



Universidad Internacional de la Rioja
(UNIR)

ESIT

Máster Universitario en Inteligencia Artificial

Deep Learning para la detección de Encefalitis Herpética (HSVE)

Trabajo Fin de Máster

Presentado por: Carlos Breuer Carrasco

Dirigido por: Guillermo Torralba Elipe

Tipo de trabajo: Comparativa de soluciones

Ciudad: Cáceres

Fecha: 16 de julio de 2023

Índice de Contenidos

1. Introducción	1
1.1. Motivación	2
1.2. Planteamiento del problema	3
1.3. Estructura del trabajo	3
2. Contexto y Estado del Arte	5
2.1. Encefalitis herpética simple (HSVE)	5
2.2. Resonancia Magnética Nuclear	7
2.3. El Machine Learning	9
2.4. El Deep Learning	11
2.4.1. Redes Neuronales Artificiales (ANN)	12
2.4.2. Redes Neuronales Convolucionales (CNN)	15
2.4.3. VGG	16
2.4.4. Redes Neuronales Residuales (ResNets)	17
2.4.5. InceptionV3	18
2.4.6. MobileNet	18
2.4.7. Transfer learning	19
2.4.8. Técnicas de regularización	20
2.5. Métricas de evaluación del modelo de clasificación	23
2.6. Investigaciones recientes relativas al tema del proyecto	27
3. Alcance y objetivos	29
3.1. Objetivo general	29
3.2. Objetivos específicos	29

3.3. Limitaciones	30
4. Material y Método	31
4.1. Fuente de datos	31
4.1.1. Conjunto de datos 1: MRI de cerebros con HSVE	32
4.1.2. Conjunto de datos 2: MRI de cerebros sanos	37
4.2. Herramientas Hardware utilizadas	40
4.3. Herramientas Software utilizadas	40
4.4. Metodología	41
4.4.1. Preprocesamiento de las imágenes	45
5. Iteración 1 de experimentos	46
5.1. Modelos implementados	46
5.1.1. Modelo 1 - Arquitectura de CNN construida desde cero	48
5.1.2. Modelo 2 - Transfer Learning con Resnet50	49
5.1.3. Modelo 3 - Transfer Learning con VGG16	49
5.1.4. Modelo 4 - Transfer Learning con VGG19	50
5.1.5. Modelo 5 - Transfer Learning con InceptionV3	50
5.1.6. Modelo 6 - Transfer Learning con MobileNet	50
5.2. Resultados obtenidos	51
5.2.1. Modelo 1 - Arquitectura de CNN desde cero	51
5.2.2. Modelo 2 - Transfer Learning con Resnet50	53
5.3. Resultados globales iteración 1	56
6. Iteración 2 de experimentos	58
6.1. Modelos implementados	58
6.1.1. Modelo 2 - Transfer Learning con ResNet50 y MRI conjunto inicial	59
6.1.2. Modelo 7 - Transfer Learning con ResNet50 y MRI subconjunto sólo plano AXIAL	59
6.1.3. Modelo 8 - Transfer Learning con ResNet50 y MRI subconjunto formatos FLAIR, T1 y T2 sólo plano AXIAL	60

6.2. Resultados obtenidos	62
6.2.1. Modelo 2 - Transfer Learning ResNet50 y conjunto inicial . . .	62
6.2.2. Modelo 7 - Modelo 2 + MRI subconjunto sólo plano AXIAL .	62
6.3. Resultados globales iteración 2	65
7. Iteración 3 de experimentos	66
7.1. Modelos implementados	66
7.1.1. Modelo 9 - Modelo 2 + <i>Dropout + Batch Normalization</i> . . .	67
7.1.2. Modelo 10 - Modelo 9 + aumento de datos	68
7.2. Resultados obtenidos	70
7.2.1. Modelo 10 - Modelo 9 + aumento de datos	71
7.3. Resultados globales iteración 3	75
8. Conclusiones y Trabajos Futuros	76
8.1. Conclusiones	76
8.2. Líneas de trabajos futuros	78
A. Relación de casos: cerebros sanos	91
B. Representación computacional de arquitecturas con keras	93
B.1. Modelo 1 - Arquitectura de CNN construida desde cero	93
B.2. Modelo 2 - Transfer Learning con Resnet50	95
B.3. Modelo 3 - Transfer Learning con VGG16	96
B.4. Modelo 4 - Transfer Learning con VGG19	97
B.5. Modelo 5 - Transfer Learning con InceptionV3	98
B.6. Modelo 6 - Transfer Learning con MobileNet	99
B.7. Modelo 9: <i>Dropout + Batch Normalization</i>	99

Índice de Figuras

2.1.	Localización de lóbulos cerebrales en zonas corticales. Nota. Adaptado del repositorio de imágenes gratuitas Pixabay, por ArtsyBee, 2015.	5
2.2.	Esquema del proceso de transmisión del virus del herpes simple (HSV-1) hasta el cerebro, produciendo la enfermedad HSVE (Marcocci et al., 2020) Nota. Adaptado de Herpes simplex virus-1 in the brain: The dark side of a sneaky infection, por Trends in Microbiology, 2020.	6
2.3.	Ejemplos de imágenes MRI del encéfalo mediante capturas en el plano coronal, sagital y horizontal (plano axial). Fuente: Material académico de UNIR (Muñoz Garrido, 2022)	8
2.4.	Tipos de sistemas de Machine Learning	10
2.5.	<i>Deep Learning</i> como subcampo del ML y de la IA	11
2.6.	Organización por capas de una red neuronal (Andrew Ng, 2021). Fuente: Programa de especialización de Deep Learning en Coursera	12
2.7.	Procesamiento sobre 1 neurona (Andrew Ng, 2021). Fuente: Programa de especialización de Deep Learning en Coursera	14
2.8.	Detección de bordes, formas y objetos mediante operaciones de convolución en diferentes regiones de una red neuronal (Andrew Ng, 2021). Fuente: Programa de especialización de Deep Learning en Coursera	16
2.9.	Arquitectura de VGG-16. Fuente: Wikimedia Commons	16
2.10.	Bloque de aprendizaje residual mediante atajos (He et al., 2015)	17
2.11.	Arquitectura de InceptionV3 (Szegedy et al., 2015). Fuente: Google Cloud	18

2.12. Arquitectura de MobileNet (Wang et al., 2020). Nota. Adaptado de A Novel Image Classification Approach via Dense-MobileNet Models, por Indawi, 2020.	19
2.13. Aplicación de un Dropout del 50 % en todas las capas de la red neuronal (Andrew Ng, 2021). Fuente: Programa de especialización de Deep Learning en Coursera	22
4.1. Imágenes MRI del cerebro con HSVE del caso 1 desde los planos AXIAL, CORONAL y SAGITAL. Fuente: <i>Radiopaedia</i> (Radiopaedia, 2023)	36
4.2. Imágenes MRI del cerebro sano del caso 4 desde los planos AXIAL, CORONAL y SAGITAL. Fuente: <i>Radiopaedia</i> (Gaillard, 2023)	39
4.3. Proceso cíclico en proyectos de Deep Learning	43
4.4. Fases del proceso seguido durante el desarrollo de los experimentos de Red Neuronal Convolucional	44
5.1. Evolución de la exactitud y la pérdida del primer modelo	51
5.2. Matriz de confusión del primer modelo	53
5.3. Evolución de la exactitud y la pérdida del segundo modelo	53
5.4. Matriz de confusión del segundo modelo	55
6.1. Evolución de la exactitud y la pérdida del séptimo modelo	62
6.2. Matriz de confusión del séptimo modelo	64
7.1. Imágenes sintéticas generadas con aumento de datos	70
7.2. Evolución de exactitud y pérdida del modelo final	71
7.3. Matriz de confusión del modelo final	73
7.4. Imágenes clasificadas incorrectamente por el modelo final	73
8.1. Descripción y resultados de experimentos por iteración	77
B.1. Arquitectura de modelo de CNN construida desde cero	94
B.2. Arquitectura de modelo de CNN con Transfer Learning de 'ResNet50' .	95
B.3. Arquitectura de modelo de CNN con Transfer Learning de 'VGG16' .	96

B.4. Arquitectura de modelo de CNN con Transfer Learning de 'VGG19' .	97
B.5. Arquitectura de modelo de CNN con Transfer Learning de 'InceptionV3'	98
B.6. Arquitectura de modelo de CNN con Transfer Learning de 'MobileNet'	99
B.7. Arquitectura de CNN con capas de 'Dropout' y 'Batch Normalization'	100

Índice de Tablas

2.1. Representación de matriz de confusión en clasificación binaria	25
4.1. Distribución de imágenes MRI de cerebros con HSVE del conjunto de datos 1	35
4.2. Distribución de imágenes MRI de cerebros sanos del conjunto de datos 2	39
4.3. Especificación de versiones de los paquetes software instalados	41
5.1. Distribución de imágenes MRI de cerebros con HSVE en los 3 conjuntos	47
5.2. Distribución de imágenes MRI de cerebros sanos en los 3 conjuntos .	47
5.3. Evaluación del primer modelo mediante la exactitud durante entrenamiento, validación y pruebas	52
5.4. Métricas de clasificación del primer modelo	52
5.5. Evaluación del segundo modelo mediante la exactitud durante entrenamiento, validación y pruebas	54
5.6. Métricas de clasificación del segundo modelo	55
5.7. Comparativa de exactitud y exhaustividad en modelos de CNN implementados en la iteración 1. Resaltado en azul más intenso se muestran los modelos de CNN con valores de exactitud más altos y en azul menos intenso aquellos con valores más bajos.	57
6.1. Subconjunto de imágenes MRI de cerebros con HSVE sólo del plano AXIAL a lo largo de los 3 conjuntos	60
6.2. Subconjunto de imágenes MRI de cerebros sanos sólo del plano AXIAL a lo largo de los 3 conjuntos	60

6.3.	Subconjunto de imágenes MRI de cerebros con HSVE sólo de los formatos FLAIR, T1 y T2 del plano AXIAL a lo largo de los 3 conjuntos	61
6.4.	Subconjunto de imágenes MRI de cerebros sanos sólo de los formatos FLAIR, T1 y T2 del plano AXIAL a lo largo de los 3 conjuntos	61
6.5.	Evaluación del séptimo modelo mediante la exactitud durante entrenamiento, validación y pruebas	63
6.6.	Métricas de clasificación del séptimo modelo	64
6.7.	Comparativa de exactitud y exhaustividad en modelos de CNN implementados en la iteración 2. Resaltado en azul más intenso se muestran los modelos de CNN con valores de exactitud más altos y en azul menos intenso aquellos con valores más bajos.	65
7.1.	Evaluación del décimo modelo mediante la exactitud durante entrenamiento, validación y pruebas	72
7.2.	Métricas de clasificación del décimo modelo	72
7.3.	Comparativa de exactitud y exhaustividad en modelos de CNN implementados en la iteración 3. Resaltado en azul más intenso se muestra el modelo de CNN con el valor de exactitud más alto y en azul menos intenso aquel con el valor más bajo.	75

Resumen

Con el propósito de desarrollar herramientas complementarias al diagnóstico médico de la enfermedad rara *encefalitis herpética* (HSVE), el presente trabajo realiza el diseño, implementación y evaluación de modelos de red neuronal convolucional (CNN) para clasificar imágenes de resonancia magnética (MRI) del cerebro humano según esté sano o presente dicha enfermedad. Se ha seguido una metodología incremental, de tal modo que la exactitud de clasificación obtenida en cada modelo trata de mejorar la obtenida en los modelos anteriores, alcanzando, finalmente, una exactitud en pruebas de 0.83 y una exhaustividad de 0.77. Se han empleado modelos pre-entrenados mediante *Transfer Learning* y aplicado técnicas de regularización para reducir el sobreajuste. Los experimentos realizados en este trabajo sientan las bases de futuros trabajos a mayor escala, en los que se entrenen y perfeccionen modelos de CNN empleando un mayor número de imágenes MRI, permitiendo obtener resultados más fiables.

