultan especialmente informativas. La presencia combinada de diferentes capas y métodos indica que la red es capaz de identificar patrones relevantes en distintos niveles de abstracción.

En el caso de Tipo2 frente a Tipo3, cobra protagonismo el *Transformer* con la técnica de oclusión aplicada en periodo diurno, lo que sugiere que las dinámicas espectrales a lo largo del día ofrecen una mayor capacidad para separar estos dos subtipos cuando se utiliza esta arquitectura. Este patrón indica que el *Transformer* puede estar captando relaciones temporales y espectrales más complejas que resultan útiles para distinguir entre el subtipo combinado y el hiperactivo.

En ambas comparativas, el umbral del 80 % aparece de forma más frecuente, lo que sugiere que un enmascaramiento más restrictivo permite resaltar con mayor claridad las regiones más relevantes para la clasificación.

Por otro lado, los *p-valores* observados en la comparación *Tipo2 vs Tipo3* tienden a ser más bajos que los obtenidos en *Tipo1 vs Tipo3*, lo que indica que estos subtipos presentan diferencias más marcadas entre sí. Esto puede deberse a mayores diferencias en la actividad entre los subtipos combinado e hiperactivo, en comparación con la que existe entre los subtipos inatento e hiperactivo, facilitando así la identificación de patrones distintivos a través de las características espectrales. Este hecho es interesante ya que aunque el Tipo2 corresponde al subtipo combinado, las diferencias observadas con el Tipo3 (hiperactivo) son más evidentes que las que se observan entre Tipo1 (inatento) y Tipo3, lo que sugiere que la combinación de síntomas en el Tipo2 genera patrones espectrales más distintivos frente al subtipo hiperactivo.

Por último, aunque los *Transformers* han mostrado una mayor precisión global que las CNN en las tareas de clasificación, como se vio en la tabla 4.4, no se observa una dominancia clara en términos de *p-valores* más bajos. Ambos enfoques, CNN y *Transformer*, aportan información valiosa y complementaria, por lo que realizar un análisis conjunto utilizando diferentes arquitecturas y métodos interpretables resulta una estrategia adecuada para obtener resultados más completos y robustos en la caracterización de los subtipos de TDAH.

### 5.2.1 ANÁLISIS MEDIANTE *boxplots*

Para completar el análisis, se van a utilizar *boxplots* que permitirán evaluar visualmente las diferencias entre las características espectrales de los subtipos. Se mostrarán y representarán los *boxplots* de las características mostradas anteriormente en las tablas 5.17 y 5.18. Asimismo, en el anexo del presente trabajo (Anexo A), se muestran todos los *boxplots* de aquellas características con *p-valores* inferiores a 0.05.

Un *boxplot* o diagrama de caja es una representación gráfica empleada en estadística y análisis de datos para mostrar la distribución de un conjunto de datos numéricos a través de sus cuartiles. En los *boxplots* se distinguen las siguientes partes:

- **Caja:** Representa la distancia entre el primer cuartil (Q1) y el tercer cuartil (Q3). La longitud de la caja refleja la dispersión de los datos en la zona central de la distribución.
- **Mediana:** Es la línea que divide la caja en dos partes iguales, y corresponde al valor central de los datos.
- **Bigotes:** Son las líneas que se extienden desde la caja hacia los extremos superior e inferior, y representan los valores de los datos por debajo de Q1 y por encima de Q3, respectivamente.

- **Outliers:** Son los puntos individuales que se encuentran fuera de los límites establecidos por los bigotes, siendo considerados valores atípicos dentro de la distribución de los datos.

### 5.2.1.1 BOXPLOTS: TIPO1 VS TIPO3

En las figura 5.1 se presentan los *boxplots* de las 10 mejores características para *Tipo1 vs Tipo3* (véase 5.17):

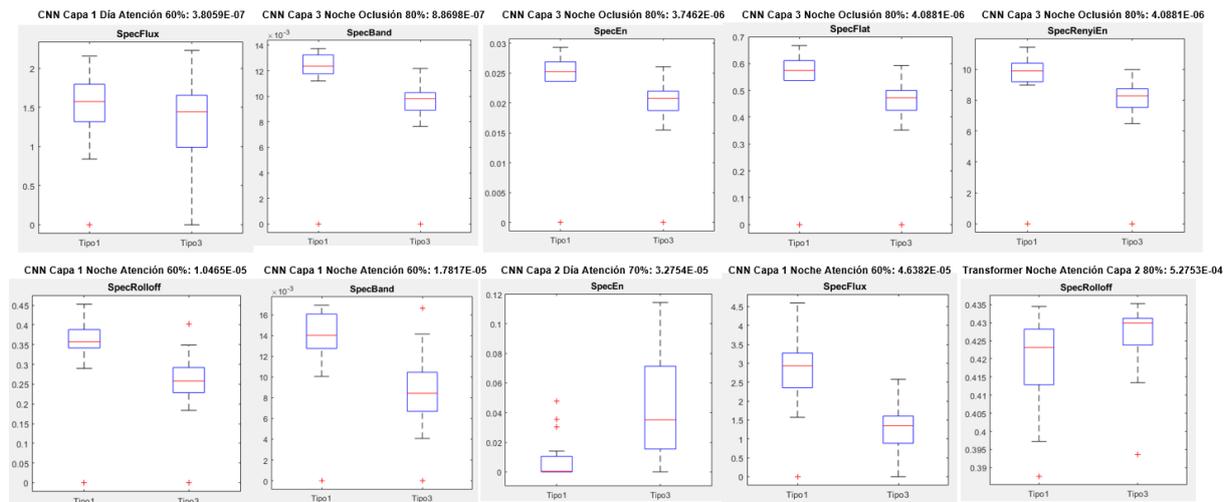


FIGURA 5.1: *Boxplots* de las 10 mejores características para *Tipo1 vs Tipo3*

A continuación, se discuten los hallazgos más destacados:

- **SpecFlux** (Atención, CNN capa 1, 60 %, día y noche): presentan valores mayores para Tipo1. SpecFlux cuantifica los cambios abruptos en la energía espectral. Un mayor SpecFlux sugiere una actividad espectral más variable en el tiempo para el Tipo1, relacionado con fluctuaciones en la atención sostenida.
- **SpecBand, SpecEn, SpecFlat, SpecRenyiEn y SpecBand** (Oclusión, CNN capa 3 y 1, 80 %, noche): muestran valores mayores para Tipo1, lo que sugiere mayor dispersión, complejidad e incertidumbre espectral. SpecBand indica un espectro más amplio, SpecRolloff refleja la presencia de frecuencias altas, y SpecFlat una distribución más uniforme de energía. Estas características, junto con la aleatoriedad capturada por SpecEn y SpecRenyiEn, apuntan a un patrón más caótico y menos organizado en Tipo1, posiblemente relacionado con dificultades para mantener la atención sostenida.
- **SpecEn** (Atención, CNN capa 2, 70 %, día) y **SpecRolloff** (Atención, *Transformer* capa 2, 80 %, noche): en estos casos, los valores son más altos para Tipo3. Esto sugiere que, según la configuración específica del mecanismo de atención, pueden resaltarse patrones distintos. Una mayor aleatoriedad (SpecEn) o presencia de frecuencias altas (SpecRolloff) en Tipo3 podría reflejar su impulsividad, visible solo en ciertos contextos analíticos que destaquen aspectos diferentes del espectro.

En conjunto, las características espectrales que diferencian a Tipo1 de Tipo3 tienden a destacar más en Tipo1, especialmente en términos de inestabilidad y complejidad espectral. No obstante, ciertas configuraciones específicas, como las basadas en mecanismos de atención, permiten identificar patrones más intensos o variables en Tipo3, posiblemente asociados a su impulsividad. Estas diferencias se alinean con los perfiles sintomáticos de ambos subtipos y se ven respaldadas por *p*-valores extremadamente bajos, lo que refuerza la solidez de los hallazgos.

### 5.2.1.2 BOXPLOTS: TIPO2 VS TIPO3

En las figura 5.2 se presentan los *boxplots* de las 10 mejores características para *Tipo2 vs Tipo3* (véase 5.18):

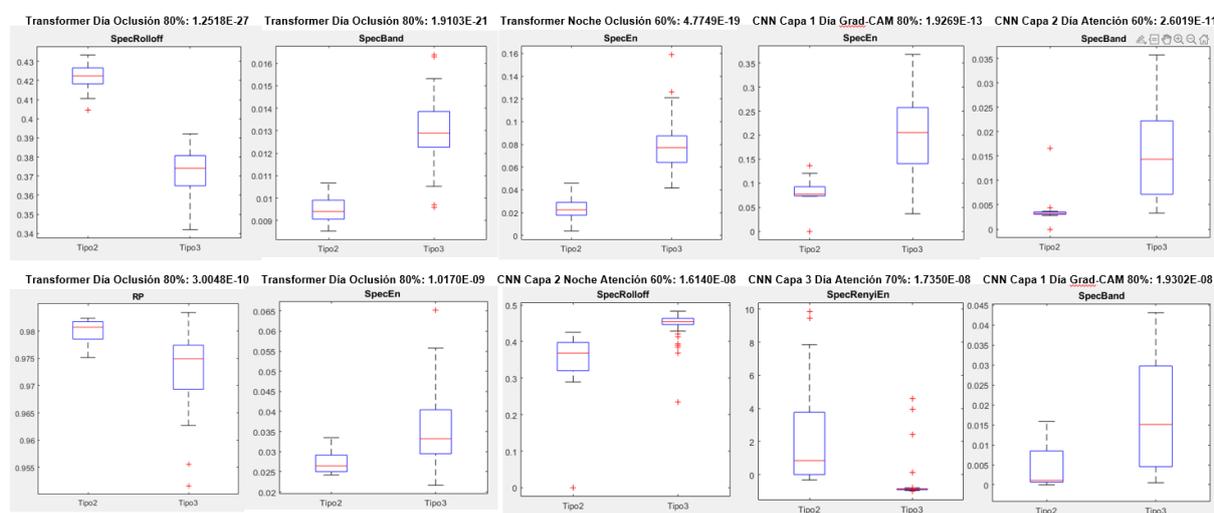


FIGURA 5.2: *Boxplots* de las 10 mejores características para *Tipo2 vs Tipo3*

A continuación, se presentan las principales conclusiones:

- SpecRolloff (Oclusión, *Transformer*, 80 %, día): se trata del *p*-valor más bajo de la comparativa ( $10^{-27}$ ), presenta valores significativamente mayores para Tipo2. Un SpecRolloff más alto indica que la mayor parte de la energía se concentra en frecuencias más elevadas. Esto sugiere una mayor presencia de componentes de alta frecuencia para el Tipo2, posiblemente por su combinación de síntomas de inatención e hiperactividad.
- SpecBand y SpecEn (Oclusión, *Transformer*, día y noche): muestran valores más altos para Tipo3. Esto indica una mayor anchura de banda y entropía espectral, lo que podría relacionarse con una conducta más impulsiva o inestable, rasgo típico del subtipo hiperactivo.
- RP (Oclusión, *Transformer*, 80 %, día): mayor para Tipo2. La Potencia Relativa (RP) indica la proporción de energía alterna. Un RP mayor sugiere mayor variabilidad en los patrones de actividad del Tipo2, reforzando la idea de un perfil mixto con patrones espectrales más amplios.

- SpecEn y SpecBand (Atención o Grad-CAM sobre CNNs, capa 1 y 2; y Oclusión, *Transformers*): son considerablemente más elevadas para Tipo3. Un mayor SpecEn y SpecBand para el Tipo3 refuerza la mayor variabilidad o dispersión en su actividad espectral, reflejando una hiperactividad desorganizada.
- SpecRenyiEn (Atención, CNN, capa 3, 70 %, día): muestra valores mayores para Tipo2. Una SpecRenyiEn más alta sugiere mayor incertidumbre o aleatoriedad espectral, indicando que el Tipo2 puede presentar mayor complejidad espectral debido a la combinación de síntomas.

Estos resultados reflejan una dualidad: algunas métricas (SpecRolloff, RP) diferencian el Tipo2 con valores superiores (altas frecuencias, mayor variabilidad), mientras que otras (SpecBand, SpecEn) identifican el Tipo3 por mayor dispersión o aleatoriedad. Esto subraya la utilidad de diversas arquitecturas y métodos para capturar los matices entre subtipos. La visualización con *boxplots* refuerza las diferencias espectrales, sugiriendo un patrón más marcado para Tipo2 y uno más desorganizado para Tipo3 fruto de la hiperactividad.

## 5.2.2 COMPARATIVA CON TRABAJOS PREVIOS

Con el objetivo de valorar la calidad y la relevancia de los resultados obtenidos en este trabajo, se ha llevado a cabo una comparación directa con los resultados presentados en el trabajo anterior, [5]. En dicho estudio se emplearon exclusivamente CNNs junto con mapas de oclusión, y las métricas utilizadas para evaluar las diferencias entre subtipos de TDAH estaban basadas en parámetros estadísticos derivados de distribuciones gaussianas. En la tabla 5.19 se muestran los *p-valores* más pequeños obtenidos en [5] para cada parámetro para las comparativas *Tipo1 vs Tipo3* y *Tipo2 vs Tipo3*:

Parámetro	<i>Tipo1 vs Tipo3</i>	<i>Tipo2 vs Tipo3</i>
$p$	0.0142	4.5822E-04
$\eta_x$	1.7325E-04	0.0372
$\eta_y$	9.5743E-05	0.0018
$\sigma_x$	3.5316E-04	4.8182E-04
$\sigma_y$	0.0019	4.2462E-05
$\rho$	0.0021	0.0841
$ p_1 - p_2 $	0.0142	1.0733E-05
$ p_1 - p_3 $	0.0695	2.1134E-04
$D_{KL}$	3.2798E-04	0.1667
$D_{KL_1}$	0.1388	0.0267
$D_{KL_2}$	0.1129	0.1520

TABLA 5.19: *P-valores* más bajos por parámetro para las comparaciones *Tipo1 vs Tipo3* y *Tipo2 vs Tipo3* obtenidos en trabajo anterior

Tal y como se puede observar, los *p-valores* obtenidos en el trabajo anterior son considerablemente más elevados que los presentados en este estudio (ver tablas 5.17 y 5.18). En este trabajo, los *p-valores* más bajos alcanzan órdenes de magnitud del  $10^{-7}$  y  $10^{-27}$  para las comparativas *Tipo1*

vs *Tipo3* y *Tipo2 vs Tipo3*, respectivamente, lo que indica una diferencia estadísticamente mucho más significativa entre los grupos.

Este claro descenso en los *p-valores* sugiere que los métodos empleados en este estudio, basados en la extracción de características espectrales y el uso de un mayor número de mapas, han permitido capturar de manera más precisa las diferencias existentes entre los subtipos de TDAH. Por tanto, se puede concluir que el enfoque propuesto en este trabajo mejora notablemente la capacidad discriminativa respecto al trabajo previo, tanto en términos de sensibilidad estadística como en la riqueza de las características analizadas.

### 5.2.3 ANÁLISIS SIN ENMASCARAMIENTO

Para finalizar esta sección, se ha llevado a cabo un análisis adicional en el que no se ha realizado el enmascaramiento de los espectrogramas. De esta manera, se podrá ver la relevancia del enmascaramiento en la obtención de unos mejores resultados.

En este análisis, se han calculado los *p-valores* de las características espectrales obtenidas de los espectrogramas, en este caso sin enmascarar sobre los mapas de oclusión en la capa convolucional 3 de la CNN.

En las tablas 5.20 y 5.21 se representan los *p-valores* de las características frecuenciales de los mapas de oclusión calculados en la capa convolucional 3 de la CNN, en este caso, sin realizar el enmascaramiento de los espectrogramas.

<i>Tipo1 vs Tipo3</i>		
Mapas de oclusión: Capa convolucional 3 CNN		
Sin enmascarar		
Parámetros	Noche	Día
RP	0.5981	0.5843
SpecBand	0.5647	0.5880
SpecCrestFac	0.2723	0.6612
SpecEn	0.4736	0.6965
SpecFlat	0.5890	0.5783
SpecFlux	0.3527	0.4515
SpecRenyiEn	0.9109	0.2885
SpecRolloff	0.5771	0.5907

TABLA 5.20: *p-valores* de las características frecuenciales de los mapas de oclusión de la CNN obtenidos en la capa convolucional 3 de la CNN sin enmascarar para *Tipo1 vs Tipo3*

<i>Tipo2 vs Tipo3</i>		
Mapas de oclusión: Capa convolucional 3 CNN		
Sin enmascarar		
Parámetros	Noche	Día
RP	0.1668	0.0985
SpecBand	0.1588	0.1658
SpecCrestFac	0.1024	0.1804
SpecEn	0.1038	0.9064
SpecFlat	0.1625	0.1636
SpecFlux	0.1253	0.1601
SpecRenyiEn	0.7188	0.1459
SpecRolloff	0.1640	0.7280

TABLA 5.21: *p-valores* de las características frecuenciales de los mapas de oclusión de la CNN obtenidos en la capa convolucional 3 de la CNN sin enmascarar para *Tipo2 vs Tipo3*

Se puede observar de manera clara que los resultados obtenidos sin la aplicación de enmascaramiento son notablemente peores, ya que los *p-valores* de las características extraídas son significativamente más altos en comparación con los obtenidos al aplicar el enmascaramiento (tablas 5.1, 5.2, 5.9 y 5.10). En particular, no se obtiene ningún *p-valor* inferior a 0.05, lo que sugiere que las características no tienen un impacto estadísticamente significativo. Este hecho resalta la importancia del enmascaramiento como un factor determinante para mejorar la calidad del análisis, al permitir la eliminación de información redundante o irrelevante en el espectrograma.