Palabras Clave: *Deep Learning*, Encefalitis herpética (HSVE), Imágenes de Resonancia magnética (MRI), Red neuronal convolucional (CNN), *Transfer Learning*.

Abstract

To develop tools that support the medical diagnosis of the rare disease *herpetic encephalitis* (HSVE), this study design, implement and evaluate convolutional neural network (CNN) models to classify magnetic resonance images (MRI) of human brain as healthy or with this disease. The study has followed an incremental methodology, in which each model tries to improve the classification accuracy of the previous models. These CNN models use pre-trained models with *Transfer Learning* and apply regularization techniques to reduce overfitting. The latest model implemented achieved a test accuracy of 0.83 and a recall of 0.77. The experiments of this study can help in future large-scale works, in which CNN models can be trained and fine-tuned with a larger number of MRI images to obtain more reliable results.

Key words: Convolutional Neural Network (CNN), Deep Learning, Herpetic Encephalitis (HSVE), Magnetic Resonance Images (MRI), Transfer Learning.

Agradecimientos

En esta sección quiero agradecer, en primer lugar, a mi director del Trabajo Fin de Máster Guillermo Torralba Elipe. Su orientación, seguimiento y comentarios me han servido de guía y de gran ayuda durante el desarrollo del trabajo para la mejora de los modelos de clasificación construidos y entrenados. También me ha ayudado a adquirir buenas prácticas en la realización de trabajos de carácter académico o de investigación.

A mis compañeros del Máster por el compañerismo, buen ambiente y ayuda que me han proporcionado siempre que lo he necesitado. Especialmente gracias a mis amigos Emilio Delgado Muñoz y Arturo Barriga Corchero por ser excelentes compañeros y buenas personas.

A mis amigos y a mi familia por ser los pilares fundamentales de mi vida y por enseñarme los valores que me caracterizan a día de hoy. Especialmente a mis padres, Willi y Cristina, por haberme apoyado en todas las decisiones que he tomado durante mi etapa académica.

De verdad, muchísimas gracias a todos ellos.

Capítulo 1

Introducción

Las enfermedades raras son aquellas que afectan a un porcentaje muy pequeño de la población. Según el Reglamento del Parlamento y del Consejo Europeo (BOE, 2015), se considera una enfermedad rara a aquella que tiene una frecuencia menor a 5 casos por cada 10.000 habitantes. No obstante, esta definición varía dependiendo del país en cuestión.

En este contexto, la encefalitis herpética simple (HSVE) es una enfermedad rara que afecta a 1 de cada 250.000 o 500.000 habitantes. Esta enfermedad es causada por el virus del herpes simple y suele manifestarse en forma de fiebre, cefalea, sensación de cansancio, disfunciones cognitivas en el lenguaje y en la memoria o trastornos en el comportamiento. La ausencia de tratamiento o un tratamiento tardío sobre esta enfermedad puede llevar a un estado de salud grave. De hecho, ante ausencia del tratamiento la tasa de mortalidad de esta enfermedad es del 70 %, aunque se puede reducir hasta el 28 % después del tratamiento (Orphanet, 2015).

La llegada de las redes neuronales convolucionales con el trabajo de Yann LeCun para el reconocimiento de caracteres (Lecun et al., 1998), junto con la mejora en la velocidad de procesamiento, memoria y paralelismo en las Unidades de Procesamiento Gráfico (GPU's) ofertadas por la empresa NVidia en la última década, han permitido lograr numerosos avances en el campo del *Deep Learning* y, concretamente, en el campo de la visión por computador.

Gracias a estos avances, en los últimos años se han publicado numerosas investigaciones en el ámbito médico relativas a la aplicación de Inteligencia Artificial (IA) mediante el uso de técnicas de visión por computador para la detección de ciertas enfermedades. Concretamente, muchos de estos trabajos han aplicado modelos de red neuronal convolucional para clasificar imágenes de resonancia magnética (MRI) de cerebros humanos y poder así detectar enfermedades tales como tumores cerebrales (Khairandish et al., 2022) o trastornos del sistema nervioso como es el caso del Alzheimer. (Yamanakkavar et al., 2020).

1.1. Motivación

Dado el gran número de aplicaciones que tienen las técnicas de visión por computador mediante el uso de redes neuronales convolucionales en combinación con imágenes médicas de resonancia magnética, surge un gran interés en conocer el impacto que pueden tener estas técnicas en el diagnóstico de enfermedades raras tales como la HSVE. Dada la baja frecuencia de aparición de esta enfermedad en la población, existen pocos datos públicos disponibles asociados a casos diagnosticados de la enfermedad.

Por otro lado, la alta tasa de mortalidad de la enfermedad, si no es tratada esta de forma adecuada, hace que sea de vital importancia para el paciente un diagnóstico rápido y eficaz. Para ello, el desarrollo de sistemas artificiales que detecten la presencia de la enfermedad en imágenes MRI puede ayudar al especialista sanitario en el diagnóstico de la enfermedad, agilizando y acelerando así el proceso de su detección en la medida de lo posible, o incluso corrigiendo falsos diagnósticos ocasionados por el error humano.

1.2. Planteamiento del problema

En general, las enfermedades raras complican el estado clínico de las personas que las contraen de forma indefinida (Cortés, 2015). Además, estas enfermedades suelen presentar una elevada tasa de mortalidad. Éste es el caso de la HSVE, que presenta una mortalidad del 70 % antes del tratamiento y del 28 % después del tratamiento (Orphanet, 2015). Esto hace que un correcto diagnóstico del paciente, ante la posible presencia de esta enfermedad, sea un aspecto fundamental para que esta pueda ser debidamente tratada en el futuro. De este modo, se hace necesario el desarrollo de soluciones basadas en Inteligencia Artificial que identifiquen la presencia de enfermedades raras tales como la encefalitis herpética, permitiendo así ayudar o complementar el diagnóstico realizado de la enfermedad por el especialista médico.

1.3. Estructura del trabajo

Este proyecto se divide en 8 secciones principales con sus respectivos sub-apartados incluidos en ellas. Estas secciones son las siguientes:

- **Capítulo 1, Introducción:** en este capítulo introductorio se exponen las motivaciones y justificación del trabajo y se presenta el planteamiento del problema.
- **Capítulo 2, Contexto y Estado del Arte:** se contextualiza el trabajo describiendo aquellos conceptos y conocimientos necesarios para su correcto entendimiento. También se presentan los avances y resultados obtenidos en trabajos o investigaciones recientes relacionadas con el tema del presente proyecto.
- **Capítulo 3, Alcance y Objetivos:** en él se describe el objetivo general del proyecto, así como los objetivos específicos. También se describen las limitaciones del proyecto.
- **Capítulo 4, Material y Método:** en esta sección se describe el material utilizado, la metodología empleada y el proceso seguido para el diseño del

conjunto de experimentos realizados.

- **Capítulo 5, Iteración 1 de experimentos:** se describen los modelos de clasificación implementados y las técnicas aplicadas sobre los modelos de la primera iteración. Además, se discuten los resultados obtenidos en dichos modelos, comentando las diferencias y mejoras existentes entre ellos.
- **Capítulo 6, Iteración 2 de experimentos:** se describen los modelos de clasificación implementados y las técnicas aplicadas sobre los modelos de esta segunda iteración. Además, se discuten los resultados obtenidos sobre dichos modelos, comentando las diferencias y mejoras existentes entre ellos.
- **Capítulo 7, Iteración 3 de experimentos:** se describen los modelos de clasificación implementados y las técnicas aplicadas sobre los modelos de esta tercera y última iteración. Además, se discuten los resultados obtenidos sobre dichos modelos, comentando las diferencias y mejoras existentes entre ellos.
- **Capítulo 8, Conclusiones y Trabajos Futuros:** se indican las conclusiones del trabajo en base a los resultados obtenidos, así como los posibles trabajos que se podrían aplicar en el futuro.

Capítulo 2

Contexto y Estado del Arte

2.1. Encefalitis herpética simple (HSVE)

La encefalitis herpética es una enfermedad rara causada por el virus del herpes simple y que produce una inflamación en el cerebro (MayoClinic, 2023). Afecta a 1 de cada 250.000 o 500.000 personas y representa el 20 % de casos totales de encefalitis. Respecto a las zonas cerebrales implicadas, en un gran número de los casos diagnosticados la HSVE afecta a los lóbulos frontal y temporal, generalmente de forma asimétrica (Orphanet, 2015). De este modo, la enfermedad puede provocar disfunciones cognitivas en el lenguaje y en la memoria, así como trastornos en el comportamiento y en la personalidad.

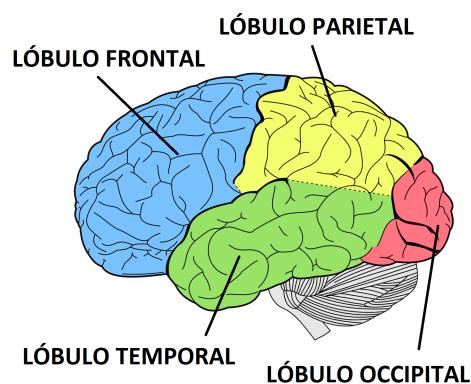


Figura 2.1: Localización de lóbulos cerebrales en zonas corticales. Nota. Adaptado del repositorio de imágenes gratuitas Pixabay, por ArtsyBee, 2015.

En la figura 2.1 se puede apreciar la distribución y disposición anatómica de los cuatro lóbulos que componen las áreas corticales del cerebro humano.