Cuando no se aplica el enmascaramiento, las características extraídas se presentan de forma más dispersa, lo que introduce interferencias en el análisis estadístico y, como consecuencia, se obtienen resultados menos precisos. En cambio, al aplicar el enmascaramiento, el análisis se enfoca en las regiones del espectrograma que contienen la información más relevante, definida esta en los mapas. Esto permite la extracción de características más fiables y útiles para diferenciar los subtipos de TDAH.

A la luz de los resultados obtenidos, se puede concluir que la técnica de enmascaramiento de espectrogramas ha demostrado ser crucial para mejorar la precisión del análisis. Por lo tanto, se establece que el enmascaramiento es una herramienta esencial para optimizar la extracción de características y, en consecuencia, para avanzar en la diferenciación de los subtipos de TDAH.

### 5.3 DISCUSIÓN DE LOS RESULTADOS

Los resultados obtenidos en este estudio han permitido avanzar en la caracterización y diferenciación de los diferentes subtipos de TDAH, concretamente del Tipo3 o hiperactivo. Además, se ha podido ver la utilidad de diferentes mapas de activación así como la efectividad del enmascaramiento en la mejora del análisis espectral. A partir de los datos recopilados y de los distintos métodos empleados, se pueden extraer varias conclusiones:

- El uso de características espectrales obtenidas a partir de los diferentes mapas de activación

ha permitido revelar diferencias claras entre los subtipos, destacando especialmente la distinción entre Tipo1 y Tipo3, y entre Tipo2 y Tipo3, siendo más marcadas las diferencias entre Tipo2 y Tipo3, a pesar de tratarse del subtipo combinado.

- En la comparativa entre Tipo1 y Tipo3, destaca la CNN, especialmente en capas tempranas y profundas durante la noche para distinguir Tipo1 de Tipo3, usando atención y oclusión. Esto sugiere que la red extrae patrones importantes en distintos niveles, que reflejan diferencias claves en los subtipos.
- En el caso de *Tipo2 vs Tipo3*, el Transformer con oclusión aplicada en el día muestra mejores resultados, indicando que sus capacidades para captar relaciones temporales y espectrales complejas permiten identificar dinámicas relevantes entre estos subtipos.
- En general, los enmascaramientos más restrictivos (80 %) resaltan con claridad las regiones relevantes para la clasificación, mejorando la detección de diferencias espectrales entre subtipos.
- Tanto CNN como Transformers aportan información complementaria y valiosa. Si bien los Transformers ofrecen mayor precisión global, ambos métodos deben combinarse para un análisis más completo y robusto.
- Los *boxplots* han sido de utilidad a la hora de observar las diferencias estadísticamente significativas facilitando la interpretación visual de las características espectrales. En concreto:
  - *Tipo1 vs Tipo3*: los *boxplots* muestran que Tipo1 se caracteriza por una mayor inestabilidad y complejidad espectral, mientras que Tipo3 presenta patrones más intensos y variables, probablemente relacionados con su impulsividad.
  - *Tipo2 vs Tipo3*: Tipo2 se define por patrones más estructurados y consistentes, mientras que Tipo3 muestran uno más caótico y desorganizado fruto de su naturaleza hiperactiva. Se destaca la importancia de emplear diversas arquitecturas y métodos para captar la complejidad del fenómeno.
- La comparación con estudios previos muestra que la metodología utilizada en este trabajo mejora significativamente la sensibilidad estadística, evidenciada en *p-valores* mucho más bajos, lo que respalda la robustez de los resultados obtenidos.
- Los resultados obtenidos al analizar la efectividad del enmascaramiento han demostrado que esta técnica es crucial para mejorar la discriminación entre los subtipos de TDAH. Sin el enmascaramiento, los *p-valores* de las características espectrales obtenidas son significativamente más altos, lo que indica una menor relevancia estadística en la diferenciación de los subtipos.

## CONCLUSIONES Y LÍNEAS FUTURAS

### 6.1 CONCLUSIONES

---

El objetivo del presente Trabajo de Fin de Máster ha sido el desarrollo de un sistema de diagnóstico objetivo del Trastorno por Déficit de Atención e Hiperactividad (TDAH), orientado especialmente a la caracterización del subtipo hiperactivo (TDAH-H), mediante el empleo de técnicas de aprendizaje profundo y herramientas de Inteligencia Artificial Explicable (XAI). Para ello, se han utilizado redes neuronales convolucionales (CNN) y *Vision Transformers* (ViT) entrenados sobre espectrogramas generados a partir de señales actigráficas de pacientes, sectorizados por subtipo y periodo (diurno/nocturno).

Con el fin de interpretar las salidas de las redes, se han aplicado distintas técnicas de XAI como mapas de oclusión, Grad-CAM y mecanismos de atención, que permiten visualizar las zonas del espectrograma más relevantes para la red. A partir de los mapas generados, se llevó a cabo un proceso de enmascaramiento y posterior análisis de características espectrales, evaluando diferencias estadísticas mediante el cálculo de p-valores y la representación visual con diagramas tipo *boxplot*.

Las conclusiones extraídas a partir del análisis de los resultados obtenidos son las siguientes:

- El uso de redes neuronales profundas como CNN y Vision Transformers, entrenadas sobre espectrogramas enmascarados de señales actigráficas, ha demostrado ser eficaz para identificar patrones distintivos asociados a los distintos subtipos de TDAH.
- La aplicación de técnicas de interpretabilidad como mapas de oclusión, Grad-CAM y atención aporta información valiosa sobre las zonas relevantes para la clasificación, facilitando una interpretación más transparente del funcionamiento interno de las redes.
- El análisis estadístico mediante características espectrales de los espectrogramas enmascarados permite identificar diferencias significativas entre los subtipos, concretamente entre el hiperactivo y los tipos inatento y combinado.
- La utilización combinada de múltiples métodos, junto con técnicas de enmascaramiento, resulta fundamental para una diferenciación más precisa y robusta entre los subtipos de

TDAH. Esta aproximación multidisciplinar permite captar patrones complementarios que ningún método por sí solo podría identificar con igual eficacia.

- La caracterización del subtipo hiperactivo frente a los subtipos inatento y combinado revela diferencias claras y consistentes en las características espectrales. Mientras que el subtipo hiperactivo (Tipo3) presenta patrones más intensos, variables y caóticos; el inatento (Tipo1) se asocia con mayor inestabilidad y complejidad espectral; y el combinado (Tipo2) con patrones más estructurados y consistentes, lo que refuerza la coherencia de los resultados obtenidos con las características clínicas y neurofisiológicas asociadas a cada subtipo.
- La metodología propuesta en este trabajo supera en sensibilidad estadística a estudios previos [5], lo que sugiere que la integración de modelos profundos con XAI y análisis espectral enmascarado constituye un avance significativo para la caracterización objetiva y diferenciada de los subtipos de TDAH, especialmente del subtipo hiperactivo.

## 6.2 LIMITACIONES Y LÍNEAS FUTURAS

---

A lo largo del desarrollo del trabajo se han identificado ciertas limitaciones, entre las que destaca la escasez de muestras de Tipo3 (TDAH-H), así como la necesidad de realizar la comparativa faltante entre TDAH-I vs TDAH-I para completar el análisis de todos los subtipos.. Esto dificulta los procesos de entrenamiento de los modelos, limitando su capacidad para generalizar correctamente las clases. No obstante, el uso de las técnicas XAI vistas, permitieron complementar el análisis obteniendo unos resultados satisfactorios.

Como línea de investigación futura, sería interesante realizar la comparativa faltante entre TDAH-I vs TDAH-I para completar el análisis de todos los subtipos, ampliar el número de muestras disponibles para mejorar la robustez de los clasificadores, o explorar técnicas de aumento de datos (*data augmentation*) que permitan equilibrar las clases sin introducir sesgos. Además, podría estudiarse la implementación de un modelo de clasificación multiclase que permitan comparar simultáneamente los tres tipos de TDAH (inatento, hiperactivo e impulsivo-combinado), así como utilizar otros modelos de red neuronal o técnicas XAI no utilizados hasta ahora.

En definitiva, los resultados obtenidos en este Trabajo de Fin de Máster resultan prometedores de cara al objetivo planteado, si bien es necesario continuar investigando para que las metodologías aplicadas puedan tener una futura utilidad en contextos clínicos y contribuir al diagnóstico automatizado de los subtipos de TDAH.

## Apéndice A

# BOXPLOTS PARA CARACTERÍSTICAS ESPECTRALES SIGNIFICATIVAS

En este apéndice se presentan todos los *boxplots* de aquellas características espectrales de interés, es decir, aquellas para las que se obtiene un *p-valor* inferior a 0.05. En primer lugar, se muestran los gráficos correspondientes a la comparativa entre los subtipos Tipo1 (inatento) y Tipo3 (hiperactivo), y a continuación, se presentan los de la comparativa entre los subtipos Tipo2 (combinado) y Tipo3 (hiperactivo).

### A.1 *Boxplots: Tipo1 vs Tipo3*

---

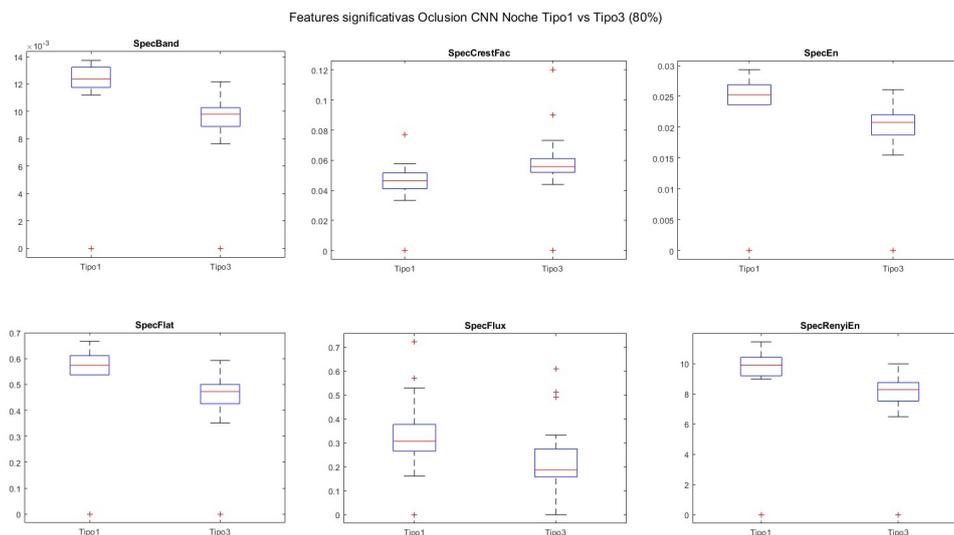


FIGURA A.1: *Boxplots* relevantes CNN Capa 3 Noche Oclusión *Tipo1 vs Tipo3*

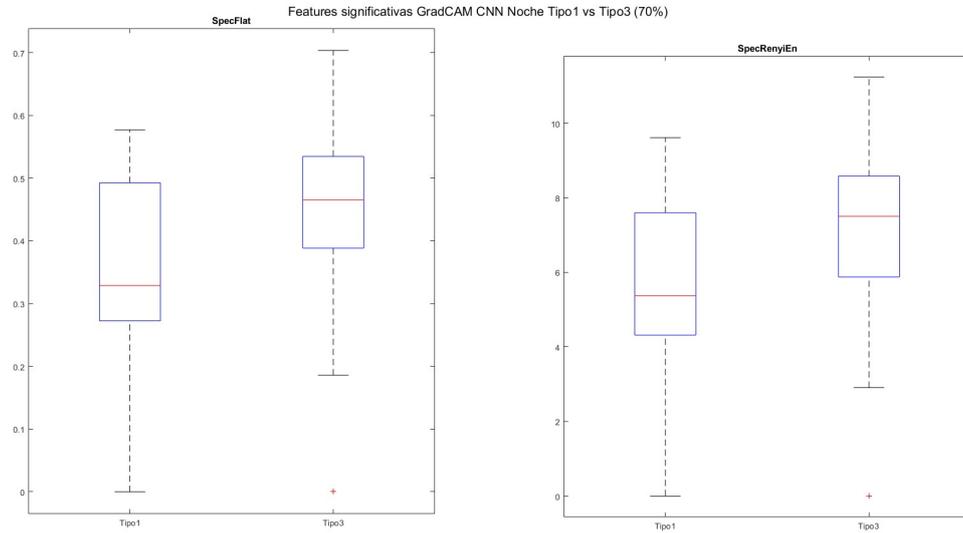


FIGURA A.2: *Boxplots* relevantes CNN Capa 3 Noche GradCAM *Tipo1 vs Tipo3*

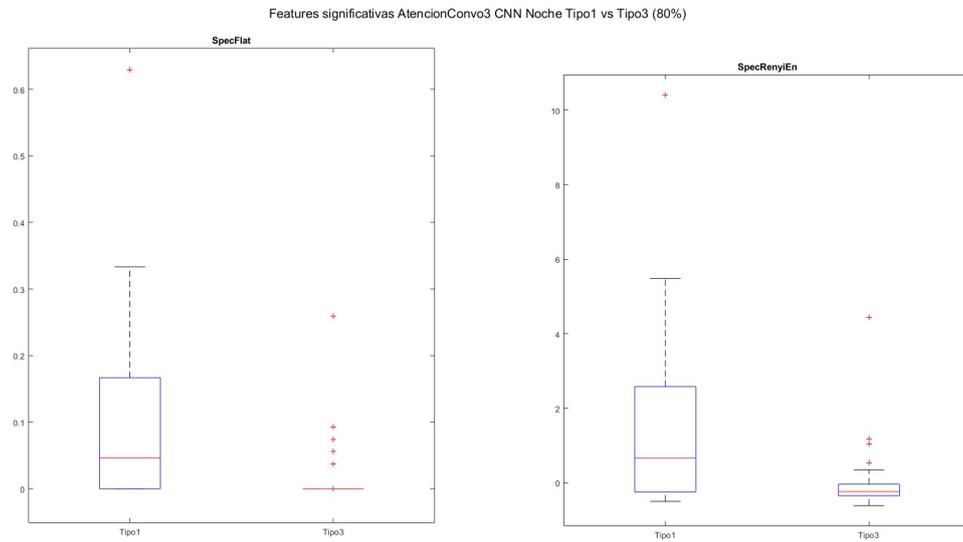


FIGURA A.3: *Boxplots* relevantes CNN Capa 3 Noche Atención *Tipo1 vs Tipo3*

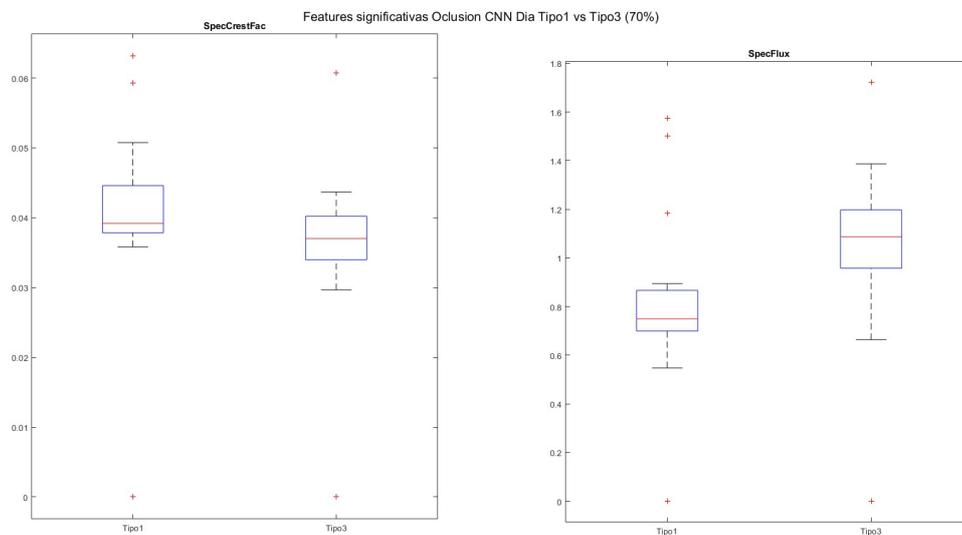


FIGURA A.4: *Boxplots* relevantes CNN Capa 3 Día Oclusión *Tipo1 vs Tipo3*

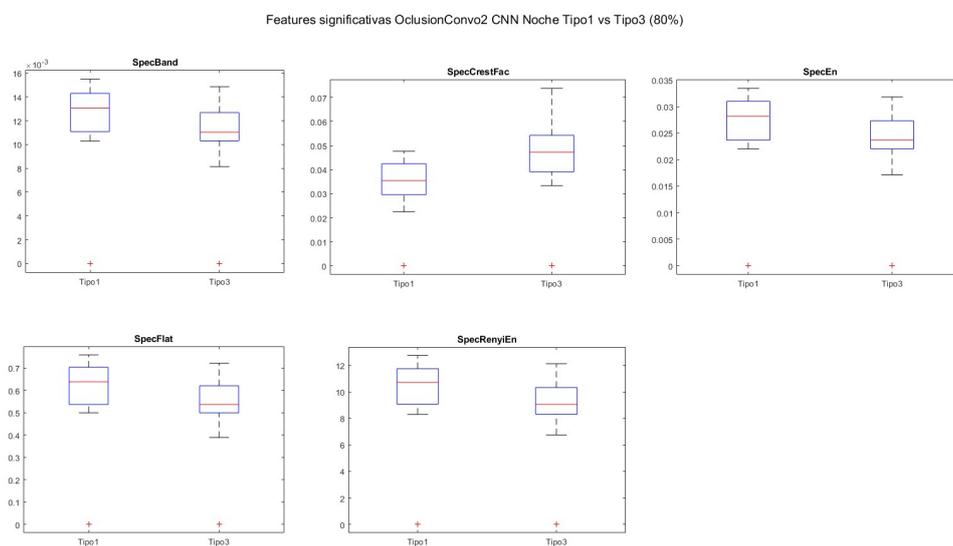


FIGURA A.5: *Boxplots* relevantes CNN Capa 2 Noche Oclusión *Tipo1 vs Tipo3*

Features significativas GradCAMConvo2 CNN Noche Tipo1 vs Tipo3 (80%)