Por otro lado, cabe destacar que existen dos tipos de HSVE, según el tipo del virus del herpes que origine la encefalitis. El virus del herpes simple de tipo 1 (HSV-1), que representa el 90 % de los casos de la enfermedad, y el virus del herpes simple de tipo 2 (HSV-2), compuesto por el 10 % de los casos restantes. El virus de tipo 1 (HSV-1) se transmite principalmente mediante el contacto del virus en la zona de los ojos o de los labios de la boca, a través de úlceras o la saliva. Sin embargo, el segundo tipo de herpes se transmite principalmente mediante relaciones sexuales por el contacto de la piel, las superficies genitales o los líquidos de la persona contagiada. (WHO, 2023).

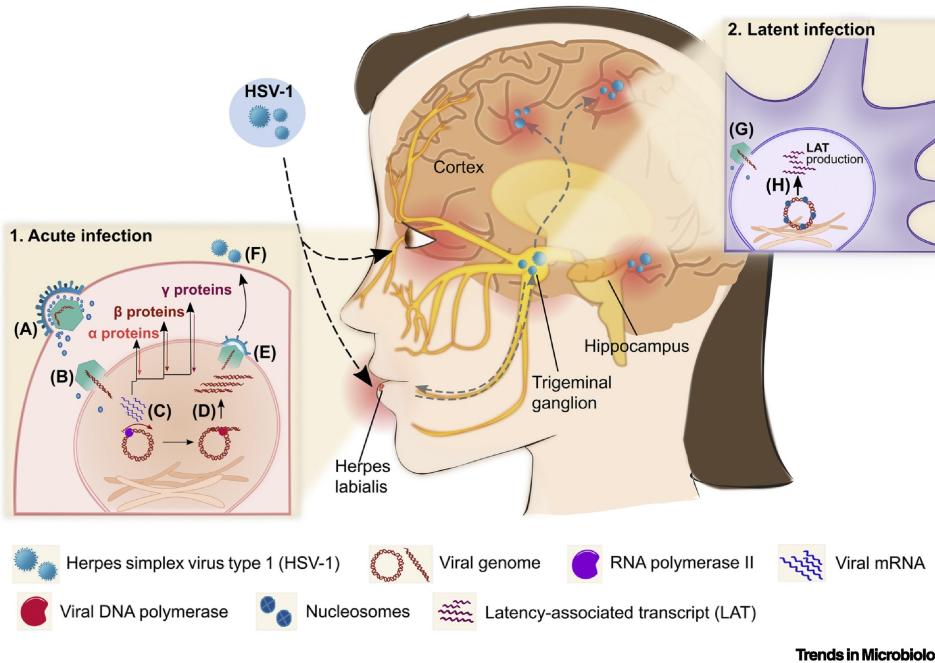


Figura 2.2: Esquema del proceso de transmisión del virus del herpes simple (HSV-1) hasta el cerebro, produciendo la enfermedad HSVE (Marcocci et al., 2020) Nota. Adaptado de Herpes simplex virus-1 in the brain: The dark side of a sneaky infection, por Trends in Microbiology, 2020.

En la figura 2.2 se expone de forma esquemática el proceso de infección del virus HSV-1 en seres humanos, llegando éste en ciertos casos hasta las regiones cerebrales del sistema nervioso y causando la HSVE. En primer lugar, el virus se transmite

a través de infecciones oculares o labiales en la boca (infección primaria). Después de dicha infección primaria, el virus se aloja en las neuronas sensoriales del sistema nervioso periférico (SNP), cercanas a la zona de la infección primaria. Desde las neuronas, el virus puede desplazarse sucesivamente a otras neuronas, propagándose a través del axón. De este modo, la replicación masiva del virus, llegando éste hasta el sistema nervioso central, es lo que causa la enfermedad HSVE (Maria Elena Marcocci and Chiara, 2020).

2.2. Resonancia Magnética Nuclear

Para estudiar el sistema nervioso humano, existen técnicas de diferente naturaleza. Dentro de las técnicas anatómicas, que permiten la identificación de las estructuras cerebrales implicadas en diferentes funciones y procesos cognitivos, se encuentran las técnicas de neuroimagen estructural, tales como la Resonancia Magnética Nuclear.

La Resonancia Magnética Nuclear (RMN) se basa en la aplicación de una onda electromagnética sobre los protones de los núcleos de hidrógeno del oxígeno presente en el cerebro, alterando así su orientación espacial. Inicialmente, estos átomos se encuentran orientados al azar. Sin embargo, tras aplicar la onda electromagnética sobre los protones, estos orientan sus ejes en la misma dirección. Una vez que los núcleos de hidrógeno pierden la energía que les ha hecho orientarse, un sensor de campo magnético altamente sensible y preciso capta las variaciones de energía emitidas por los núcleos de hidrógeno. A partir de estos datos de la señal recibida, se construyen representaciones en 2D y 3D del cerebro, donde los diferentes niveles de gris se asocian con las diferencias de composición y densidad entre diferentes áreas nerviosas. A estas representaciones se les conoce por el nombre de Imágenes de Resonancia Magnética (MRI) (Muñoz Garrido, 2022). Existen 3 tipos de MRI del cerebro dependiendo del plano que se utilice para capturar cada una de las secciones del cerebro. Estos planos son el plano coronal o frontal, el sagital y el axial u horizontal.

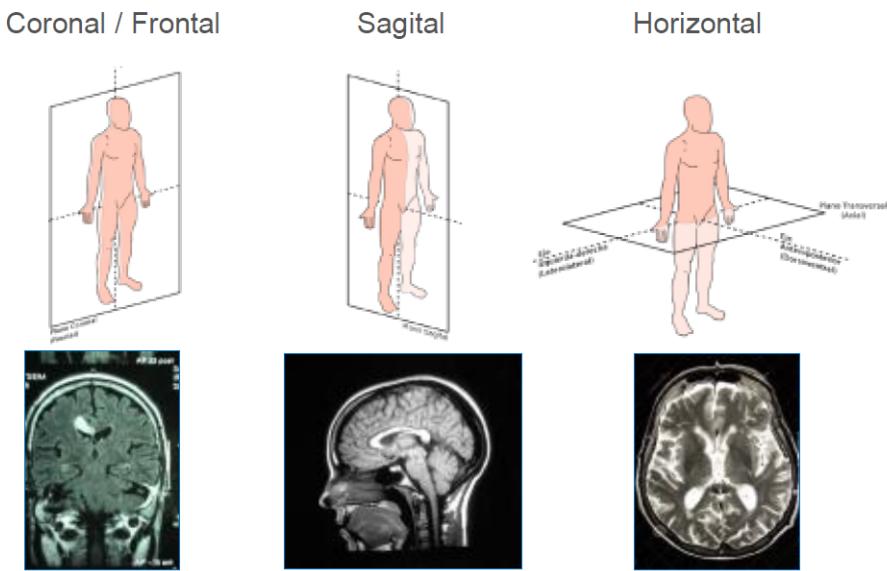


Figura 2.3: Ejemplos de imágenes MRI del encéfalo mediante capturas en el plano coronal, sagital y horizontal (plano axial). Fuente: Material académico de UNIR (Muñoz Garrido, 2022)

En la figura 2.3 se pueden apreciar imágenes MRI de ejemplo del encéfalo humano capturadas sobre los planos coronal, sagital y axial u horizontal respectivamente.

En el ámbito médico, estas imágenes de resonancia magnética son de gran utilidad ya que permiten detectar lesiones o anomalías en el organismo. Concretamente, las imágenes MRI obtenidas del cerebro son un factor clave en la detección de enfermedades, como tumores cerebrales, o trastornos, como es el caso del Alzheimer. Además, en el ámbito de la Inteligencia Artificial estas imágenes son empleadas para la construcción de sistemas que permitan su segmentación, para la distinción de las diferentes partes de la imagen, o para su clasificación, como es el caso del presente trabajo. De este modo, es posible extraer información relevante relativa a dichas enfermedades y poder así dar soporte en el diagnóstico médico del paciente.

2.3. El Machine Learning

El *Machine Learning* (ML) o aprendizaje automático es un subcampo de la Inteligencia Artificial (IA) dedicado a la construcción, entrenamiento y evaluación de modelos algorítmicos que permitan a los computadores identificar patrones sobre conjuntos de datos, así como la elaboración de predicciones (Geron, 2019).

Dependiendo del nivel de supervisión del aprendizaje que tenga el sistema durante el entrenamiento, los sistemas de ML se pueden clasificar en los siguientes grupos:

1. **Aprendizaje supervisado:** en este tipo de aprendizaje, el modelo de ML es entrenado empleando un conjunto de datos conocidos y correctamente etiquetados, es decir, identificados con su salida. El objetivo del entrenamiento del modelo será el de evaluar dicho modelo prediciendo la salida de datos desconocidos.

Los 2 algoritmos más empleados en el aprendizaje supervisado son la clasificación y la regresión. El presente trabajo se centra en un problema de clasificación, en el que se entrena un modelo mediante un conjunto de entrenamiento, se valida el entrenamiento mediante un conjunto de validación y, finalmente, se evalúa el modelo entrenado con un conjunto de pruebas (test). Dependiendo del número de clases que pretenda clasificar el modelo, la clasificación será binaria (2 clases) o multi-clase (3 o más clases).

2. **Aprendizaje no supervisado:** en este nivel de supervisión, los datos del conjunto de entrenamiento se encuentran sin etiquetar, es decir, no se identifica la salida o resultado asociado a cada muestra. El objetivo del sistema en este tipo de aprendizaje consiste en encontrar patrones o relaciones desconocidas en los datos sin la necesidad de que alguien o algo le ayude etiquetando los datos de entrada, agrupando así las muestras en grupos o clústers con características similares (“clustering”).

3. **Aprendizaje semi-supervisado:** en este nivel de supervisión, se etiquetan parcialmente los datos de entrenamiento del modelo. Debido a que el proceso de

etiquetado de los datos suele ser un proceso costoso, etiquetarlos parcialmente puede disminuir este coste.

4. **Aprendizaje por refuerzo:** resulta bastante diferente con respecto a los tipos de aprendizaje mencionados previamente. El sistema de aprendizaje parte de un agente que observa el entorno y que selecciona y ejecuta una acción siguiendo una política de entre un conjunto de acciones disponibles (por ejemplo, cada acción tiene un peso diferente). En función de la acción ejecutada, el agente recibirá una recompensa positiva, si elige bien la acción, o una recompensa negativa o multa, en caso de que se equivoque. De este modo, el agente aprende por sí mismo trazando una estrategia en la que maximiza la recompensa obtenida a lo largo del tiempo. Una política será la encargada de indicar la acción que debe elegir el agente ante una determinada situación.

En la figura 2.4 se aprecian en forma de diagrama los diferentes tipos de sistemas de *Machine Learning* descritos previamente. Además, se señalan los algoritmos de clasificación, dado que son los empleados en los experimentos del presente proyecto.

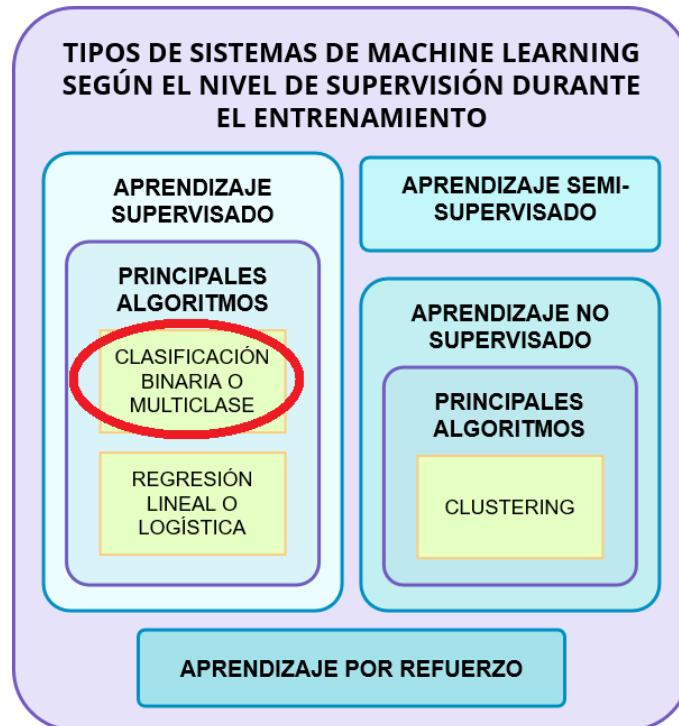


Figura 2.4: Tipos de sistemas de Machine Learning

2.4. El Deep Learning

El aprendizaje profundo o *Deep Learning* (DL) es un subcampo de la Inteligencia Artificial y, a su vez, del *Machine Learning* dedicado principalmente al desarrollo de sistemas predictivos que encuentren patrones sobre grandes cantidades de datos, empleando para ello arquitecturas algorítmicas basadas en redes neuronales artificiales con un elevado número de capas. Además, se aplica un enfoque *end-to-end*, el cual permite abstraer gran parte del procesamiento en las capas internas del algoritmo y centrar el problema en el diseño de los experimentos. En este diseño intervienen aspectos tales como la preparación de los datos, la elección o construcción de la arquitectura de red y el ajuste de hiper-parámetros, entre otros. Estos sistemas (una vez han sido entrenados) reciben una entrada, tal como una imagen o una pista de audio, y devuelven una salida con el resultado de predicción de la imagen o con una posible transcripción del audio recibido en la entrada.

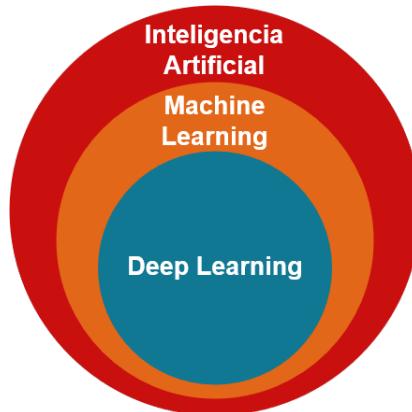


Figura 2.5: *Deep Learning* como subcampo del ML y de la IA

La principal diferencia con respecto al *Machine Learning* es que los algoritmos de *Deep Learning* logran identificar patrones sobre mayores cantidades de datos mediante diferentes representaciones de los mismos (Chollet, 2018). Estas diferentes representaciones sobre los datos se obtienen a través de las múltiples capas por las que está constituida una red neuronal (*neural network* en inglés). En la actualidad, existe un gran número de arquitecturas de red neuronal. Destacan, por ejemplo, las redes neuronales convolucionales (CNN), ampliamente utilizadas en el campo de la visión por computador, o las redes neuronales recurrentes (RNN), dedicadas al

procesamiento de secuencias temporales y empleadas, por ejemplo, en el campo del procesamiento natural del lenguaje.

Gracias al progreso logrado en las últimas décadas en el ámbito de las redes neuronales, se han desarrollado numerosos y exitosos sistemas dedicados al reconocimiento de imágenes, reconocimiento facial, sistemas de traducción, segmentación de imágenes, etc.,.

La siguiente sección se centrará en explicar brevemente el funcionamiento de las redes neuronales convoluciones, al tratarse de la arquitectura de red neuronal empleada en los experimentos de este proyecto.

2.4.1. Redes Neuronales Artificiales (ANN)

Una red neuronal artificial es un algoritmo de *Deep Learning* empleado para resolver una gran variedad de tareas predictivas, generalmente tareas de clasificación o regresión, aunque también se pueden emplear para segmentar imágenes o procesar lenguaje natural, entre otras aplicaciones.

Una red neuronal está constituida por 1 capa de entrada, al menos una capa oculta y una capa de salida. Esta distribución de capas de la red neuronal se ilustra en la figura 2.6.

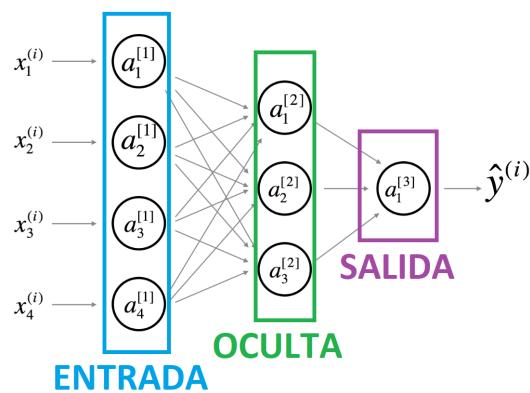


Figura 2.6: Organización por capas de una red neuronal (Andrew Ng, 2021).

Fuente: Programa de especialización de Deep Learning en Coursera

Cada capa de la red está formada por un número diferente de neuronas. Además, cada neurona tiene los siguientes parámetros:

1. Un sesgo “b”
2. Tantos pesos “W” como entradas reciba la neurona en cuestión

Tras conocer los parámetros de cada neurona, resulta necesario comentar cómo se realiza el entrenamiento de la red neuronal. Éste se divide en 2 fases:

Forward-propagation

Durante esta fase, todos y cada uno de los datos de entrenamiento se desplazan a través de las neuronas de las distintas capas de la red, desde la primera capa hasta la última. En cada neurona se realiza un procesamiento dividido en 2 fases. En la primera fase, se calcula el valor de la variable “z” de la siguiente forma:

$$z = W^T \cdot x + b \quad (2.1)$$

Una vez hecho esto, en la segunda fase se aplica la función de activación, denotada por “g”, sobre el resultado anterior obtenido en “z”, obteniendo la salida sobre la variable “a”, tal y como se muestra en la siguiente ecuación:

$$a = g(z) \quad (2.2)$$

Cabe destacar que existen diferentes tipos de funciones de activación tales como la reLu, Leaky reLu, Sigmoide, Tangente, prelu, softmax y muchas otras. Las funciones de activación son aquellas funciones que proporcionan no linealidad al modelo predictivo. Una vez se ha aplicado la función de activación sobre el valor “z”, se desplaza el resultado denotado por “a”, a la entrada de las neuronas de la siguiente capa. La ecuación general que se aplica en cada neurona sobre cada dato de entrada, denotado por “x”, es la que se muestra a continuación:

$$a = g^{[l]}(W_x \cdot x^{(i)} + b_1) = g^{[l]}(z_1) \quad (2.3)$$

Resulta necesario aclarar que $g^{[l]}$ denota la función de activación a aplicar sobre la capa “l”. Este procesamiento se ilustra en la figura 2.7.

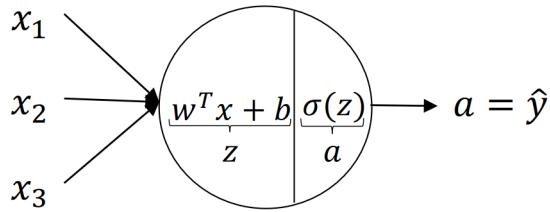


Figura 2.7: Procesamiento sobre 1 neurona (Andrew Ng, 2021). Fuente: Programa de especialización de Deep Learning en Coursera

Sobre cada dato o muestra que es procesado por la red neuronal de izquierda a derecha se realiza una predicción en la capa de salida de la red. Una vez obtenidas todas las predicciones sobre cada uno de los datos, se obtiene el valor de una función de pérdida. Este valor medirá el error cometido por el modelo en las predicciones, evaluando la diferencia entre el valor real y el valor predicho. Existen diferentes funciones de error o de pérdida tales como el error absoluto medio (*Mean Absolute Error (L1 Loss)*) o el error cuadrático medio (*Mean Squared Error (L2 Loss)*).

Backward-propagation

En esta segunda fase del entrenamiento de la red se modifican los parámetros de la red (pesos y sesgos de cada neurona) con el objetivo de mejorar el valor de las predicciones en la siguiente iteración del entrenamiento. Para calcular el valor de los nuevos parámetros, se realiza el cálculo del gradiente a partir de la función de pérdida y las funciones de activación de cada capa. Dado que la mayoría de las funciones de activación son derivables, es posible calcular su gradiente, permitiendo propagar el error calculado sobre la función de pérdida hacia atrás de derecha a izquierda por la red. De este modo, se van modificando los valores de los parámetros de la red siguiendo el valor del gradiente. Cuanto mayor sea el gradiente obtenido, mayor será la modificación de los parámetros de la neurona en cuestión.

Una vez terminadas las dos fases sobre todos los datos de entrenamiento, se da por concluida la primera época (*epoch*) del entrenamiento de la red neuronal. Un

hiper-parámetro fundamental de la red es el número de *epochs* que dedicará la red a entrenar el modelo, teniendo en cuenta que cada *epoch* ejecutará los 2 procesos mencionados previamente, partiendo de los parámetros (pesos W y sesgos b) definidos en la época anterior. Una inicialización muy frecuente de los parámetros de la red es la inicialización aleatoria. No se deben inicializar los parámetros con el valor 0 dado que, de este modo, al calcular los gradientes de las funciones de activación, los pesos se modificarían en cada iteración por igual, lo cual impediría encontrar patrones sobre los datos. Para encontrar dichos patrones, la inicialización aleatoria de los parámetros modifica en mayor medida unos parámetros y en menor medida otros, proporcionando así mayor relevancia a ciertos parámetros y permitiendo la búsqueda de patrones en los datos.