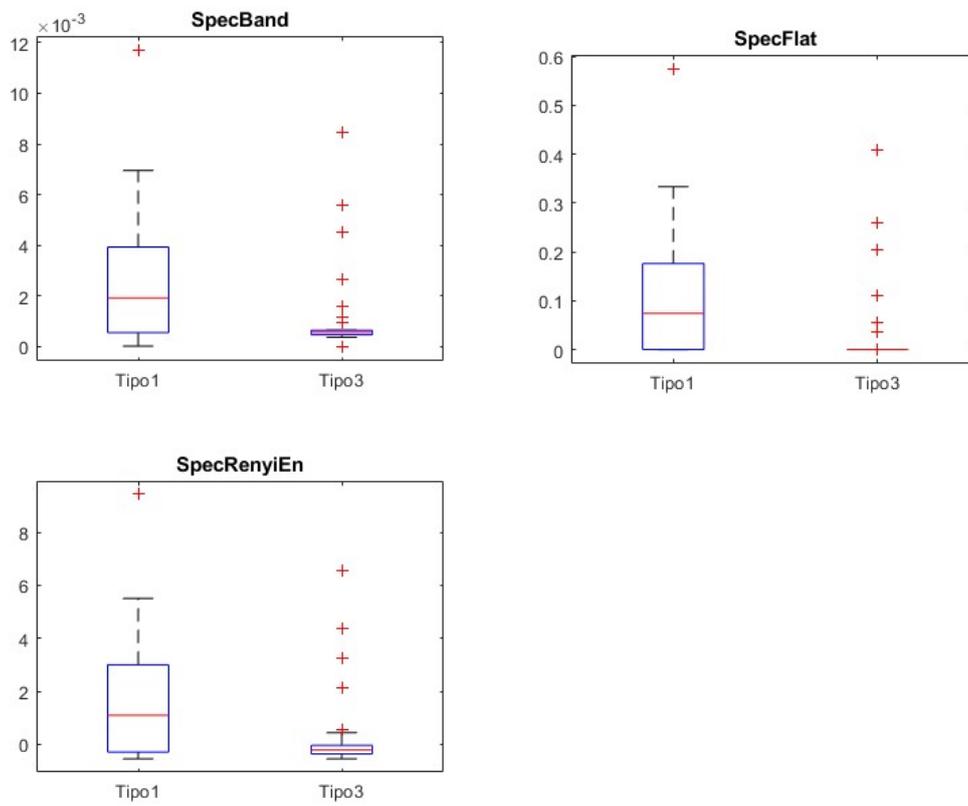


FIGURA A.6: *Boxplots* relevantes CNN Capa 2 Noche GradCAM Tipo1 vs Tipo3

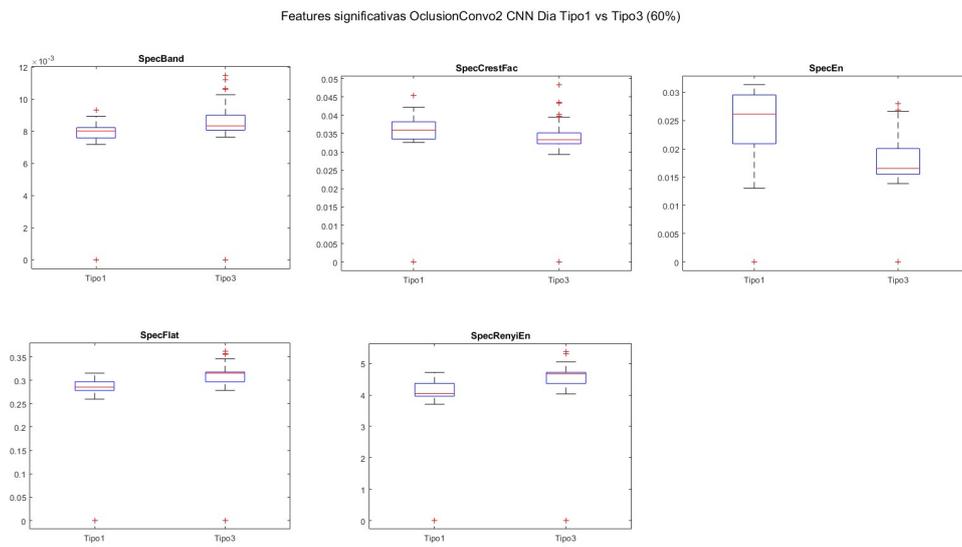


FIGURA A.7: *Boxplots* relevantes CNN Capa 2 Día Oclusión *Tipo1 vs Tipo3*

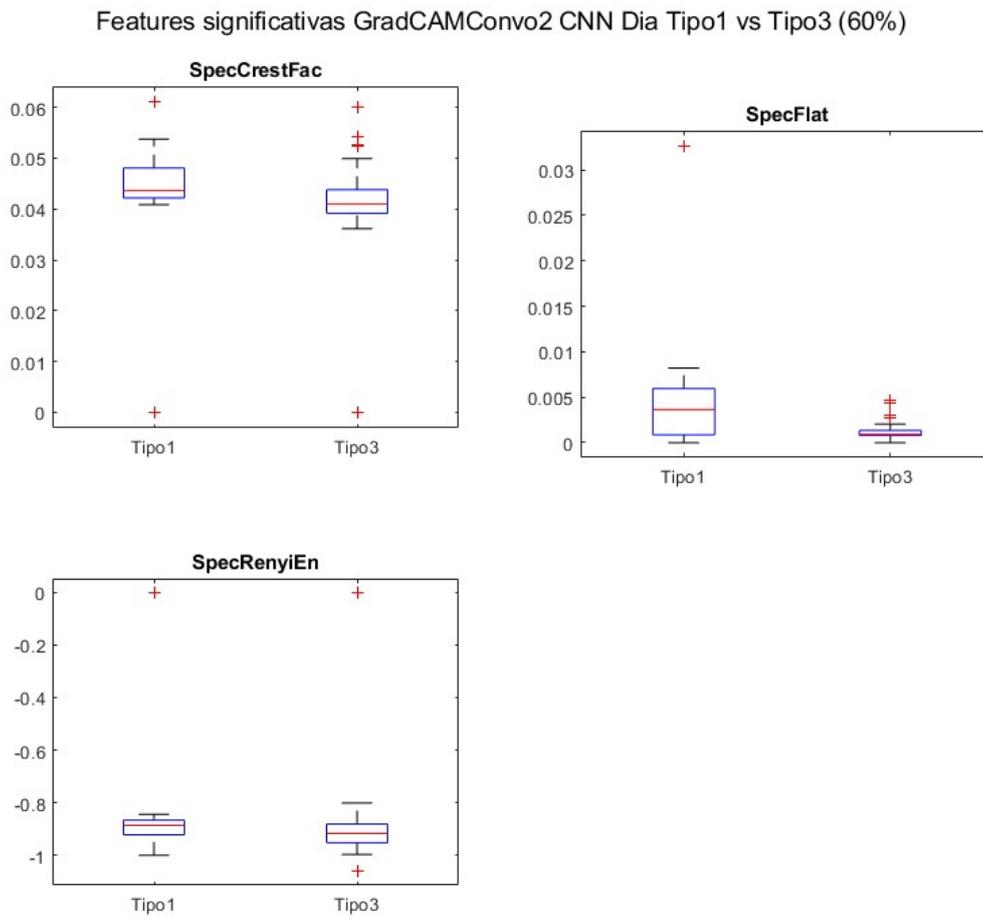


FIGURA A.8: *Boxplots* relevantes CNN Capa 2 Día GradCAM *Tipo1 vs Tipo3*

Features significativas AtencionConvo2 CNN Dia Tipo1 vs Tipo3 (70%)

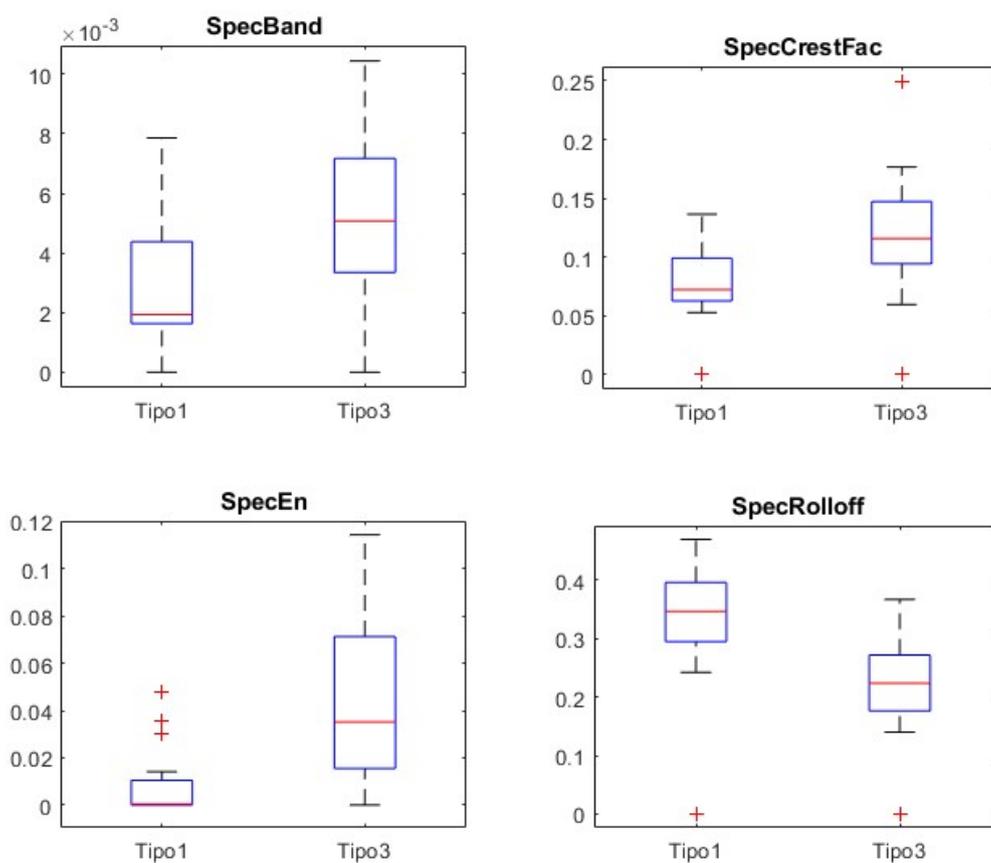


FIGURA A.9: *Boxplots* relevantes CNN Capa2 Día Atención *Tipo1 vs Tipo3*

]Secciones/Imágenes/CNNCapa1NocheOclusionTipo1Tipo3.jpg

FIGURA A.10: *Boxplots* relevantes CNN Capa 1 Noche Oclusión *Tipo1 vs Tipo3*

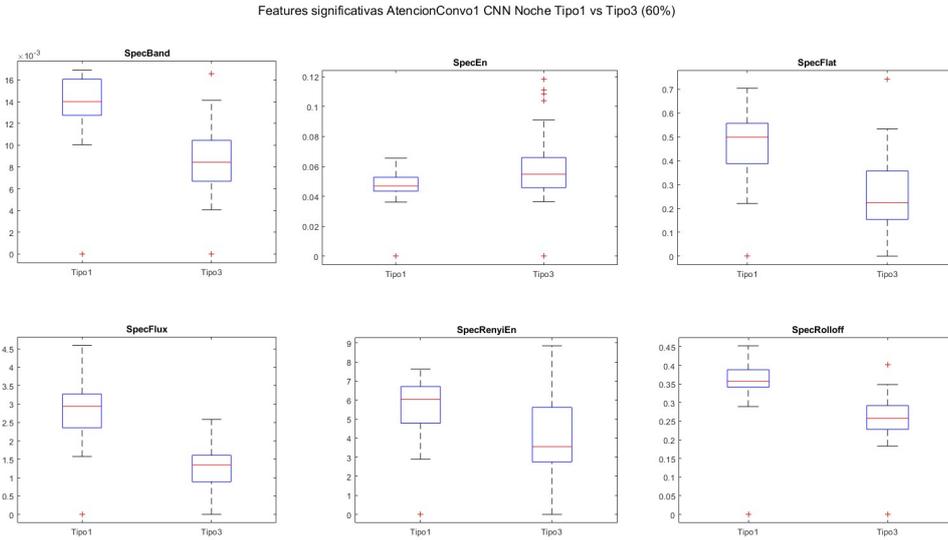


FIGURA A.11: *Boxplots* relevantes CNN Capa 1 Noche Atención *Tipo1 vs Tipo3*

Features significativas OclusionConvo1 CNN Dia Tipo1 vs Tipo3 (70%)