2.4.2. Redes Neuronales Convolucionales (CNN)

Las redes neuronales convolucionales son un tipo de red neuronal artificial que trata de simular el funcionamiento de las neuronas de la corteza visual del sistema nervioso humano. Estas redes son ampliamente utilizadas en problemas de clasificación y segmentación de imágenes, y constituyen un pilar fundamental en la construcción de sistemas de visión por computador para la búsqueda de patrones en las imágenes, permitiendo la identificación de objetos, personas, animales o de cualquier otra entidad que sea objeto de ser identificada.

Para detectar patrones en las imágenes, estas redes aplican operaciones de convolución y de *pooling* mediante el uso de filtros. Estas operaciones se emplean a nivel de capa durante el proceso de construcción de la red neuronal convolucional, de tal modo que existen capas de convolución y capas de *pooling*. Un aspecto fundamental a tener en cuenta en este tipo de arquitectura es que las primeras capas de la red suelen dedicarse a la detección de bordes mediante el uso de filtros paso alto (características de bajo nivel), las capas intermedias a la detección de formas u objetos parciales y las últimas capas de la red a la detección de los objetos completos (características de alto nivel). Esta idea es ilustrada en la figura 2.8.

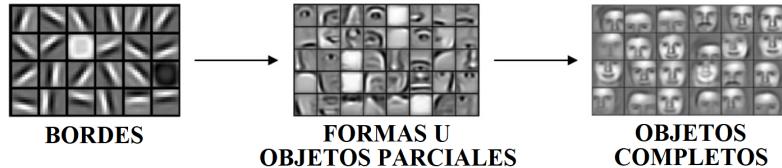


Figura 2.8: Detección de bordes, formas y objetos mediante operaciones de convolución en diferentes regiones de una red neuronal (Andrew Ng, 2021). Fuente: Programa de especialización de Deep Learning en Coursera

2.4.3. VGG

Visual Geometry Group o VGG (Simonyan and Zisserman, 2014) es un tipo de red neuronal convolucional que presenta un alto rendimiento en la construcción de sistemas de reconocimiento de objetos. Concretamente, esta arquitectura ha logrado superar el 92 % de exactitud en la clasificación de objetos del famoso conjunto de datos *ImageNet* (Deng et al., 2009), el cual está compuesto por más de 14 millones de imágenes etiquetadas en 1000 clases diferentes. De este modo, la arquitectura de red profunda VGG se ha convertido en una de las más populares en tareas de clasificación de imágenes. Existen dos principales implementaciones, la arquitectura VGG-16 y la arquitectura VGG-19. La arquitectura VGG-16, la cual se ilustra en la figura 2.9 está compuesta por un total de 16 capas (13 capas de convolución + 3 capas *fully connected* o completamente conectadas). La única diferencia entre VGG-16 y VGG-19 es que la arquitectura VGG-19 presenta 3 capas convolucionales más (16) respecto a las que presenta VGG-16 (13).

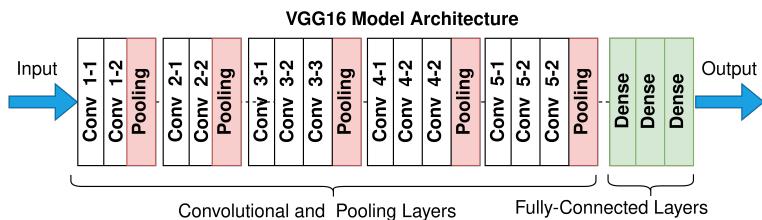


Figura 2.9: Arquitectura de VGG-16. Fuente: Wikimedia Commons

2.4.4. Redes Neuronales Residuales (ResNets)

Uno de los problemas que puede aparecer en las redes neuronales artificiales durante la fase de *back-propagation* es el problema del desvanecimiento del gradiente (*gradient vanishing problem*). Este problema surge en aquellas neuronas en las que la función de activación hace que los gradientes presenten valores muy bajos. Estos valores de gradiente tan bajos y cercanos a 0, tienen como consecuencia una baja modificación en los valores de los parámetros de la neurona asociada y, por lo tanto, estas neuronas aprenderán muy poco o nada.

Con el objetivo de solventar este error, en 2015 surgen las *Residual Neural Networks* o ResNets (He et al., 2015). Estas redes profundas implementan lo que se denominan atajos, *skip connections* o *shortcuts*. Estos atajos consisten en saltos entre 2 o incluso 3 capas, permitiendo así alimentar una capa con las salidas, no sólo de la capa anterior, sino de otras capas anteriores. Esta situación se puede visualizar en la figura 2.10.

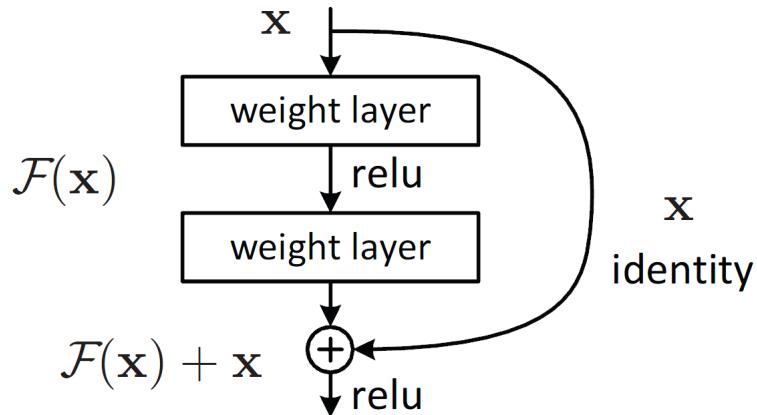


Figura 2.10: Bloque de aprendizaje residual mediante atajos (He et al., 2015)

Gracias al uso de las ResNets es posible, no sólo solventar el problema del desvanecimiento del gradiente, sino también optimizar la red neuronal y proporcionar mecanismos que faciliten el flujo de la información a través de las capas de forma no lineal (mediante saltos entre capas).

2.4.5. InceptionV3

La arquitectura de red neuronal convolucional InceptionV3, desarrollada por investigadores de Google en 2015 (Szegedy et al., 2015), está constituida por una serie de bloques de convolución llamados módulos “Inception”. Estos bloques, que se pueden apreciar en la figura 2.11, aplican en paralelo varias operaciones convolucionales de diferentes tamaños del filtro junto con una operación de *pooling* para extraer características pertenecientes a diferentes niveles de abstracción. Por ejemplo, un filtro de tamaño 3x3 es posible que sea capaz de detectar un borde horizontal (característica de bajo nivel), mientras que un filtro de 5x5 es posible que reconozca la forma completa de la boca de una persona (característica de alto nivel) a la que pertenecía dicho borde horizontal.

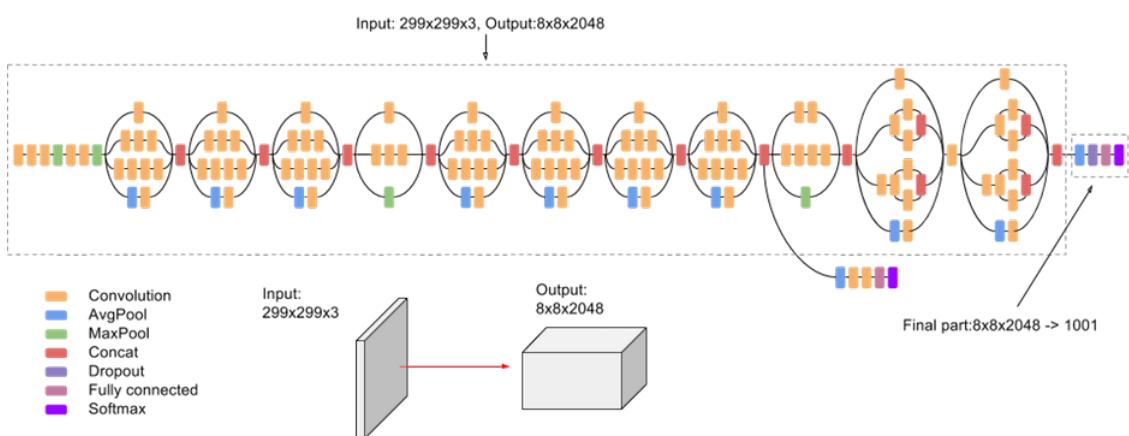


Figura 2.11: Arquitectura de InceptionV3 (Szegedy et al., 2015). Fuente: Google Cloud

La arquitectura *InceptionV3* es utilizada principalmente para la resolución de problemas relacionados con la clasificación de imágenes y detección de objetos.

2.4.6. MobileNet

La arquitectura de red neuronal convolucional *MobileNet*, desarrollada también por investigadores de Google en 2017 (Howard et al., 2017) (2 años después de lanzar InceptionV3), es ampliamente utilizada en tareas de visión por computador en dispositivos móviles. Esto se debe al uso eficiente que realiza de los recursos limitados

en estos dispositivos. A diferencia de otras arquitecturas vistas previamente como ResNet o VGG16, MobileNet logra reducir en gran medida el número de parámetros de la red mediante la aplicación de varias técnicas, entre las que destaca el uso de operaciones de convolución separables en profundidad (*depth-wise convolution* en inglés). Este tipo de operación de convolución consta de tres fases: en primer lugar, se aplica un filtro diferente para cada canal de la imagen (convolución de tipo *Depthwise*), posteriormente se combinan los resultados de la convolución anterior mediante la operación de convolución *Pointwise* y, finalmente, se aplica una función de activación a los resultados de la operación anterior. Estas operaciones se pueden apreciar de forma esquemática en la figura 2.12, donde se ilustra la arquitectura MobileNet.

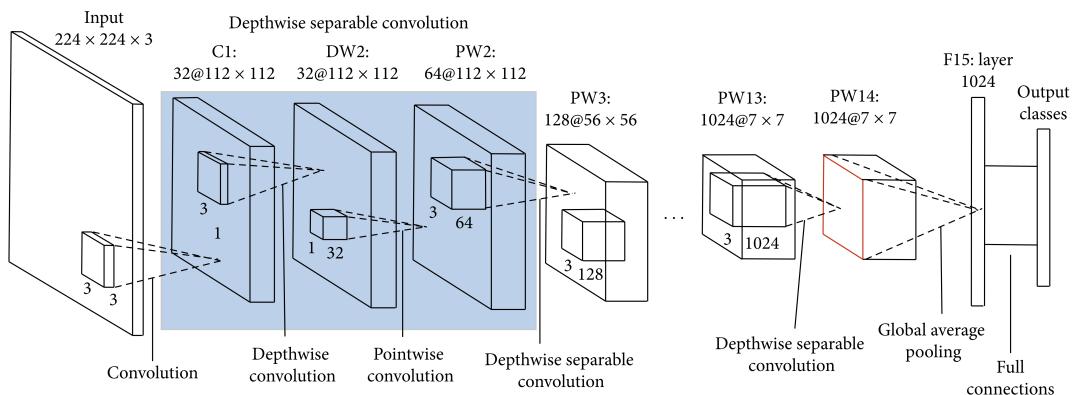


Figura 2.12: Arquitectura de MobileNet (Wang et al., 2020). Nota. Adaptado de A Novel Image Classification Approach via Dense-MobileNet Models, por Indawi, 2020.

De este modo, la arquitectura *MobileNet* logra alcanzar, de forma eficiente y mediante recursos limitados, buenos resultados en problemas de clasificación de imágenes, detección de objetos, reconocimiento facial o segmentación semántica.

2.4.7. Transfer learning

La técnica conocida como *transfer learning* (“transferencia de aprendizaje” en español) consiste en importar las características extraídas de un modelo de red neuronal previamente entrenado y aprovecharlas en el entrenamiento de una red neuro-

nal de un problema similar (Chollet, 2018). Para ello, resulta necesario aplicar los siguientes pasos:

1. Importar un modelo base pre-entrenado con sus correspondientes pesos, obtenidos mediante un entrenamiento previo de dicho modelo.
2. Bloquear o congelar (*freeze*) estos pesos pre-entrenados para así mantener intactas estas características extraídas.
3. Eliminar la última capa del modelo base importado y añadir nuevas capas sobre dicho modelo base.
4. Finalmente, re-entrenar la red con el nuevo conjunto de datos. De este modo, sólo se actualizarán los pesos de las nuevas capas añadidas.

Esta técnica es ampliamente utilizada para problemas en los que se dispone de pocos datos, como es el caso del presente proyecto en el que se dispone de un total de 745 imágenes MRI. De este modo, se aprovechan las características extraídas durante el entrenamiento previo realizado sobre otro problema de clasificación.

2.4.8. Técnicas de regularización

El sobreajuste u *overfitting* es una situación que se produce cuando el modelo se adapta tan bien a los datos de entrenamiento que no es capaz de generalizar adecuadamente, fallando al predecir la clase de ciertos ejemplos de los conjuntos de validación o de pruebas. Para reducir este sobre-ajuste en redes neuronales, se utilizan las técnicas de regularización. Algunas de estas técnicas son la regularización de tipo L1 y L2 (*L1/L2 Regularization*), *Dropout*, *Batch Normalization*, aumento de datos (*Data Augmentation*) y *Early Stopping*. En esta sección se describen 4 de estas técnicas de regularización, utilizadas para el desarrollo de este proyecto:

Early Stopping

Esta técnica permite detener el entrenamiento de un modelo de red neuronal artificial en caso de que se satisfaga una determinada condición. En el presente

proyecto, se ha empleado esta técnica para detener el entrenamiento si la exactitud o la pérdida en el conjunto de validación no mejora en 10 épocas (*epochs*) consecutivas.

Batch Normalization

Esta técnica consiste en normalizar la salida obtenida en una capa de la red neuronal antes de pasarla a la capa siguiente. Al ajustar los valores de la salida de dicha capa para que tengan una media cercana a cero y una desviación estándar cercana a 1, se resuelve el problema del desvanecimiento del gradiente, es decir, se evita que se realicen cambios cada vez más pequeños en los parámetros de la red a medida que se va entrenando el modelo. Además, al normalizar los valores de las salidas de las capas en cuestión de la red neuronal, *Batch Normalization* permite que dichos valores no adquieran valores demasiado grandes o pequeños, facilitando así tanto el aprendizaje como la convergencia del modelo.

Dropout

Esta técnica consiste en la desactivación aleatoria de un porcentaje de las neuronas de una capa oculta de la red neuronal. Para ello, resulta necesario definir previamente un valor comprendido entre 0 y 1 donde, por ejemplo, un valor de 0.5 significará que la salida del 50 % de neuronas aleatorias de la capa oculta en cuestión tendrá la salida con el valor 0 (se desactiva aleatoriamente el 50 % de las neuronas de la capa donde se emplee dicho Dropout). Además, para cada capa oculta de la red, es posible definir un valor de *Dropout* diferente. Durante la evaluación del modelo, se tendrán en cuenta todas las neuronas de la red, por lo que no se desactivará ninguna neurona como sí aplica Dropout durante el entrenamiento (Chollet, 2018).

En el ejemplo que se puede apreciar en la figura 2.13, se ha definido el mismo valor 0.5 de Dropout para cada una de las capas de la red neuronal.

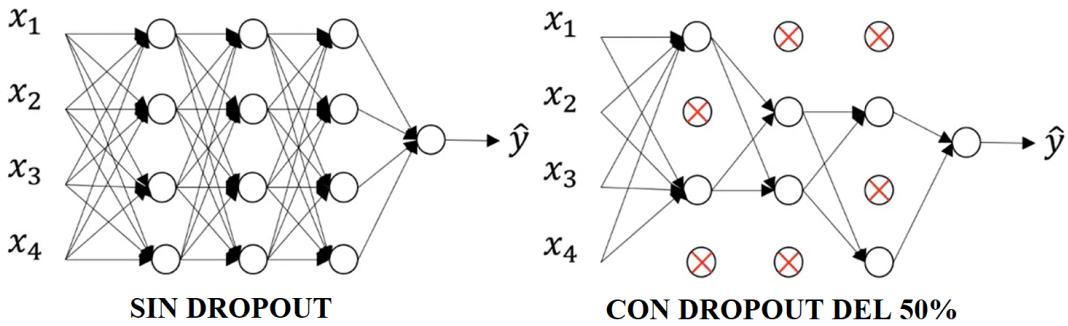


Figura 2.13: Aplicación de un Dropout del 50 % en todas las capas de la red neuronal (Andrew Ng, 2021). Fuente: Programa de especialización de Deep Learning en Coursera

Esta técnica es empleada para evitar que ciertas neuronas memoricen patrones relativos a detalles no relevantes en los datos de entrada, que es precisamente lo que sucede con el sobre-ajuste u *overfitting* de un modelo, el cual se intente reducir mediante el uso de estas técnicas.

Aumento de datos

Esta técnica consiste en la aplicación de distintas transformaciones sobre las muestras originales con el objetivo de generar nuevas muestras ligeramente diferentes y que serán utilizadas para entrenar el modelo de red neuronal.

Esta técnica resulta de gran utilidad en visión artificial, donde se aplican distintas transformaciones sobre las imágenes, o incluso combinaciones de dichas transformaciones, para generar nuevas imágenes llamadas imágenes sintéticas. Es importante destacar que estas nuevas imágenes sólo se deben utilizar para entrenar el modelo de red neuronal convolucional y en ningún caso para evaluar dicho modelo. De este modo, el aumento de datos ayuda a disponer de un mayor número de imágenes y conseguir así un modelo con un mayor nivel de generalización, capaz de predecir un mayor número de imágenes desconocidas. Algunas de las transformaciones más utilizadas aplicadas sobre las imágenes originales son las siguientes:

1. Voltrear la imagen en horizontal / vertical.

2. Desplazar un porcentaje de la imagen o una serie de píxeles de la imagen en el eje horizontal o vertical (mover la imagen hacia arriba, abajo, izquierda o derecha).
3. Recortar, añadir relleno, redimensionar, etc.
4. Aplicar deformaciones de perspectiva.
5. Disminuir o aumentar el brillo de la imagen.
6. Introducir ruido sobre la imagen.
7. Rotar la imagen X grados.
8. Aplicar combinaciones de las transformaciones anteriores.

Cabe destacar que resulta de vital importancia seleccionar correctamente las transformaciones o combinaciones de transformaciones a aplicar en el aumento de datos. Esto se debe a que el objetivo en todo momento es generar imágenes similares a las originales. En caso de que se deformasen lo suficiente como para modificar las características de las imágenes originales, se podría llegar a introducir ruido en el modelo, empeorando así la tarea de clasificación sobre nuevas imágenes.

2.5. Métricas de evaluación del modelo de clasificación

Para poder evaluar los modelos de clasificación implementados en este trabajo, resulta fundamental utilizar métricas que permitan analizar los resultados de forma adecuada. Antes de describir las métricas de clasificación más importantes, se procede a explicar los siguientes términos. Para ello, se toma como ejemplo el problema clásico de clasificación binaria en el que se pretende clasificar cada imagen en una de 2 posibles clases en función de si existe algún gato en la imagen (clase positiva o '1') o no (clase negativa '0'):

1. **Verdaderos positivos (*True Positive*)**: se trata de aquellas muestras o datos que han sido predichos con la clase positiva por el modelo de clasificación, perteneciendo realmente estas a dicha clase. Esta métrica también se hace referencia con el término 'TP'. En el ejemplo se trataría de las predicciones que han clasificado correctamente las imágenes en las que aparece algún gato con la clase positiva, indicando así que efectivamente existe algún gato en la imagen.
2. **Verdaderos negativos (*True Negative*)**: se trata de aquellas muestras o datos que han sido predichos con la clase negativa por el modelo de clasificación, perteneciendo realmente estas a dicha clase. Esta métrica también se hace referencia con el término 'TN'. En el ejemplo se trataría de las predicciones que han clasificado correctamente las imágenes en las que no aparece ningún gato con la clase negativa, indicando así que efectivamente no existen gatos en dichas imágenes.
3. **Falsos positivos (*False Positive*)**: se trata de aquellas muestras o datos que han sido predichos con la clase positiva por el modelo de clasificación pero realmente no pertenecen a dicha clase. Esta métrica también se hace referencia con el término 'FP'. En el ejemplo se trataría de las predicciones que han clasificado incorrectamente las imágenes, indicando mediante la clase positiva que sí existe algún gato, a pesar de que en ellas no aparece gato alguno.
4. **Falsos negativos (*False Negative*)**: se trata de aquellas muestras o datos que han sido predichos con la clase negativa por el modelo de clasificación pero realmente no pertenecen a dicha clase. Esta métrica también se hace referencia con el término 'FN'. En el ejemplo se trataría de las predicciones que han clasificado incorrectamente las imágenes, indicando mediante la clase negativa que no existe gato alguno, a pesar de que en ellas sí aparece algún gato.

En la tabla 2.1, se ilustran de forma esquemática cada uno de los términos mencionados previamente, teniendo en cuenta que existen datos correspondientes a 2

clases sobre los que se realizan predicciones: la clase '0' y la clase '1'. Esta tabla, en la cual se relacionan las predicciones realizadas por un modelo de clasificación sobre los datos y las etiquetas reales de dichos datos, recibe el nombre de matriz de confusión y supone una gráfica fundamental para la evaluación de modelos de clasificación en aprendizaje supervisado.