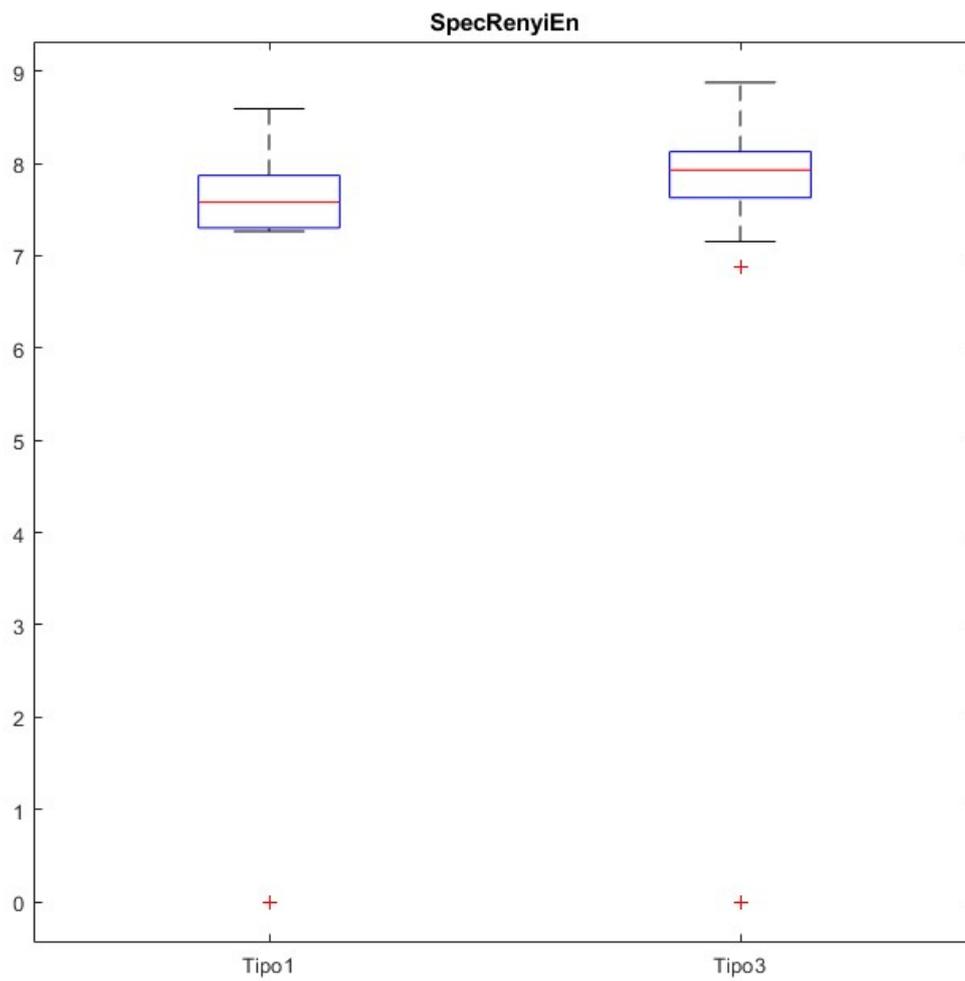


FIGURA A.12: *Boxplots* relevantes CNN Capa 1 Día Oclusión *Tipo1 vs Tipo3*

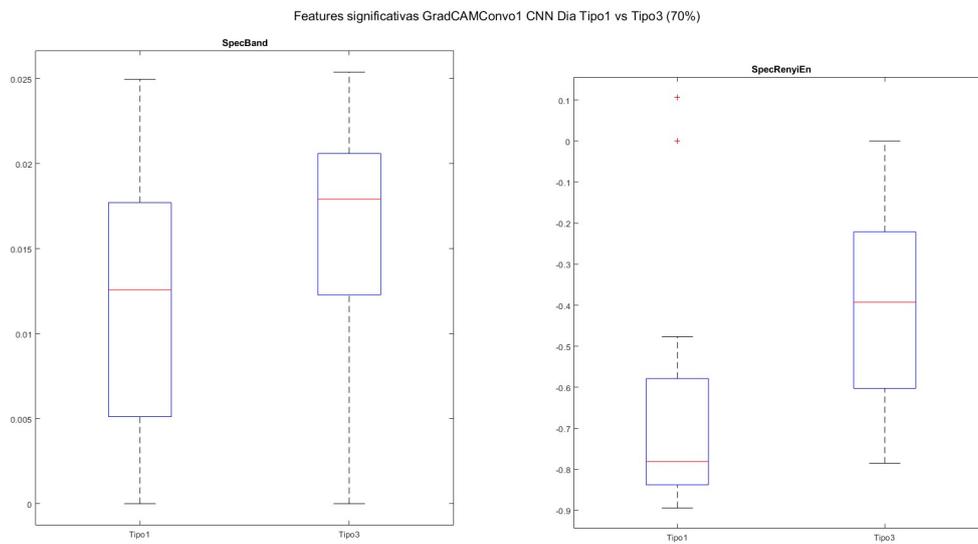


FIGURA A.13: *Boxplots* relevantes CNN Capa 1 Día GradCAM *Tipo1 vs Tipo3*

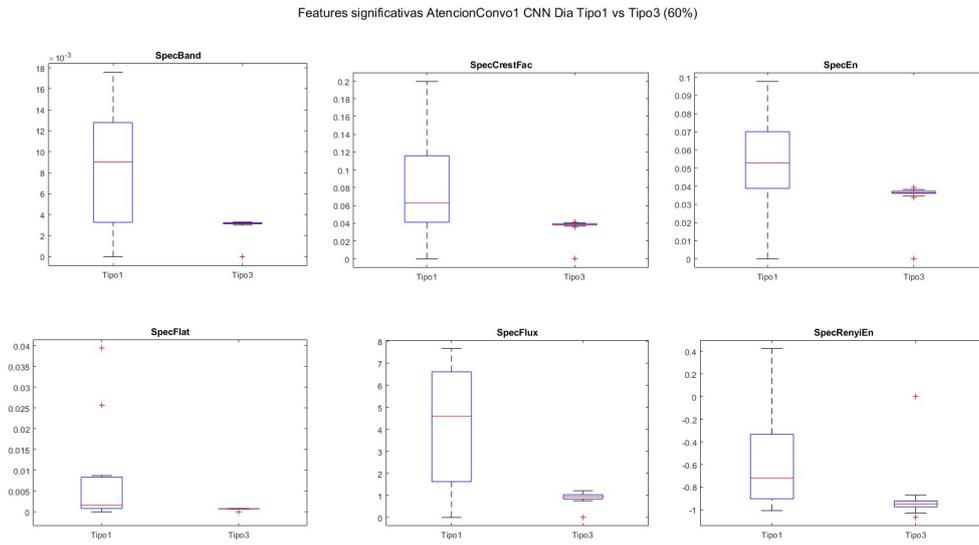


FIGURA A.14: *Boxplots* relevantes CNN Capa 1 Día Atención *Tipo1 vs Tipo3*

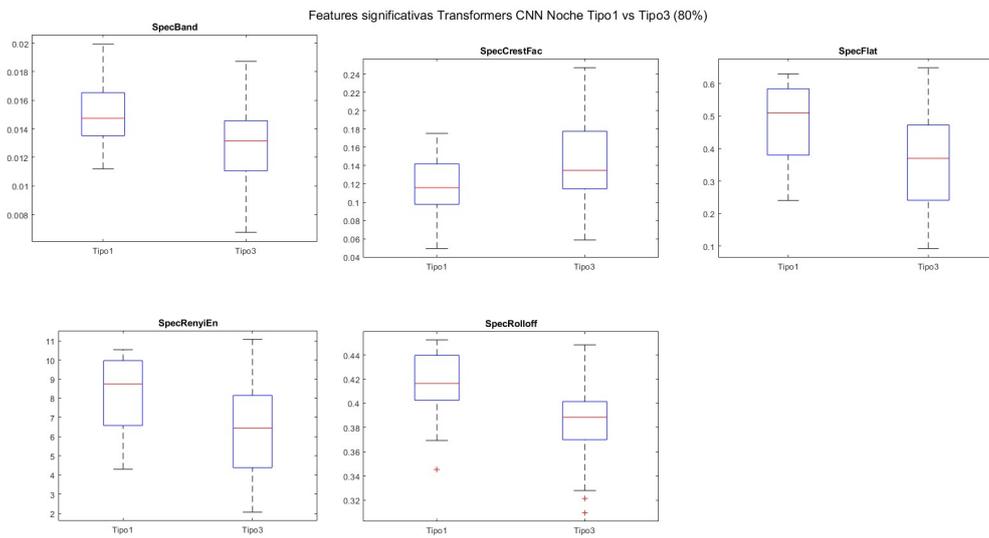


FIGURA A.15: *Boxplots* relevantes Transformers Noche Oclusión *Tipo1 vs Tipo3*

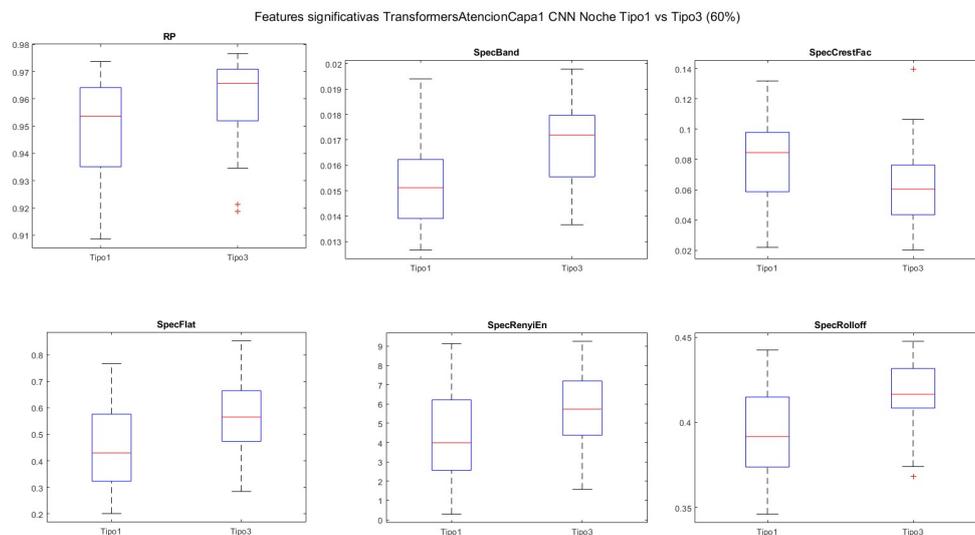


FIGURA A.16: Boxplots relevantes Transformers Noche Atención Capa 1 Tipo1 vs Tipo3

Features significativas TransformersAtencionCapa2 CNN Noche Tipo1 vs Tipo3 (80%)

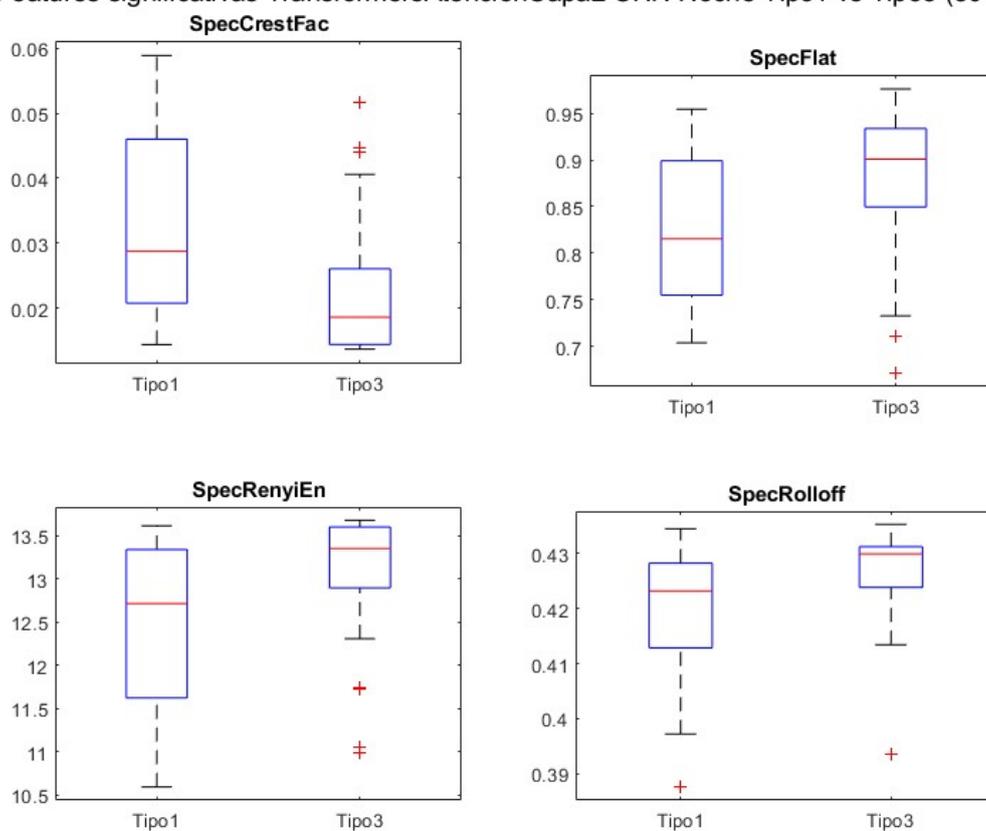


FIGURA A.17: *Boxplots* relevantes Transformers Noche Atención Capa 2 Tipo1 vs Tipo3

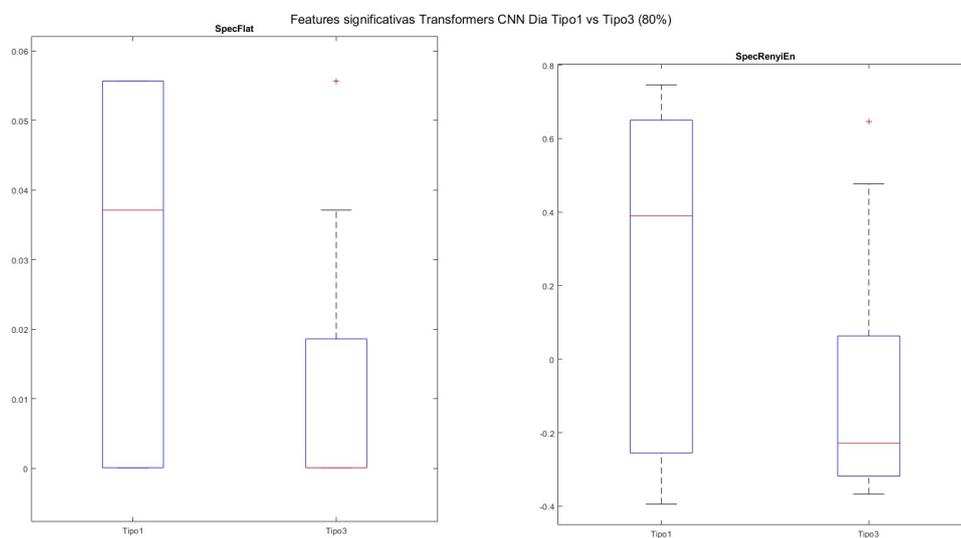


FIGURA A.18: *Boxplots* relevantes Transformers Día Oclusión *Tipo1 vs Tipo3*

Features significativas TransformersAtencionCapa1 CNN Dia Tipo1 vs Tipo3 (70%)

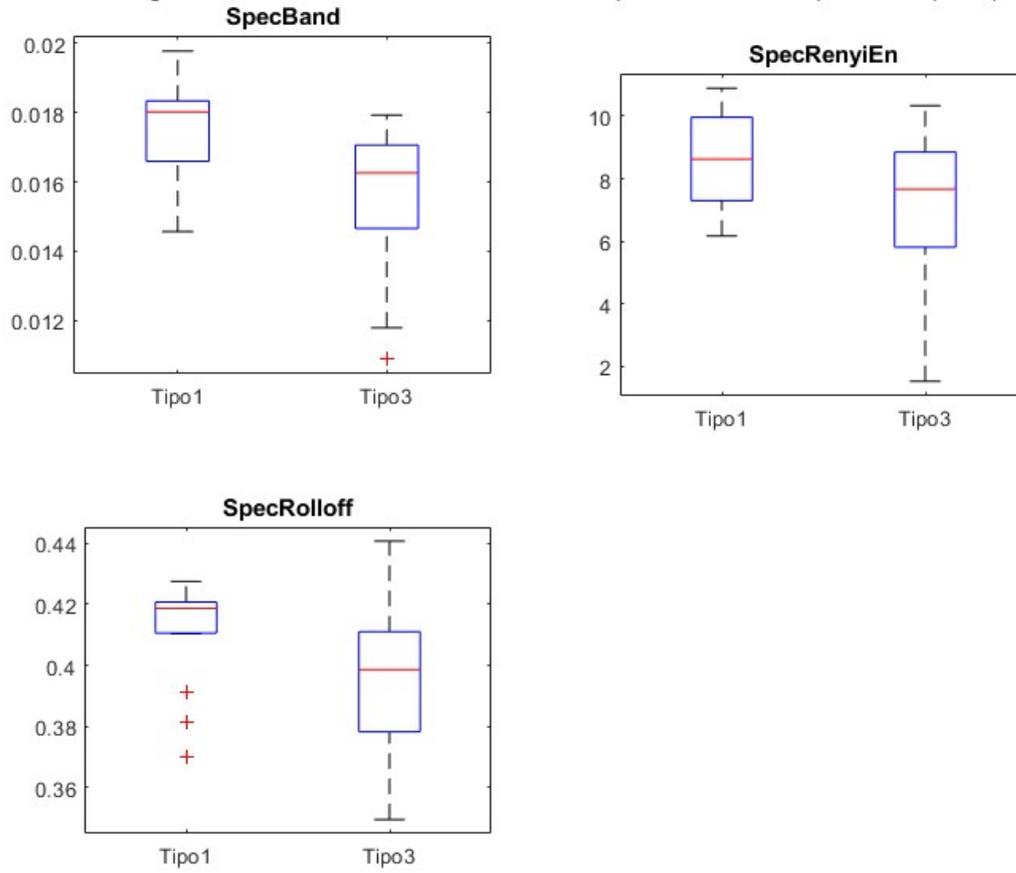


FIGURA A.19: *Boxplots* relevantes Transformers Día Atención Capa 1 *Tipo1 vs Tipo3*

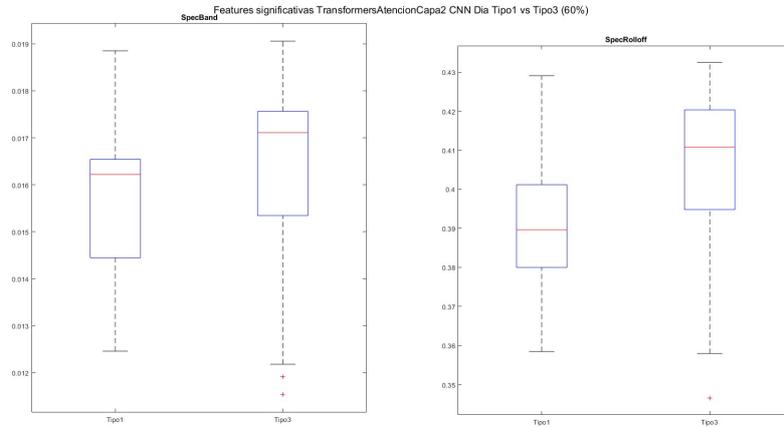


FIGURA A.20: *Boxplots* relevantes Transformers Día Atención Capa 2 *Tipo1 vs Tipo3*

## A.2 *Boxplots: Tipo2 vs Tipo3*

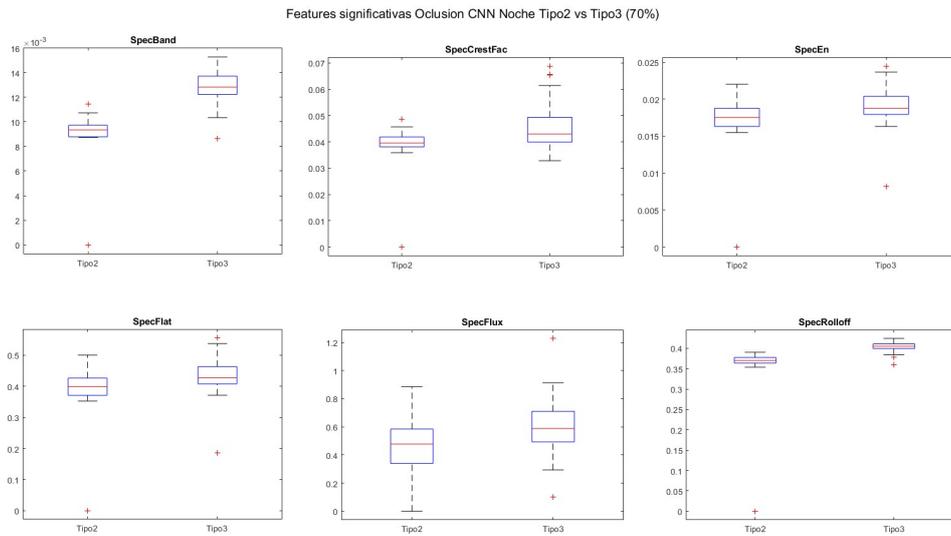


FIGURA A.21: *Boxplots* relevantes CNN Capa 3 Noche Oclusión *Tipo2 vs Tipo3*

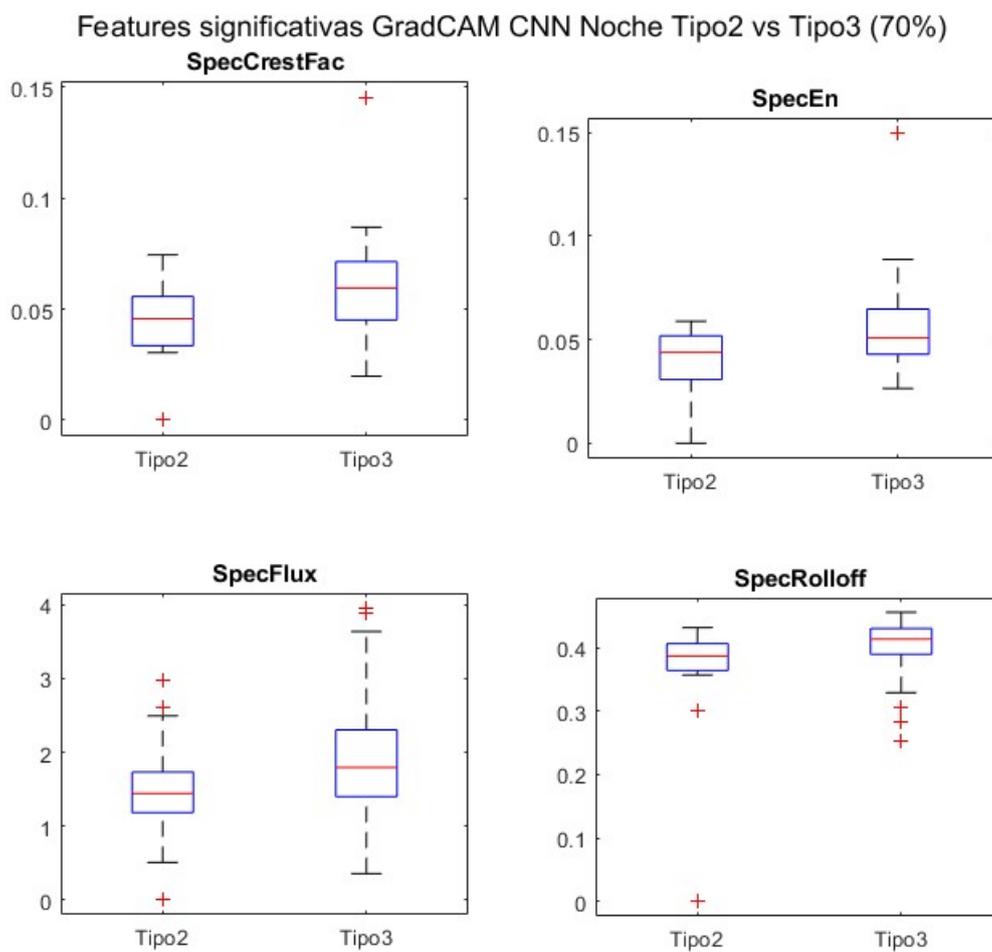


FIGURA A.22: *Boxplots* relevantes CNN Capa 3 Noche GradCAM *Tipo2 vs Tipo3*

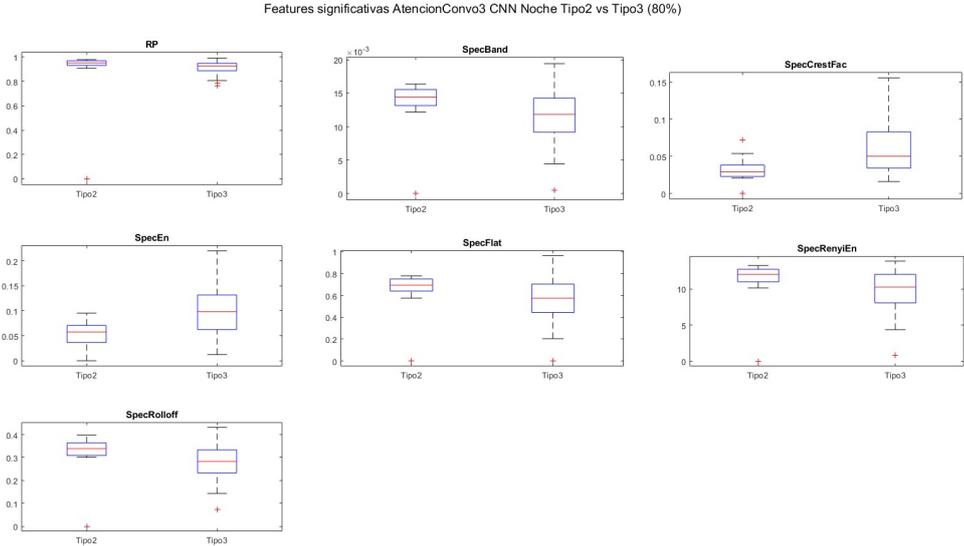


FIGURA A.23: *Boxplots* relevantes CNN Capa 3 Noche Atención *Tipo2 vs Tipo3*

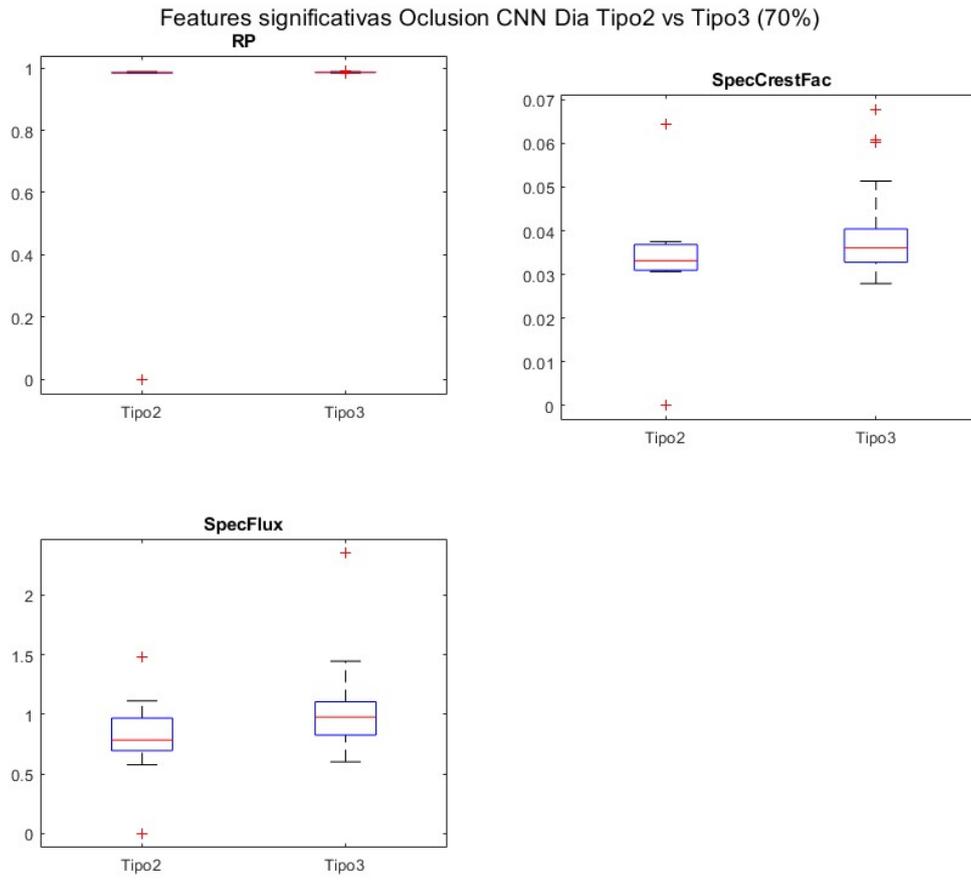


FIGURA A.24: *Boxplots* relevantes CNN Capa 3 Día Oclusión *Tipo2 vs Tipo3*

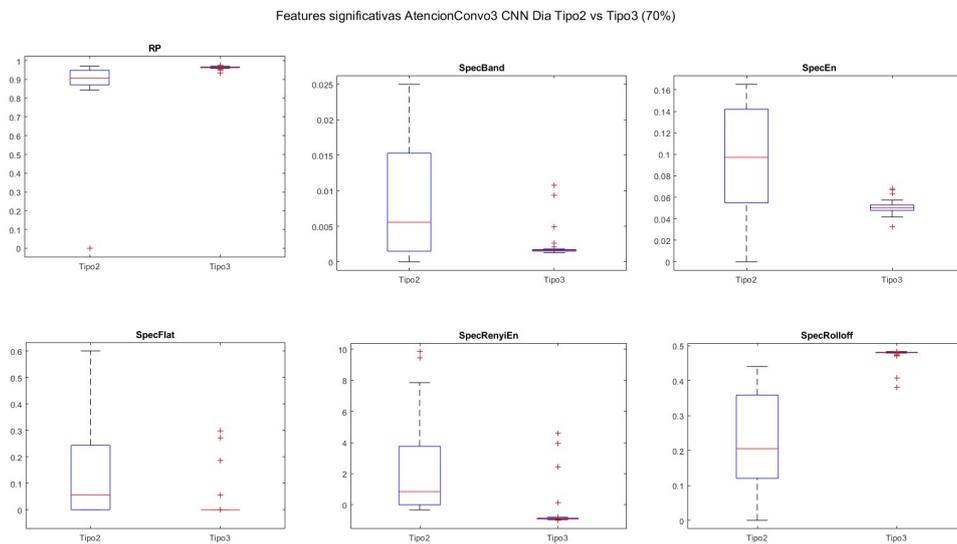


FIGURA A.25: *Boxplots* relevantes CNN Capa 3 Día Atención *Tipo2 vs Tipo3*

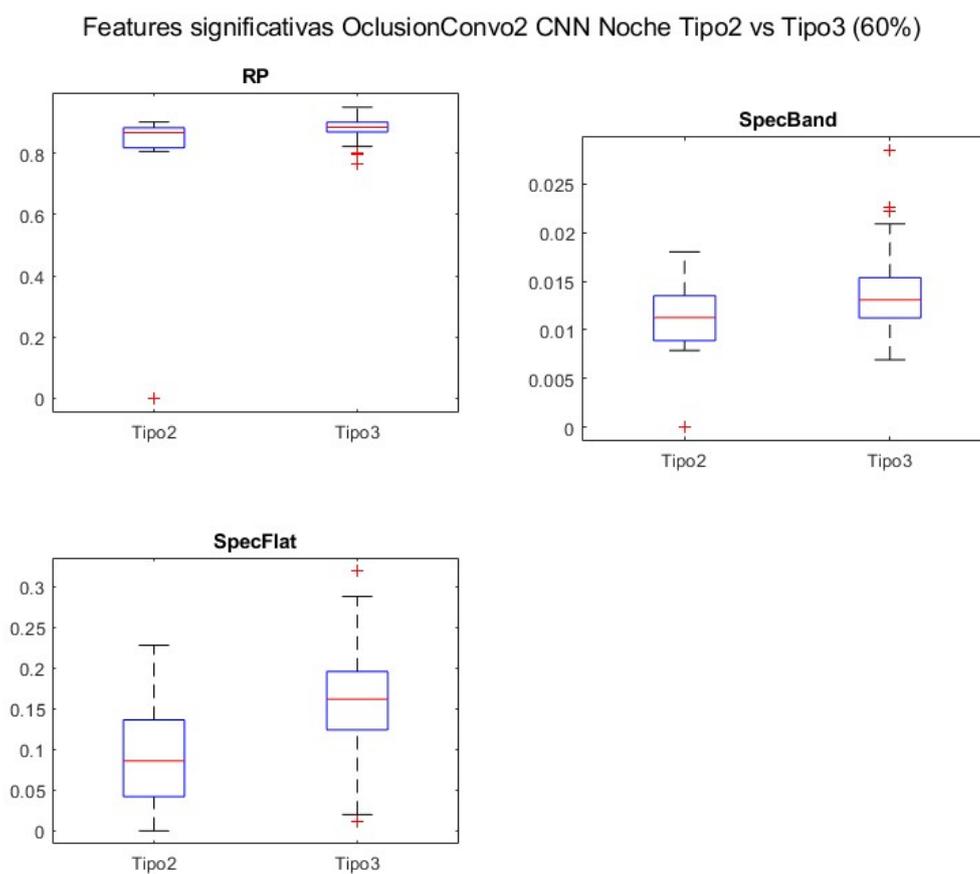


FIGURA A.26: *Boxplots* relevantes CNN Capa 2 Noche Oclusión *Tipo2 vs Tipo3*

Features significativas GradCAMConvo2 CNN Noche Tipo2 vs Tipo3 (80%)

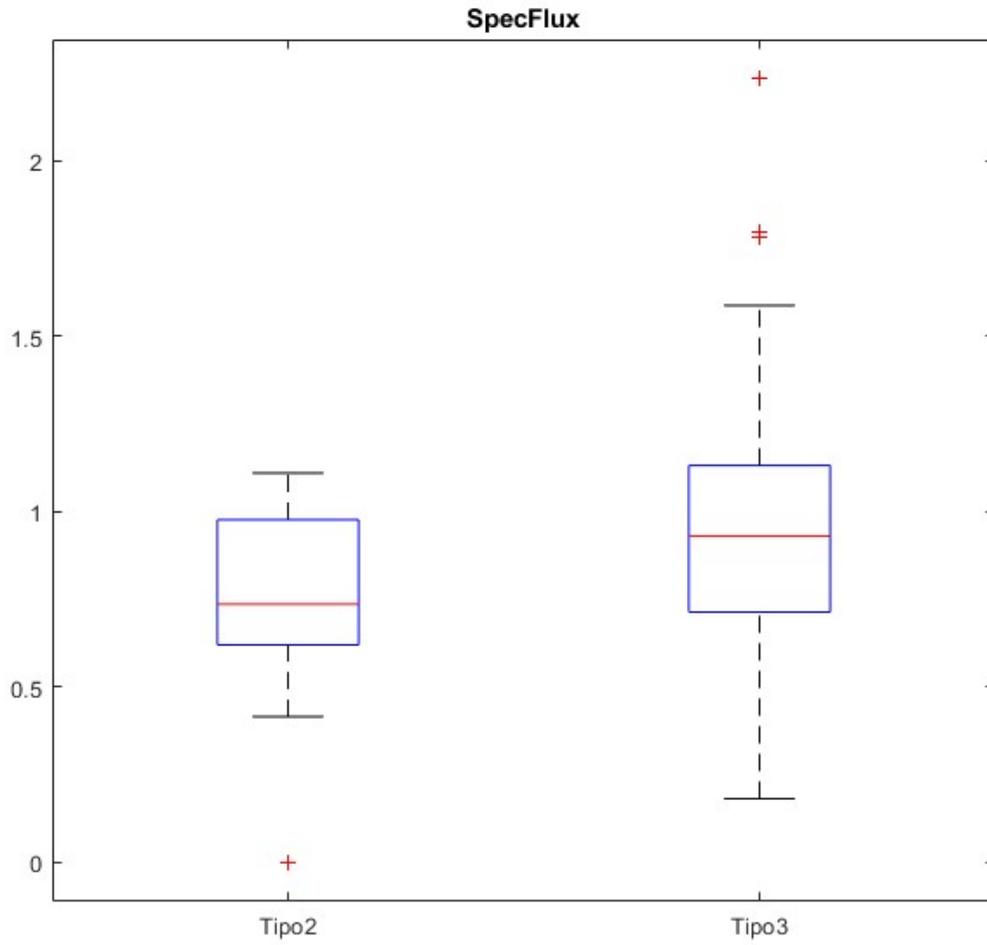


FIGURA A.27: *Boxplots* relevantes CNN Capa 2 Noche GradCAM *Tipo2 vs Tipo3*

Features significativas AtencionConvo2 CNN Noche Tipo2 vs Tipo3 (60%)