		PREDICCIÓN	
		0	1
REALIDAD	0	TN	FP
	1	FN	TP

Tabla 2.1: Representación de matriz de confusión en clasificación binaria

A continuación, se describen las principales métricas de clasificación:

1. **Exactitud (Accuracy)**: indica la proporción de muestras que el modelo de clasificación ha predicho o acertado correctamente con respecto al total. Se calcula como la división del número de muestras predichas correctamente (verdaderos positivos + verdaderos negativos) entre el número de muestras totales de la siguiente forma:

$$\text{Exactitud (Accuracy)} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.4)$$

2. **Pérdida (Loss)**: esta métrica de evaluación mide el error existente en las predicciones del modelo, es decir, la diferencia entre las predicciones realizadas y los valores reales de las etiquetas de las muestras. Para calcular esta métrica, se emplea una función de pérdida, cuya implementación puede variar dependiendo del problema en cuestión. Durante el entrenamiento del modelo, se tratará de minimizar el valor de pérdida lo máximo posible.
3. **Precisión (Precision)**: indica la proporción de muestras o datos de una clase que han sido predichas correctamente. Se calcula como la división del número de muestras de la clase X predichas correctamente (verdaderos positivos) entre el número de muestras predichas con dicha clase (positivos) del siguiente modo:

$$\text{Precisión (Precision)} = \frac{TP}{TP + FP} \quad (2.5)$$

4. **Exhaustividad (Recall)**: mide la capacidad del modelo de clasificar muestras de una clase. Se calcula como la división del número de muestras de la clase X predichas correctamente (verdaderos positivos) entre el número de muestras que realmente pertenecen a dicha clase (verdaderos positivos + falsos negativos). La ecuación es la siguiente:

$$\text{Exhaustividad (Recall)} = \frac{TP}{TP + FN} \quad (2.6)$$

Junto a la exactitud, la exhaustividad será una de las principales métricas a tener en cuenta al evaluar los modelos de CNN de este trabajo. Al tratarse de un problema médico, con el objetivo de evitar el error de no diagnosticar

la enfermedad HSVE, lo cual podría llegar a ser la causa de la muerte de una persona, resulta fundamental tener en cuenta los falsos negativos del modelo en forma de penalización. De este modo, cuanto mayor sea el número de falsos negativos, la exhaustividad adquirirá un valor menor, mientras que si el número de falsos negativos es menor, la exhaustividad adquirirá un valor mayor.

5. **Valor F (*F1-score*)**: este valor se obtiene combinando los valores de las métricas de precisión y exhaustividad en un sólo valor. De este modo, mediante la combinación de la precisión y de la exhaustividad resulta más fácil comparar el rendimiento de varios modelos. *F1-score* se calcula haciendo la media armónica entre la precisión y la exhaustividad de la siguiente forma:

$$\text{Valor } F \text{ (} F1\text{-score) } = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.7)$$

2.6. Investigaciones recientes relativas al tema del proyecto

En los últimos años, numerosas investigaciones en el campo de la visión por computador (*Computer Visión*) han permitido la construcción de sistemas capaces de clasificar imágenes de resonancia magnética del cerebro humano mediante el uso de redes neuronales convolucionales.

En el ámbito de la clasificación de encefalitis, destaca el trabajo realizado por Yayun Xiang junto a otros investigadores (Xiang et al., 2021). En dicho trabajo se implementaron varios modelos de *Deep Learning* en los que, mediante el uso de imágenes MRI multisecuencia, se logró su identificación y clasificación obteniendo una exactitud en pruebas de 0.831 sobre las imágenes MRI de encefalitis autoinmune, de 0.882 sobre MRI de encefalitis herpética y de 0.892 sobre MRI de sujetos de control sanos.

Otras investigaciones recientes han construido modelos de CNN con el propósito de clasificar imágenes MRI de cerebros humanos según diferentes tipos de tumores cerebrales. Un ejemplo de ello es el estudio realizado por Chetana Srinivas junto a otros

investigadores (Srinivas et al., 2022), en el que se emplean modelos pre-entrenados de red convolucional de diferentes arquitecturas, haciendo uso de la técnica *Transfer Learning*, con el objetivo de evaluar su rendimiento en la clasificación binaria de imágenes MRI de cerebros sanos o con tumores. El dataset empleado en esta investigación fue obtenido del repositorio público Kaggle. Una vez importado correctamente el conjunto de datos, se aplicó una fase de preprocesamiento en la que se reescalaron las imágenes para que todas tuvieran las mismas dimensiones 256×256 y se normalizaron cada una de ellas. También se empleó la técnica de regularización “aumento de datos” con el objetivo de disponer de un mayor número de imágenes para entrenar los modelos. Una vez entrenados cada uno de los modelos de CNN, aquellos que proporcionaron un mejor rendimiento en la clasificación de imágenes fueron ResNet50 y VGG-16, con valores de exactitud (*accuracy*) de 0.96 y 0.95 respectivamente. Dados estos resultados tan favorables, el presente proyecto tratará de evaluar el rendimiento de estos modelos pre-entrenados (arquitecturas ResNet50 y VGG16), además de otros, sobre el problema de clasificación de imágenes MRI para la detección de la enfermedad rara HSVE.

Sin embargo, se debe tener en cuenta que el rendimiento puede variar dependiendo de los conjuntos de datos utilizados, el problema de clasificación planteado sobre una enfermedad dada, la arquitectura de la red neuronal, los métodos de pre-procesamiento y otros factores. Además, se deberán concebir los sistemas basados en visión por computador como herramientas de apoyo para el diagnóstico realizado por el experto médico, en ningún caso como un sistema para automatizar el diagnóstico del paciente.

Capítulo 3

Alcance y objetivos

3.1. Objetivo general

El principal objetivo del proyecto consiste en emplear técnicas de *Deep Learning* aplicadas a la clasificación de imágenes de resonancia magnética del cerebro para la detección de la enfermedad rara HSVE. La clasificación consiste en identificar si una imagen MRI es de un cerebro con HSVE o si, por el contrario, es de un cerebro sano. Para alcanzar este objetivo, a continuación se definen un conjunto de objetivos específicos o requisitos del proyecto.

3.2. Objetivos específicos

1. Adquirir conocimientos sobre diferentes técnicas de *Deep Learning* aplicadas a algoritmos de clasificación que empleen redes convolucionales y documentar de forma clara y objetiva este conocimiento mediante la presente memoria.
2. Revisar y documentar las investigaciones previas relacionadas con la aplicación de redes convolucionales a la clasificación de imágenes MRI médicas.
3. Implementar modelos de CNN capaces de clasificar imágenes MRI del cerebro humano según presente encefalitis herpética o no (cerebro sano).
4. Realizar diferentes experimentos conforme a dichos modelos implementados donde además se apliquen técnicas de *Deep Learning*. Se deberá realizar una

evaluación de cada modelo entrenado, obtener métricas que midan sus resultados y realizar comparativas con los resultados de clasificación de cada modelo.

5. Obtener conclusiones del trabajo realizado y proponer posibles líneas de trabajo futuro.

3.3. Limitaciones

Las limitaciones detectadas al realizar el proyecto fueron las siguientes:

1. Para realizar los experimentos se ha empleado una Unidad Gráfica de Procesamiento (GPU) DE 4GB de memoria cuyo modelo es “NVIDIA GeForce GTX 1050”. La disposición y configuración de esta GPU en el equipo de trabajo ha permitido acelerar el proceso de entrenamiento de los modelos de clasificación basados en redes convolucionales, ya que generalmente el tiempo de entrenamiento se reduce significativamente con las tarjetas gráficas frente al uso de la Unidad Central de Procesamiento (CPU). No obstante, al emplear grandes arquitecturas de redes convolucionales, el largo tiempo de entrenamiento de sus modelos sigue siendo un aspecto que limita el desarrollo de este proyecto.
2. En relación al conjunto de datos empleado, el estudio se encuentra limitado a un total de 745 imágenes MRI correspondientes a 10 casos de la enfermedad HSVE, diagnosticados y publicados en el repositorio público *Radiopaedia* (<https://radiopaedia.org/?lang=us>). Cabe destacar que el número de imágenes MRI disponibles que se han obtenido en cada caso es diferente, variando este número desde 1 imagen (mínimo) hasta 85 imágenes (máximo) por caso. En la sección 4.1 del siguiente capítulo se detallan todas las características relativas a los conjuntos de datos empleados.

Capítulo 4

Material y Método

4.1. Fuente de datos

El presente proyecto ha empleado un total de 745 imágenes de secuencias de Resonancia Magnética (MRI) del cerebro de diferentes pacientes. Estas imágenes se clasifican en dos grupos dependiendo del estado del cerebro del paciente en cuestión. Las imágenes de estos conjuntos presentan las siguientes características comunes:

1. Tienen una resolución muy variable, abarcando imágenes con la resolución mínima de 256x256x1 píxeles hasta imágenes con la mayor resolución de 1000x840x4 píxeles.
2. Son todas imágenes de resonancia magnética (MRI).
3. Presentan 1, 2 ó 4 canales. De este modo, pueden existir imágenes con un sólo canal (p.e.: 256x256x1) o imágenes con hasta 4 canales (p.e.: 1000x840x4). A pesar de esta variación en el número de canales, todas las imágenes emplean la escala de grises como escala de color, como es propio de las MRI.
4. Constituyen una secuencia de cortes o secciones capturadas a partir de los planos frontal o CORONAL, SAGITAL y AXIAL u horizontal.
5. Los archivos presentan alguno de los siguientes formatos de imagen: .png, .jpg o .jpeg. Estos formatos aplican compresión en la imagen, lo que se traduce en una reducción de la calidad y el tamaño del archivo.

En las siguientes sub-secciones, se describen las particularidades de cada uno de los conjuntos de imágenes utilizados. Cada uno de estos conjuntos de datos se corresponden con una clase diferente en el algoritmo de clasificación.

4.1.1. Conjunto de datos 1: MRI de cerebros con HSVE

Este conjunto de datos está constituido por un total de 366 imágenes de secuencias MRI de cerebros con la enfermedad rara HSVE. Estas imágenes son etiquetadas en el algoritmo de clasificación con la clase positiva '1', haciendo así referencia a que el cerebro ha sido afectado por la enfermedad rara HSVE.

Este conjunto de datos ha sido obtenido del repositorio *Radiopaedia* por medio de una búsqueda de casos por el nombre “HSV” (Herpes Simplex Virus). A partir de esta búsqueda, se ha aplicado un filtro que permita seleccionar aquellos casos que presenten imágenes MRI y cuyos diagnósticos habían finalizado (Radiopaedia, 2023). De esta búsqueda de casos, se han seleccionado 10 casos de HSVE. Dependiendo de la disponibilidad de imágenes en cada caso, se ha obtenido un menor o mayor número de imágenes pertenecientes a alguno de los siguientes tipos de secuencias MRI: FLAIR, T1, T2, T1C+ o T2C+ (Preston, 2016). Además, las secuencias MRI pueden haber sido capturadas desde alguno de los siguientes tres planos: CORONAL, SAGITAL o AXIAL. A continuación, se describen cada uno de los 10 casos obtenidos.