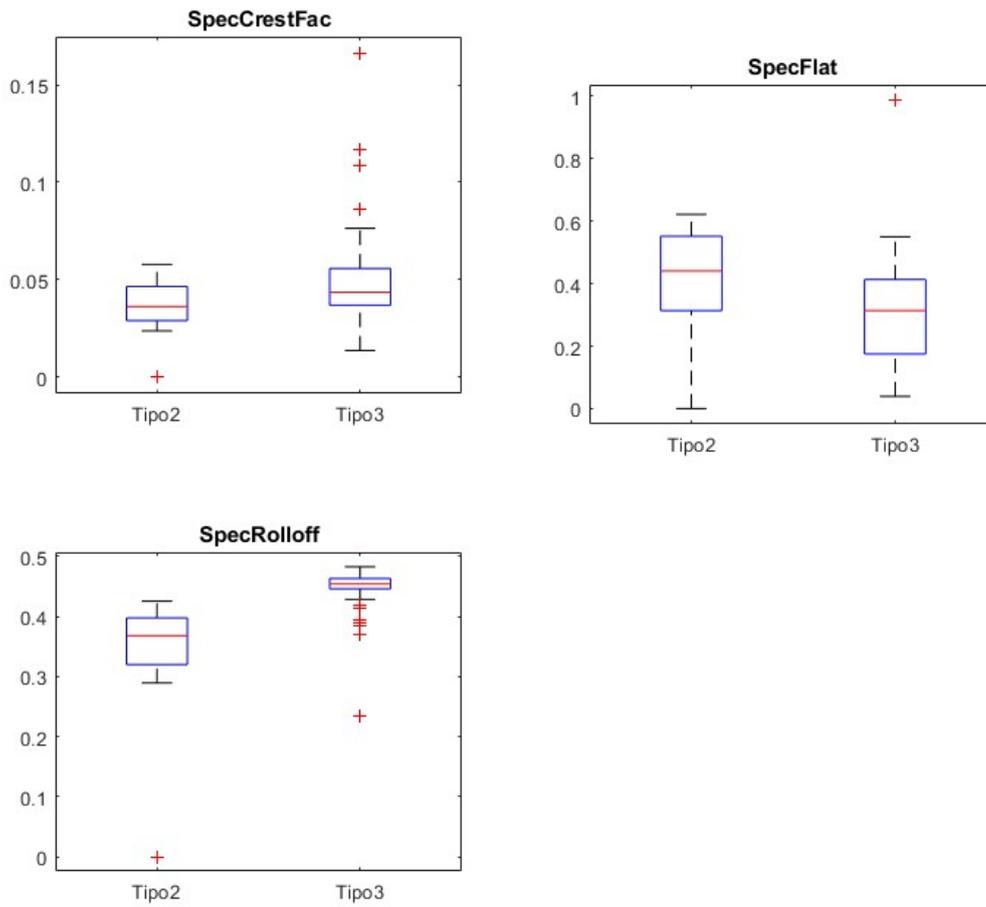


FIGURA A.28: *Boxplots* relevantes CNN Capa 2 Noche Atención *Tipo2 vs Tipo3*

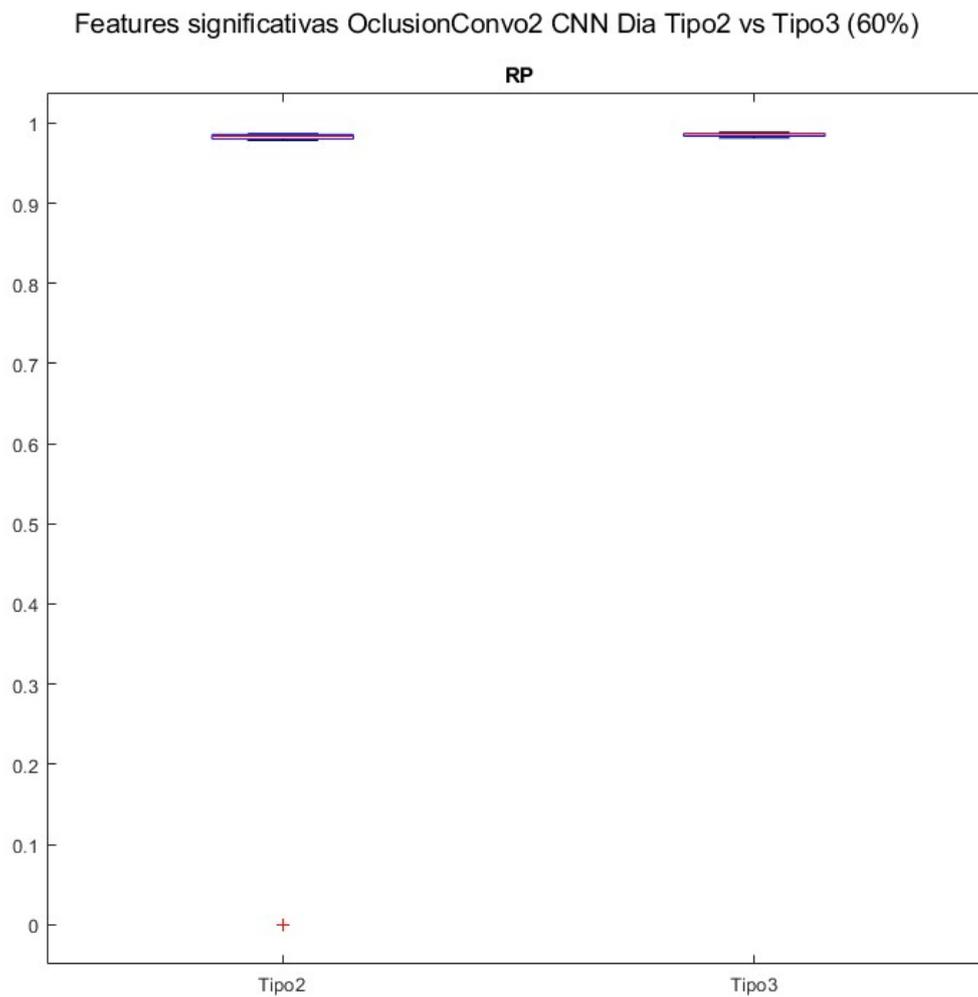


FIGURA A.29: *Boxplots* relevantes CNN Capa 2 Día Oclusión *Tipo2 vs Tipo3*

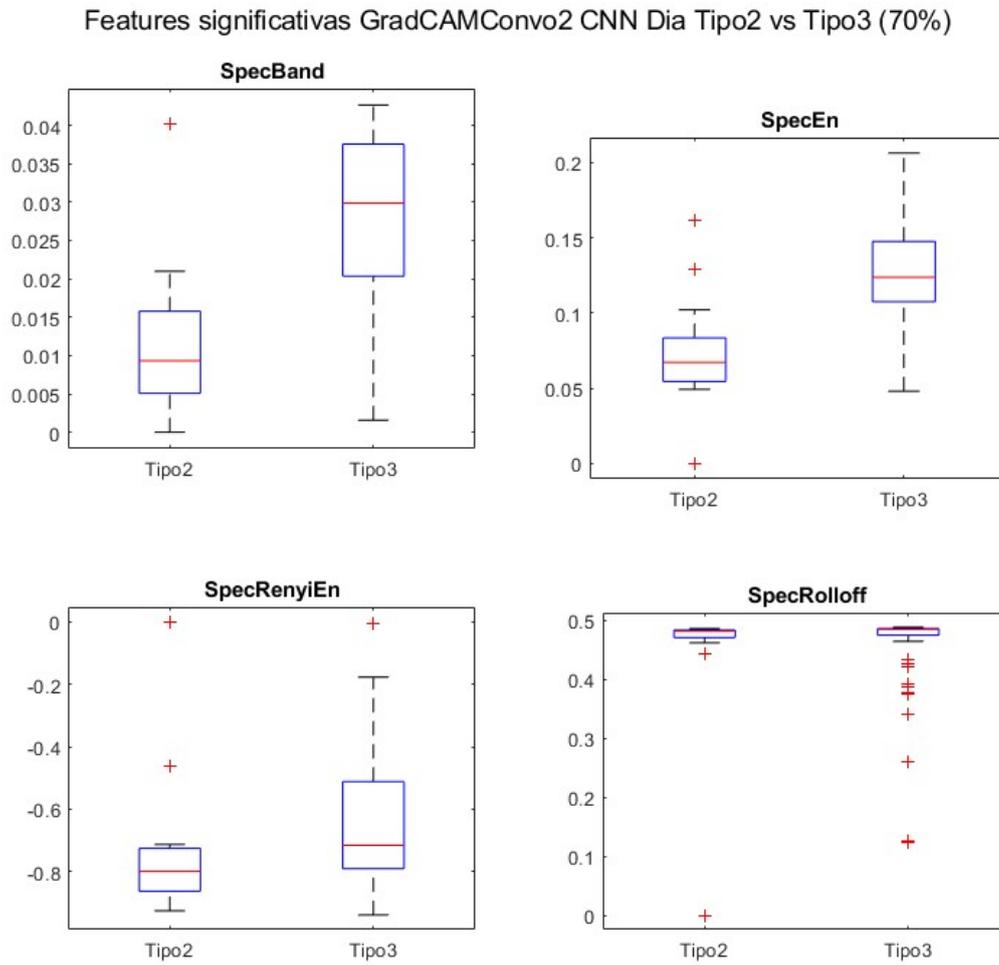


FIGURA A.30: *Boxplots* relevantes CNN Capa 2 Día GradCAM Tipo2 vs Tipo3

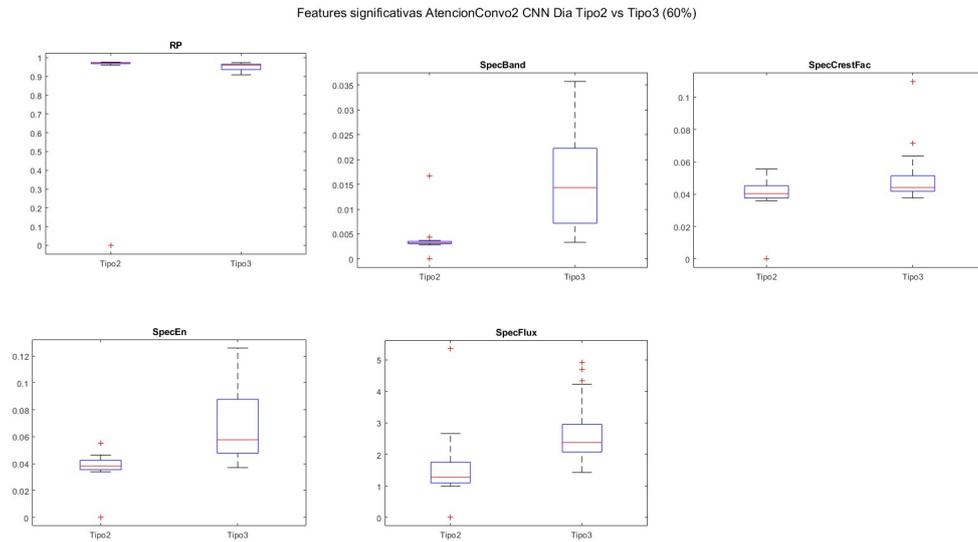


FIGURA A.31: *Boxplots* relevantes CNN Capa 2 Día Atención *Tipo2 vs Tipo3*

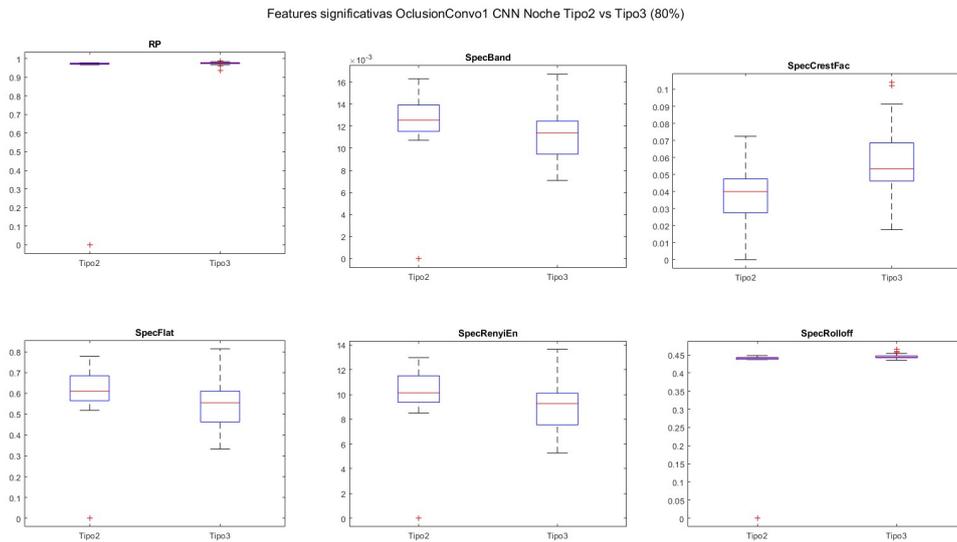


FIGURA A.32: *Boxplots* relevantes CNN Capa 1 Noche Oclusión *Tipo2 vs Tipo3*

Features significativas GradCAMConvo1 CNN Noche Tipo2 vs Tipo3 (80%)

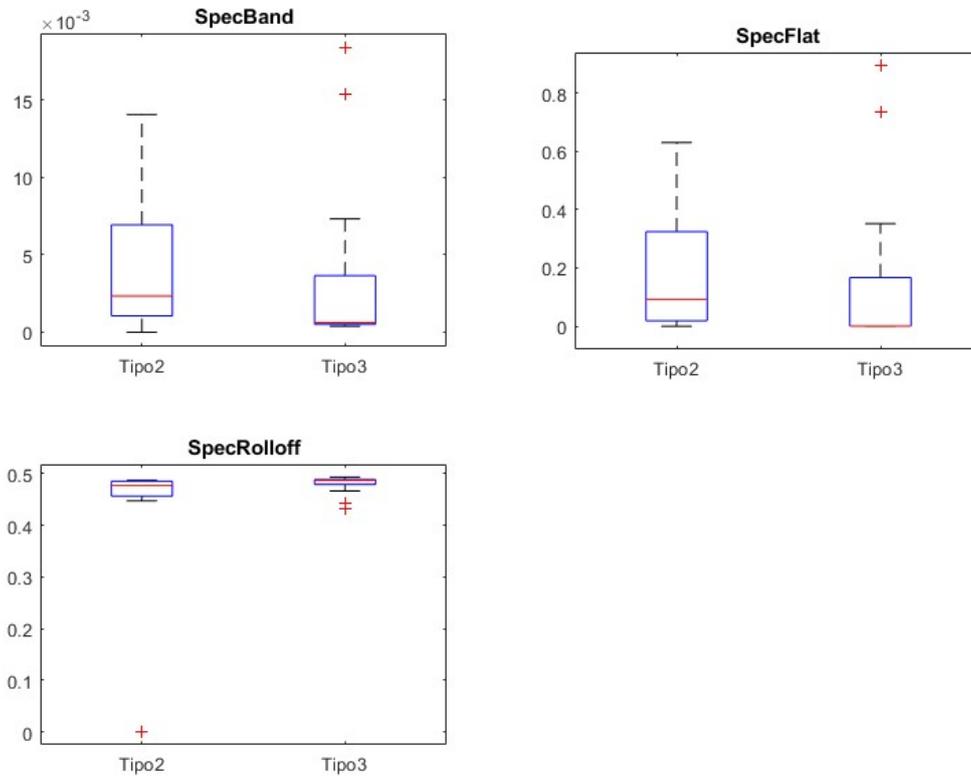


FIGURA A.33: *Boxplots* relevantes CNN Capa 1 Noche GradCAM Tipo2 vs Tipo3

Features significativas AtencionConvo1 CNN Noche Tipo2 vs Tipo3 (80%)

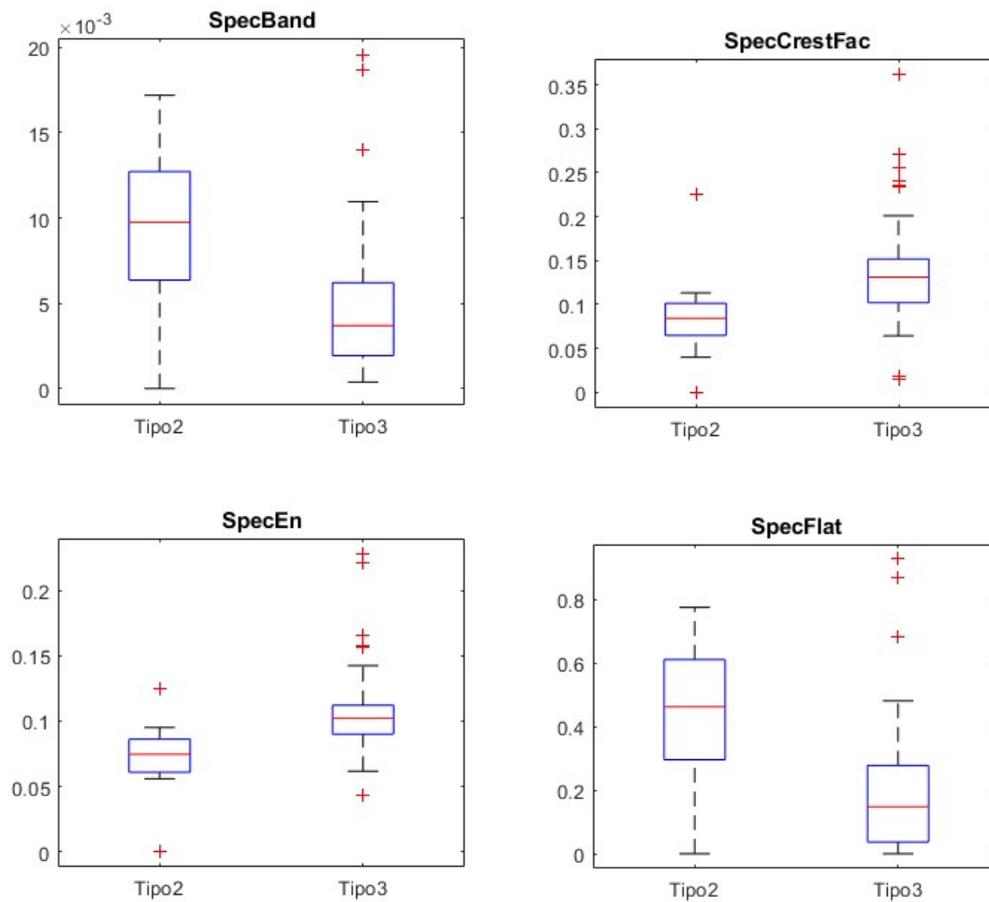


FIGURA A.34: *Boxplots* relevantes CNN Capa 1 Noche Atención *Tipo2 vs Tipo3*

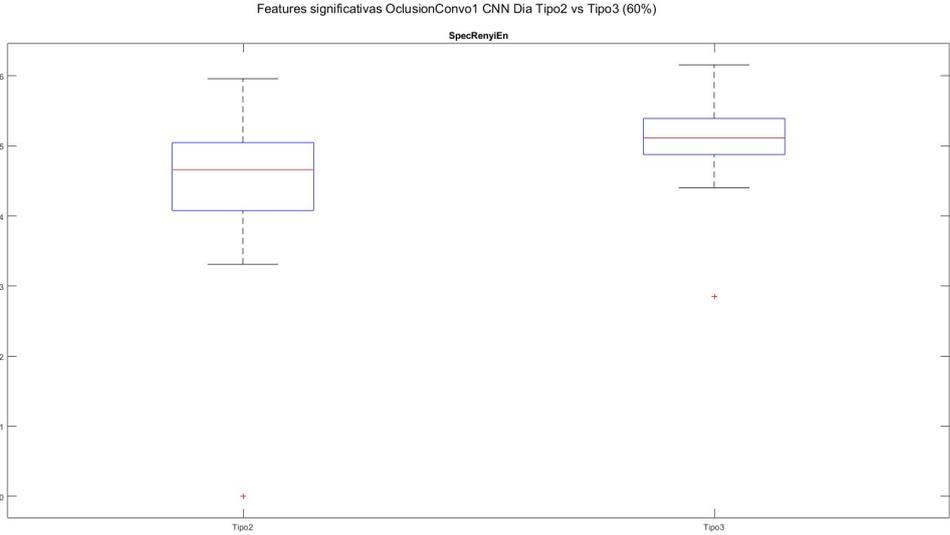


FIGURA A.35: *Boxplots* relevantes CNN Capa 1 Día Oclusión *Tipo2 vs Tipo3*

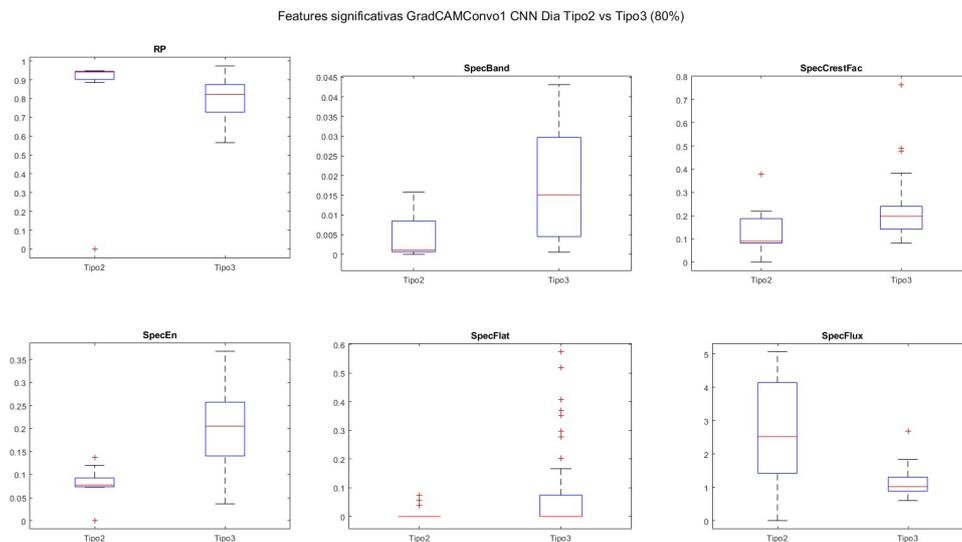


FIGURA A.36: *Boxplots* relevantes CNN Capa 1 Día GradCAM Tipo2 vs Tipo3

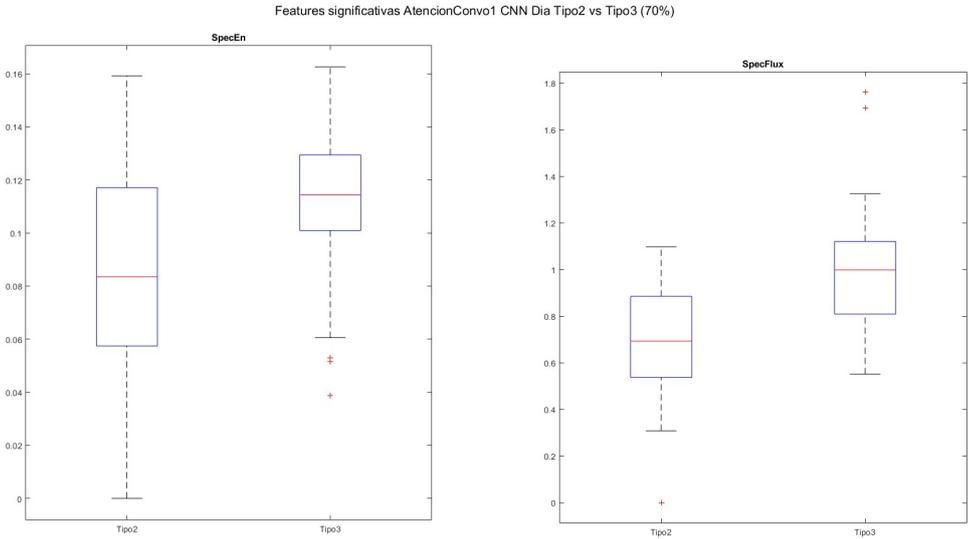


FIGURA A.37: *Boxplots* relevantes CNN Capa 1 Día Atención Tipo2 vs Tipo3

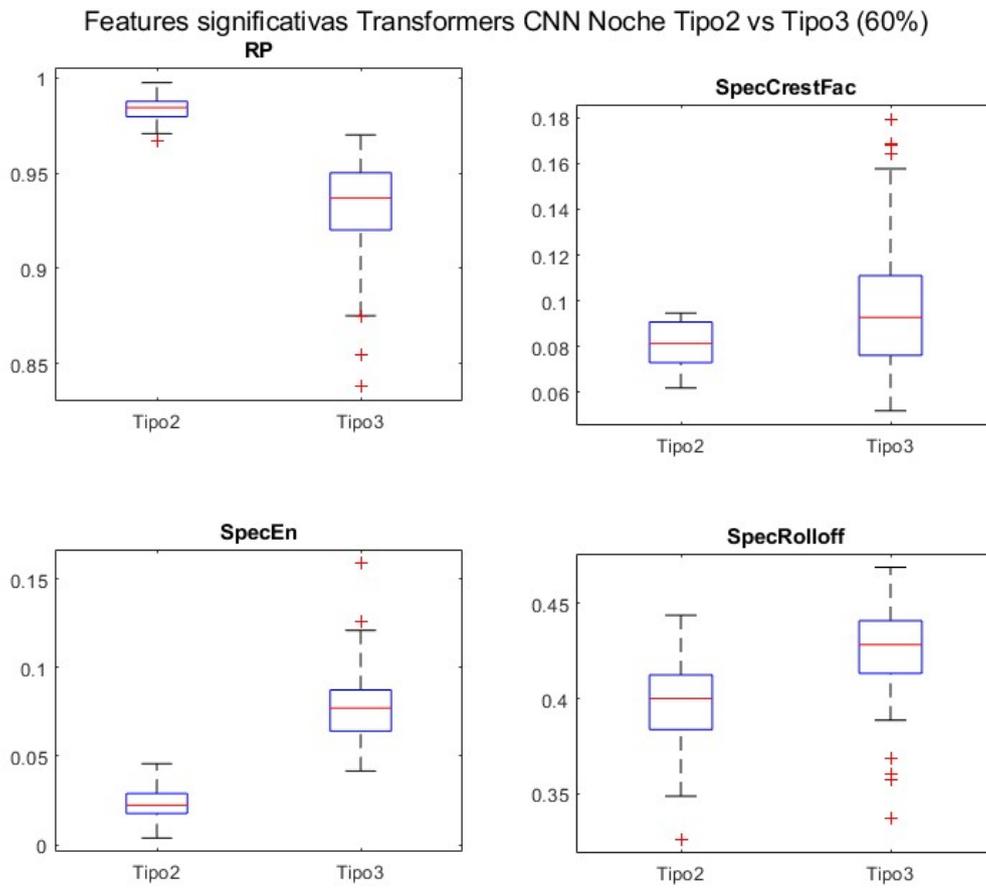


FIGURA A.38: *Boxplots* relevantes Transformers Noche Oclusión *Tipo2 vs Tipo3*

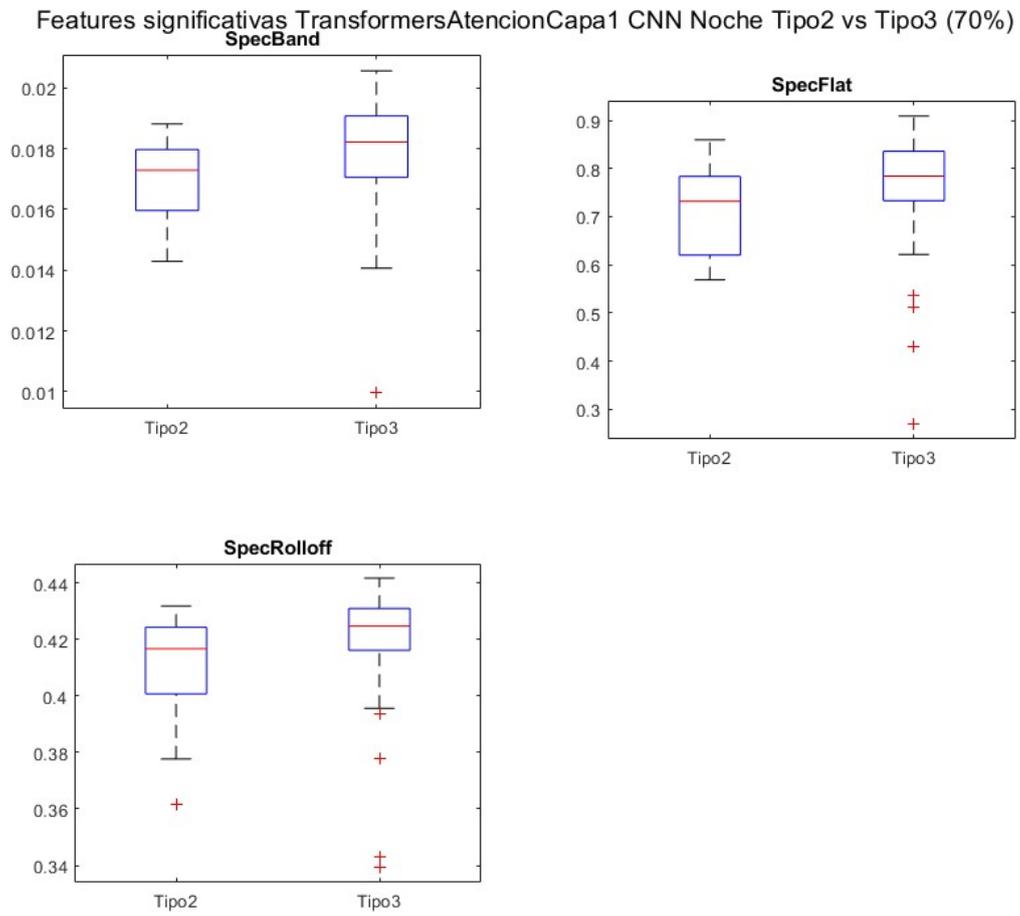


FIGURA A.39: *Boxplots* relevantes Transformers Noche Atención Capa 1 *Tipo2 vs Tipo3*

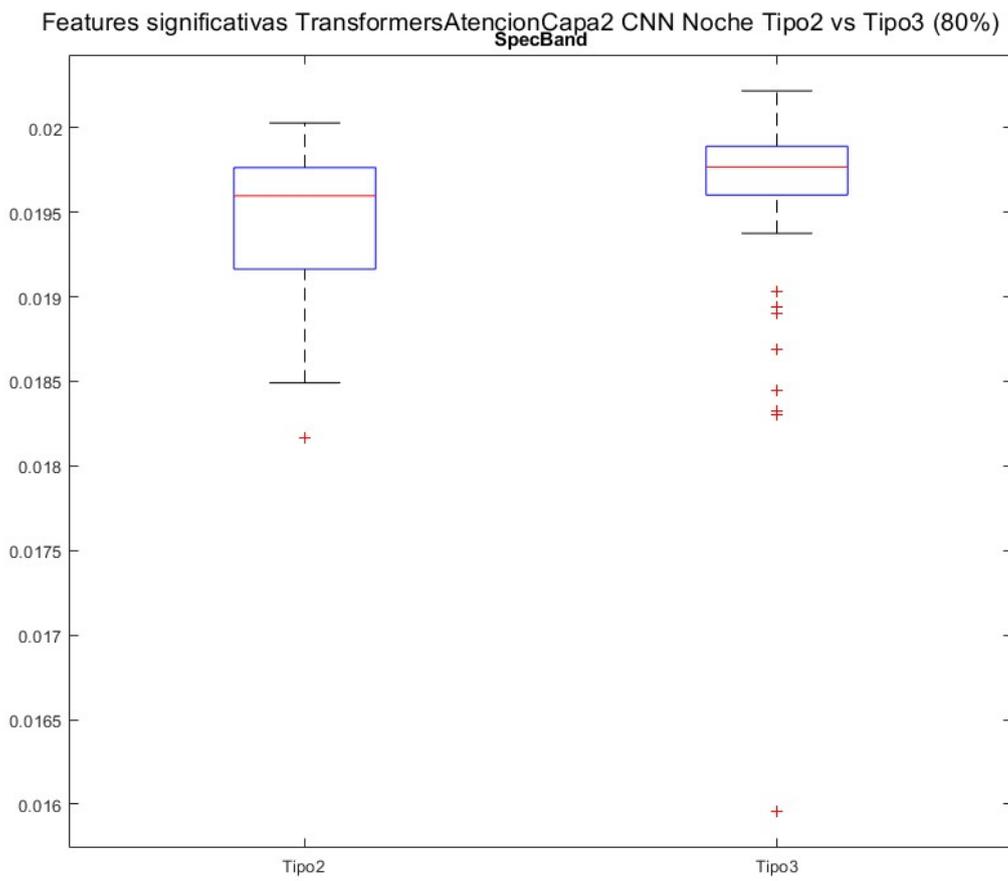


FIGURA A.40: *Boxplots* relevantes Transformers Noche Atención Capa 2 Tipo2 vs Tipo3

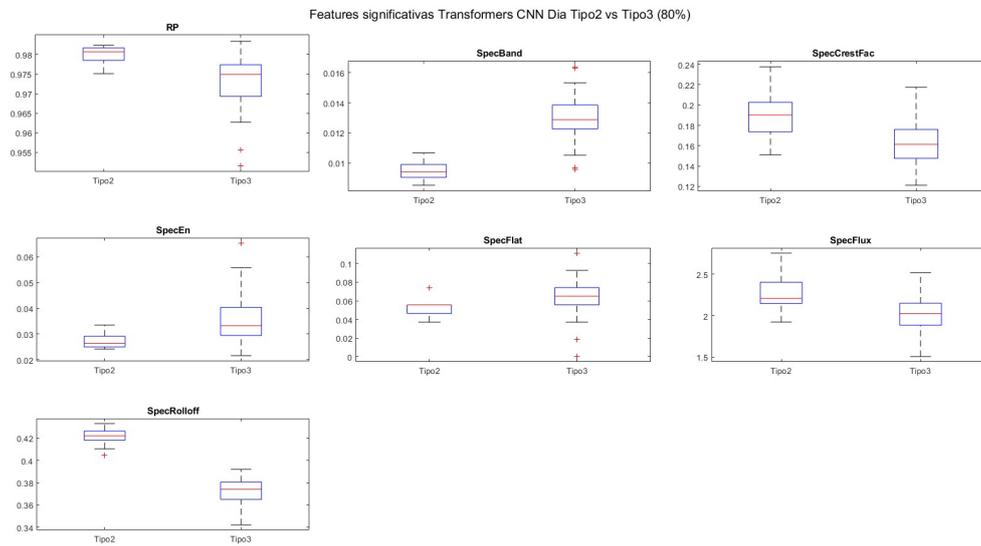


FIGURA A.41: *Boxplots* relevantes Transformers Día Oclusión *Tipo2 vs Tipo3*

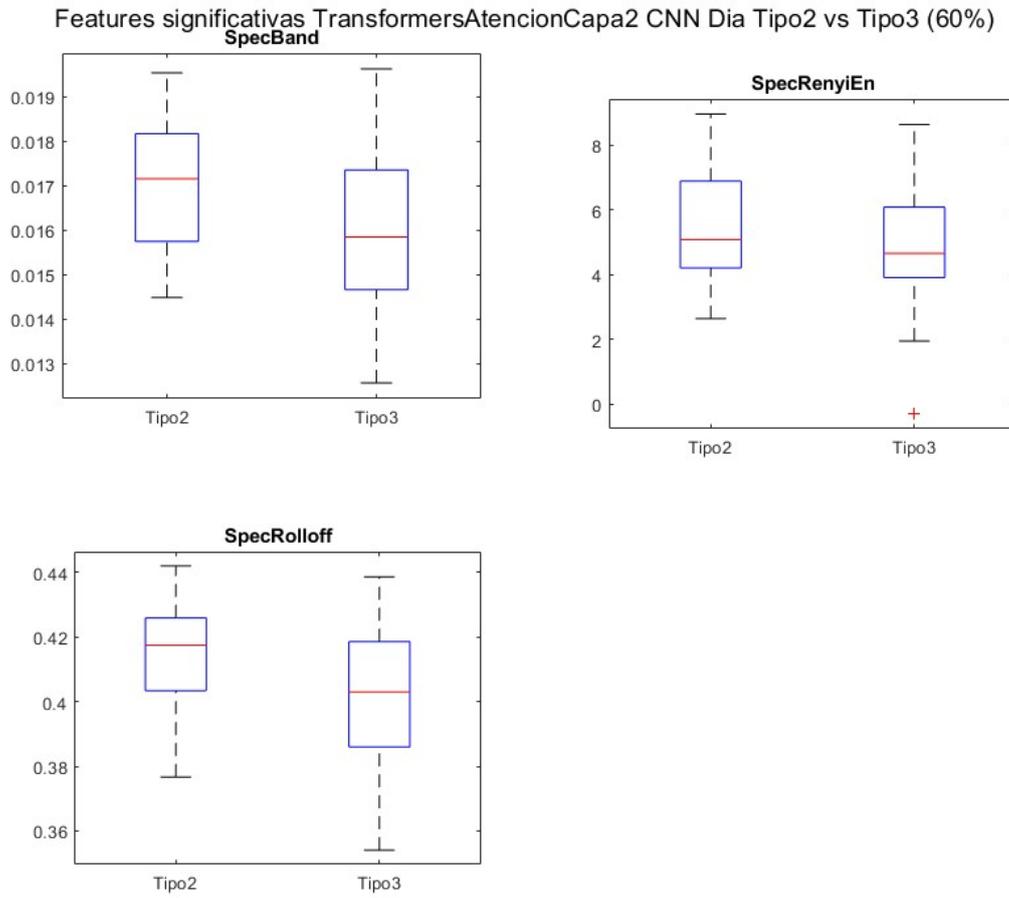


FIGURA A.42: *Boxplots* relevantes Transformers Día Atención Capa 2 *Tipo2 vs Tipo3*

## BIBLIOGRAFÍA

- [1] American Psychiatric Association. Diagnostic And Statistical Manual Of Mental Disorders, (DSM-V). *American Journal of Psychiatry*, 5ª edición, 2013.
- [2] Patricia Amado-Caballero, Pablo Casaseca-de-la Higuera, Susana Alberola-López, Jesus María Andrés-de Llano, Jose Antonio López-Villalobos, Jose Ramón Garmendia-Leiza, and Carlos Alberola-López. Objective adhd diagnosis using convolutional neural networks over daily-life activity records. *IEEE journal of biomedical and health informatics*, 24(9):2690–2700, 2020.
- [3] Patricia Amado-Caballero et al. Ayuda al diagnóstico de TDAH en la infancia mediante técnicas de procesamiento de señal y aprendizaje. 2019.
- [4] Paula Álvarez Tavera. Análisis de mapas de oclusión en una red convolucional para la obtención de patrones interpretables en pacientes de TDAH. 2022.
- [5] Guillermo Chico Delgado. Interpretación de redes neuronales profundas como herramientas para diferenciar subtipos de TDAH a partir de patrones de actividad. 2023.
- [6] Patricia Amado-Caballero, Pablo Casaseca de-la Higuera, Susana Alberola-López, Jesús María Andrés de Llano, José Antonio López-Villalobos, and Carlos Alberola-López. Insight into adhd diagnosis with deep learning on actimetry: Quantitative interpretation of occlusion maps in age and gender subgroups. *Artificial Intelligence in Medicine*, 143: 102630, 2023. ISSN 0933-3657. doi: <https://doi.org/10.1016/j.artmed.2023.102630>. URL <https://www.sciencedirect.com/science/article/pii/S0933365723001446>.
- [7] Patricia Amado-Caballero. Contribuciones a la detección automática de TDAH en la infancia mediante soluciones basadas en Deep Learning. 2021.
- [8] Erik G Willcutt. The prevalence of dsm-iv attention-deficit/hyperactivity disorder: a meta-analytic review. *Neurotherapeutics*, 9(3):490–499, 2012.
- [9] MathWorks MATLAB for Deep Learning. <https://es.mathworks.com/>, Último acceso Enero 2025.
- [10] Stefan Kottwitz. *LaTeX - Beginner's Guide*. Packt Publishing, 2011.
- [11] TensorFlow. Documentación oficial de tensorflow, Último acceso Enero 2025. URL [https://www.tensorflow.org/api\\_docs/python/tf/all\\_symbols](https://www.tensorflow.org/api_docs/python/tf/all_symbols).

- [12] Keras. Documentación oficial de keras, Último acceso Enero 2025. URL <https://keras.io/>.
- [13] NumPy. Documentación oficial de numpy, Último acceso Enero 2025. URL <https://numpy.org/doc/stable/>.
- [14] Matplotlib. Documentación oficial de matplotlib, Último acceso Enero 2025. URL <https://matplotlib.org/stable/index.html>.
- [15] scikit-learn. Documentación oficial de scikit-learn, Último acceso Enero 2025. URL <https://scikit-learn.org/0.21/documentation.html>.
- [16] American Psychiatric Association. Diagnostic And Statistical Manual Of Mental Disorders, (DSM-III). *American Journal of Psychiatry*, 3ª edición, 1980.
- [17] Erica D Palmer and Stanley Finger. An early description of adhd (inattentive subtype): Dr alexander crichton and 'mental restlessness'(1798). *Child Psychology and Psychiatry Review*, 6 (2):66–73, 2001.
- [18] James McGough. *Adhd*. Oxford University Press, 2014.
- [19] Guilherme V Polanczyk, Giovanni A Salum, Luisa S Sugaya, Arthur Caye, and Luis A Rohde. Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of child psychology and psychiatry*, 56(3):345–365, 2015.
- [20] L Rodríguez Molinero, JA López Villalobos, M Garrido Redondo, AM Sacristán Martín, MT Martínez Rivera, and F Ruiz Sanz. Estudio psicométrico-clínico de prevalencia y comorbilidad del trastorno por déficit de atención con hiperactividad en Castilla y León (España). *Pediatría Atención Primaria*, 11(42):251–270, 06 2009. ISSN 1139-7632. URL [http://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1139-76322009000200006&nrm=iso](http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1139-76322009000200006&nrm=iso).
- [21] Yannis Paloyelis, Mitul A Mehta, Jonna Kuntsi, and Philip Asherson. Functional mri in adhd: a systematic literature review. *Expert review of neurotherapeutics*, 7(10):1337–1356, 2007.
- [22] Agatha Lenartowicz and Sandra K Loo. Use of eeg to diagnose adhd. *Current psychiatry reports*, 16:1–11, 2014.
- [23] Adeleh Dehghani Nazhvani, Reza Boostani, Somayeh Afrasiabi, and Khadijeh Sadatnezhad. Classification of adhd and bmd patients using visual evoked potential. *Clinical neurology and neurosurgery*, 115(11):2329–2335, 2013.
- [24] Astar Lev, Yoram Braw, Tomer Elbaum, Michael Wagner, and Yuri Rassovsky. Eye tracking during a continuous performance test: Utility for assessing adhd patients. *Journal of Attention Disorders*, 26(2):245–255, 2022.
- [25] Elena Martín-Gonzalez et al. Intervalos de normalidad en medidas de actimetría: un estudio para diferentes tipos de actímetros. 2017.

- [26] Pablo Casaseca-de-la Higuera, Diego Martín-Martínez, Susana Alberola-López, Jesús María Andrés-de Llano, José Antonio López-Villalobos, José Ramón-Garmendia Leiza, and Carlos 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. IEEE, 2012.
- [27] Zhenyu Mao, Yi Su, Guangquan Xu, Xueping Wang, Yu Huang, Weihua Yue, Li Sun, and Naixue Xiong. Spatio-temporal deep learning method for adhd fmri classification. *Information Sciences*, 499:1–11, 2019.
- [28] Deping Kuang and Lianghua He. Classification on adhd with deep learning. In *2014 International Conference on Cloud Computing and Big Data*, pages 27–32. IEEE, 2014.
- [29] Amirmasoud Ahmadi, Mehrdad Kashefi, Hassan Shahrokhi, and Mohammad Ali Nazari. Computer aided diagnosis system using deep convolutional neural networks for adhd subtypes. *Biomedical Signal Processing and Control*, 63:102227, 2021.
- [30] Amirali Vahid, Annet Bluschke, Veit Roessner, Sebastian Stober, and Christian Beste. Deep learning based on event-related eeg differentiates children with adhd from healthy controls. *Journal of clinical medicine*, 8(7):1055, 2019.
- [31] Aleksandar Tenev, Silvana Markovska-Simoska, Ljupco Kocarev, Jordan Pop-Jordanov, Andreas Müller, and Gian Candrian. Machine learning approach for classification of adhd adults. *International Journal of Psychophysiology*, 93(1):162–166, 2014.
- [32] Masatoshi Yamashita, Qiulu Shou, and Yoshifumi Mizuno. Unsupervised machine learning for identifying attention-deficit/hyperactivity disorder subtypes based on cognitive function and their implications for brain structure. *Psychological Medicine*, 54(14):3917–3929, 2024.
- [33] 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.
- [34] 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.
- [35] Annika Wiebe, Benjamin Selaskowski, Martha Paskin, Laura Asché, Julian Pakos, Behrem Aslan, Silke Lux, Alexandra Philipsen, and Niclas Braun. Virtual reality-assisted prediction of adult adhd based on eye tracking, eeg, actigraphy and behavioral indices: a machine learning analysis of independent training and test samples. *Translational psychiatry*, 14(1):508, 2024.
- [36] Georgios Feretzakis, Iris Boufeas, Sophia Fotakidis, Efrosyni Bazakidou, Konstantinos Kalodanis, Paraskevi Peristera, and Athanasios Anastasiou. Enhanced machine learning approach to adhd classification using actigraphy data. *Studies in health technology and informatics*, 328:136–140, 2025.

- [37] Zhihan Jiang, Adrienne YL Chan, Dawn Lum, Kirstie HTW Wong, Janice CN Leung, Patrick Ip, David Coghill, Rosa S Wong, Edith CH Ngai, and Ian CK Wong. Wearable signals for diagnosing attention-deficit/hyperactivity disorder in adolescents: A feasibility study. *JAACAP Open*, 2024.
- [38] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553): 436–444, 2015.
- [39] Stuart J Russell and Peter Norvig. *Artificial intelligence: a modern approach*. Pearson, 2016.
- [40] Rohan Gupta, Devesh Srivastava, Mehar Sahu, Swati Tiwari, Rashmi K Ambasta, and Pravir Kumar. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Molecular diversity*, 25:1315–1360, 2021.
- [41] Keyrus. Conceptos clave de la ia que todo experto en marketing debería controlar, Último acceso Julio 2025. URL <https://keyrus.com/sp/es/insights/conceptos-clave-de-la-ia-que-todo-experto-en-marketing-deberia-controlar>.
- [42] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [43] Andrej Krenker, Janez Bešter, and Andrej Kos. Introduction to the artificial neural networks. *Artificial Neural Networks: Methodological Advances and Biomedical Applications. InTech*, pages 1–18, 2011.
- [44] AD Dongare, RR Kharde, Amit D Kachare, et al. Introduction to artificial neural network. *International Journal of Engineering and Innovative Technology (IJEIT)*, 2(1):189–194, 2012.
- [45] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
- [46] Willy Alex Castañeda Sánchez, Benjamín Roldan Polo Escobar, and Fernando Vega Huincho. Redes neuronales artificiales: una medición de aprendizajes de pronósticos como demanda potencial. *Universidad, Ciencia y Tecnología*, 27(118):51–60, 2023.
- [47] Sagar Sharma, Simone Sharma, and Anidhya Athaiya. Activation functions in neural networks. *Towards Data Sci*, 6(12):310–316, 2017.
- [48] Francisco Alonso. Redes neuronales y deep learning - capítulo 2: La neurona, Último acceso Julio 2025. URL <https://www.futurespace.es/redes-neuronales-y-deep-learning-capitulo-2-la-neurona/>.
- [49] Antonio Richaud. Redes perceptrón multicapa, Último acceso Julio 2025. URL <https://antonio-richaud.com/blog/archivo/publicaciones/41-redes-perceptron-multicapa.html>.
- [50] Oscar R Dolling and Eduardo A Varas. Artificial neural networks for streamflow prediction. *Journal of hydraulic research*, 40(5):547–554, 2002.

- [51] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. Learning to learn by gradient descent by gradient descent. *Advances in neural information processing systems*, 29, 2016.
- [52] James Gareth, Witten Daniela, Hastie Trevor, and Tibshirani Robert. *An introduction to statistical learning: with applications in R*. Springer, 2013.
- [53] Xue Ying. An overview of overfitting and its solutions. In *Journal of physics: Conference series*, volume 1168, page 022022. IOP Publishing, 2019.
- [54] Daniel Berrar et al. Cross-validation., 2019.
- [55] Interactive Chaos. Validación cruzada de k iteraciones, Último acceso Julio 2025. URL <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/validacion-cruzada-de-k-iteraciones>.
- [56] Interactive Chaos. Validación cruzada aleatoria, Último acceso Julio 2025. URL <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/validacion-cruzada-aleatoria>.
- [57] Interactive Chaos. Validación cruzada dejando uno fuera, Último acceso Julio 2025. URL <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/validacion-cruzada-dejando-uno-fuera>.
- [58] Ricardo Mendoza. Aplicación del gradiente en redes neuronales, Último acceso Julio 2025. URL <https://medium.com/@ricardojmv85/aplicaci%C3%B3n-del-gradiente-en-redes-neuronales-78bff0d802d5>.
- [59] Daniel Svozil, Vladimir Kvasnicka, and Jiri Pospichal. Introduction to multi-layer feed-forward neural networks. *Chemometrics and intelligent laboratory systems*, 39(1):43–62, 1997.
- [60] Hamid Jahani and Ali Asghar Safaei. Neural signals processing using deep learning for diagnosis of cognitive disorders. In *Signal Processing Strategies*, pages 91–118. Elsevier, 2025.
- [61] Larry R Medsker, Lakhmi Jain, et al. Recurrent neural networks. *Design and Applications*, 5 (64-67):2, 2001.
- [62] Pramodya Sahan. Cnn from scratch: Building the foundation with python - part 1, Último acceso Julio 2025. URL <https://medium.com/@pramodyasahan.edu/cnn-from-scratch-building-the-foundation-with-python-part-1-826c890c1bcb>.
- [63] K O’Shea. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.
- [64] Anh H. Reynolds. Convolutional neural networks (cnns), Último acceso Julio 2025. URL <https://anhreynolds.com/blogs/cnn.html>.
- [65] Muhamad Yani, Si MT Budhi Irawan, S, and MT Casi Setiningsih, ST. Application of transfer learning using convolutional neural network method for early detection of terry’s nail. In *Journal of Physics: Conference Series*, volume 1201, page 012052. IOP Publishing, 2019.

- [66] Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. Transformers in vision: A survey. *ACM computing surveys (CSUR)*, 54(10s):1–41, 2022.
- [67] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.
- [68] Wikipedia, the free encyclopedia. transformer (deep learning architecture), Último acceso Diciembre 2024. URL [https://en.wikipedia.org/wiki/Transformer\\_\(deep\\_learning\\_architecture\)#Architecture](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)#Architecture).
- [69] Bichen Wu, Chenfeng Xu, Xiaoliang Dai, Alvin Wan, Peizhao Zhang, Zhicheng Yan, Masayoshi Tomizuka, Joseph Gonzalez, Kurt Keutzer, and Peter Vajda. Visual transformers: Token-based image representation and processing for computer vision. *arXiv preprint arXiv:2006.03677*, 2020.
- [70] Maithra Raghu, Thomas Unterthiner, Simon Kornblith, Chiyuan Zhang, and Alexey Dosovitskiy. Do vision transformers see like convolutional neural networks? *Advances in neural information processing systems*, 34:12116–12128, 2021.
- [71] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. Xai—explainable artificial intelligence. *Science robotics*, 4(37):eaay7120, 2019.
- [72] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part I 13*, pages 818–833. Springer, 2014.
- [73] MathWorks MATLAB for Deep Learning. Understand network predictions using occlusion, Último acceso Diciembre 2024. URL <https://es.mathworks.com/help/deeplearning/ug/understand-network-predictions-using-occlusion.html>.
- [74] 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.
- [75] Ramprasaath R Selvaraju, Abhishek Das, Ramakrishna Vedantam, Michael Cogswell, Devi Parikh, and Dhruv Batra. Grad-cam: Why did you say that? *arXiv preprint arXiv:1611.07450*, 2016.
- [76] Qiaoying Teng, Zhe Liu, Yuqing Song, Kai Han, and Yang Lu. A survey on the interpretability of deep learning in medical diagnosis. *Multimedia Systems*, 28(6):2335–2355, 2022.
- [77] Zhaoyang Niu, Guoqiang Zhong, and Hui Yu. A review on the attention mechanism of deep learning. *Neurocomputing*, 452:48–62, 2021.
- [78] Mariano Ntroukas, Nikolaos Gkalelis, and Vasileios Mezaris. Tame: Attention mechanism based feature fusion for generating explanation maps of convolutional neural networks. In *2022 IEEE International Symposium on Multimedia (ISM)*, pages 58–65. IEEE, 2022.

- [79] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [80] Jesús Monge-Álvarez, Carlos Hoyos-Barceló, Luis Miguel San-José-Revuelta, and Pablo Casaseca-de-la Higuera. A machine hearing system for robust cough detection based on a high-level representation of band-specific audio features. *IEEE Transactions on Biomedical Engineering*, 66(8):2319–2330, 2018.
- [81] Jesús Monge-Álvarez, Carlos Hoyos-Barceló, Paul Lesso, and Pablo Casaseca-De-La-Higuera. Robust detection of audio-cough events using local hu moments. *IEEE journal of biomedical and health informatics*, 23(1):184–196, 2018.
- [82] Arunan Ramalingam and Sridhar Krishnan. Gaussian mixture modeling using short time fourier transform features for audio fingerprinting. In *2005 IEEE International Conference on Multimedia and Expo*, pages 1146–1149. IEEE, 2005.
- [83] Sujatha Mudadla. What is the main difference between the t-test and the wilcoxon test in terms of their assumptions and calculation methods?, Último acceso Diciembre 2024. URL <https://medium.com/>. Accedido en diciembre 2024. URL <https://medium.com/@sujathamudadla1213/what-is-the-main-difference-between-the-t-test-and-the-wilcoxon-test-in-terms-of-their-assumptions-ebff4c8385aa>.
- [84] MathWorks MATLAB for Deep Learning. Matlab (ttest2), Último acceso Diciembre 2024. URL <https://es.mathworks.com/help/stats/ttest2.html#btrkaaw>.
- [85] MathWorks MATLAB for Deep Learning. Matlab (ranksum), Último acceso Diciembre 2024. URL <https://es.mathworks.com/help/stats/ranksum.html?searchHighlight=ranksu>.
- [86] MathWorks MATLAB for Deep Learning. chi2gof. prueba de bondad de ajuste de chi-cuadrado, Último acceso Septiembre 2025. URL <https://es.mathworks.com/help/stats/chi2gof.html>.