1. **Caso 1:** compuesto por un total de 67 imágenes. De este total, 12 imágenes se corresponden con secuencias FLAIR capturadas desde el plano AXIAL, 13 con secuencias T2 desde el plano AXIAL, 23 con secuencias FLAIR desde el plano CORONAL y 19 con secuencias T1 del plano SAGITAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro de una mujer de 70 años que fue encontrada en el suelo de su casa confundida, con fiebre y con un bajo nivel de conciencia. Los picos de fiebre de este paciente y la punción lumbar positiva confirmaron la encefalitis por HSV-1 (virus del herpes simple de tipo 1).

2. **Caso 2:** compuesto por una única imagen MRI con el formato T2 capturadas

desde el plano CORONAL. Concretamente, esta imagen MRI es una visualización del cerebro de una mujer de 35 años que presentaba fiebre y se encontraba agitada y confundida. Tras realizarle una biopsia cerebral, la histología confirmó el caso de encefalitis por HSV-1.

3. **Caso 3:** compuesto por una única imagen MRI con el formato FLAIR capturada desde el plano AXIAL. Concretamente, esta imagen MRI es una visualización del cerebro de un bebé de 20 meses que presentaba fiebre y convulsiones. Tras realizarle un examen, se confirmó el caso de encefalitis por HSV-1.
4. **Caso 4:** compuesto por un total de 42 imágenes. De este total, 13 imágenes se corresponden con secuencias FLAIR capturadas desde el plano AXIAL, 12 con secuencias T1 desde el plano AXIAL y 17 con secuencias T2 del plano SAGITAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro de una mujer de 90 años que presentaba fiebre y un nivel de conciencia alterado. En el examen, se confirmó el caso de encefalitis por HSV-1. En este caso, el virus apareció como afectación unilateral del lóbulo temporal medial y el sistema límbico.
5. **Caso 5:** compuesto por un total de 55 imágenes. De este total, 18 imágenes se corresponden con secuencias FLAIR capturadas desde el plano AXIAL, 18 con secuencias T2 desde el plano AXIAL y 19 con secuencias FLAIR del plano SAGITAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro de una mujer de 50 años que tuvo constantes dolores de cabeza, fiebre y cambios en el estado mental (confusión y alucinaciones). El examen que se le realizó confirmó la encefalitis por HSV-1. Las imágenes de resonancia magnética cerebral muestran efectos sobre el lóbulo temporal bilateral y sobre los lóbulos frontales.
6. **Caso 6:** compuesto por un total de 25 imágenes. De este total, 12 imágenes se corresponden con secuencias FLAIR capturadas desde el plano AXIAL y 13 con secuencias T2 desde el plano AXIAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro de un hombre de 65 años que tuvo dos

días de dolor de cabeza seguidos de convulsiones epilépticas. Tras la realización del correspondiente examen, se confirmó el caso de encefalitis por HSV-1 en el paciente.

7. **Caso 7:** compuesto por un total de 4 imágenes. De este total, 2 imágenes se corresponden con secuencias FLAIR capturadas desde el plano AXIAL, mientras que las otras 2 se corresponden con secuencias FLAIR del plano CORONAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro de un hombre de 60 años que se presentó en urgencias tras una semana de fiebre leve, malestar general, confusión y múltiples caídas. Tras examinar al paciente, se confirmó el caso de encefalitis por HSV-2. Las imágenes MRI mostraron efectos sobre el lóbulo temporal derecho, así como sobre el lóbulo parieto-occipital izquierdo. Dado que éste es el único caso de encefalitis por HSV-2 empleado en el presente trabajo, estas imágenes serán utilizadas en el conjunto de entrenamiento para ayudar al modelo a generalizar ante cualquier futuro caso de encefalitis herpética que se le presente, ya sea de tipo 1 o de tipo 2.
8. **Caso 8:** compuesto por un total de 41 imágenes. De este total, 12 imágenes se corresponden con secuencias FLAIR capturadas desde el plano AXIAL, otras 12 se corresponden con secuencias T2 del plano AXIAL y 17 se corresponden con secuencias FLAIR del plano CORONAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro de un hombre de 55 años que presentó convulsiones, agitación, fiebre y disminución del nivel de conciencia durante 4 días. Tras examinar al paciente, se confirmó el caso de encefalitis por HSV-1.
9. **Caso 9:** compuesto por un total de 78 imágenes. De este total, 25 imágenes se corresponden con secuencias T1 capturadas desde el plano AXIAL, 12 se corresponden con secuencias T2 desde el plano AXIAL, 24 con secuencias FLAIR desde el plano CORONAL y 17 se corresponden con secuencias T1 desde el plano SAGITAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro de un hombre de 50 años que tuvo dolores de cabeza y alteración del comportamiento durante cinco días. Tras examinar al paciente, se confirmó el caso de encefalitis por HSV-1.

10. Caso 10: compuesto por un total de 52 imágenes. De este total, 11 imágenes se corresponden con secuencias FLAIR capturadas desde el plano AXIAL, 12 se corresponden con secuencias T1 desde el plano AXIAL, otras 12 con secuencias T2 desde el plano AXIAL y 17 se corresponden con secuencias T2 desde el plano SAGITAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro de una mujer de 90 años que presentó fiebre leve y alteración del nivel de consciencia. Tras examinar al paciente, se confirmó el caso de encefalitis por HSV-1.

	Nº IMÁGENES MRI CEREBROS HSVE											
CASO	1	2	3	4	5	6	7	8	9	10	TOTAL	PORCENTAJE
AXIAL	25	0	1	25	36	25	2	24	37	35	210	58 %
CORONAL	23	1	0	0	0	0	2	17	24	0	67	18 %
SAGITAL	19	0	0	17	19	0	0	0	17	17	89	24 %
TOTAL	67	1	1	42	55	25	4	41	78	52	366	100 %

Tabla 4.1: Distribución de imágenes MRI de cerebros con HSVE del conjunto de datos 1

En la tabla 4.1, se ilustra la distribución de imágenes MRI de cerebros con HSVE para cada uno de los 10 casos utilizados en este conjunto de datos. Además, se identifica el tipo de secuencia de las imágenes MRI según el plano empleado en su captura (AXIAL, CORONAL o SAGITAL). Cabe destacar que todos los casos de encefalitis son del tipo HSV-1, a excepción del caso 7 que es del tipo HSV-2. Tal y como se puede apreciar en la tabla, alrededor del 58% de imágenes MRI de cerebros con HSVE se corresponden con secuencias MRI de tipo AXIAL, el 18% con secuencias de tipo CORONAL y el 24% con secuencias de tipo SAGITAL. De este modo, resulta de gran importancia destacar que la mayor parte de las imágenes se corresponden con secuencias MRI de tipo AXIAL.

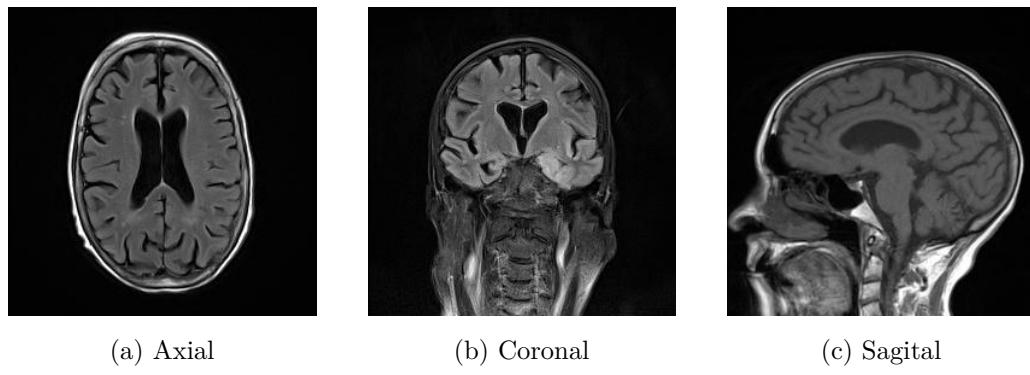


Figura 4.1: Imágenes MRI del cerebro con HSVE del caso 1 desde los planos AXIAL, CORONAL y SAGITAL. Fuente: *Radiopaedia* (Radiopaedia, 2023)

En la figura 4.1 se visualizan imágenes MRI desde los planos AXIAL, CORONAL y SAGITAL del cerebro con encefalitis herpética con el virus del tipo HSV-1 correspondiente al caso 1, cuyo paciente se encontraba confundida, con fiebre y con un bajo nivel de conciencia.

4.1.2. Conjunto de datos 2: MRI de cerebros sanos

Este conjunto de datos está constituido por un total de 379 imágenes de secuencias MRI de cerebros sanos o en estado normal de pacientes con dolor de cabeza. Estas imágenes son etiquetadas en el algoritmo de clasificación con la clase negativa '0', haciendo así referencia a que el cerebro no presenta la enfermedad rara HSVE. Este conjunto de datos ha sido obtenido a través del listado de casos de cerebros normales, elaborado y recopilado por Frank Gaillard en el repositorio público: *Radiopaedia* (Gaillard, 2023). De este listado de casos de cerebros sanos o en estado normal se han seleccionado 7 casos, cuya correspondencia ha sido especificada en el apéndice A. Dependiendo de la disponibilidad de imágenes en cada caso, se ha obtenido un menor o mayor número de imágenes pertenecientes a alguno de los siguientes tipos de secuencias MRI: T1, T2 y FLAIR (Preston, 2016). Además, las secuencias MRI pueden haber sido capturadas mediante alguno de los tres siguientes planos: CORONAL, SAGITAL o AXIAL. A continuación, se describen cada uno de los 7 casos obtenidos.

1. **Caso 1:** compuesto por un total de 67 imágenes. De este total, 12 imágenes se corresponden con secuencias FLAIR capturadas desde el plano AXIAL, otras 12 con secuencias T1 desde el plano AXIAL, 26 con secuencias T2 del plano AXIAL y 17 con secuencias T1 del plano SAGITAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro con apariencia normal de una mujer joven de 20 años que tenía dolores de cabeza leves.
2. **Caso 2:** compuesto por un total de 62 imágenes. De este total, 19 imágenes se corresponden con secuencias FLAIR capturadas desde el plano AXIAL, 22 con secuencias T2 del plano AXIAL y 21 con secuencias T1 del plano SAGITAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro con apariencia normal de una mujer de 30 años que tenía dolores de cabeza leves.
3. **Caso 3:** compuesto por un total de 31 imágenes. De este total, 15 imágenes se corresponden con secuencias FLAIR capturadas desde el plano AXIAL y 16

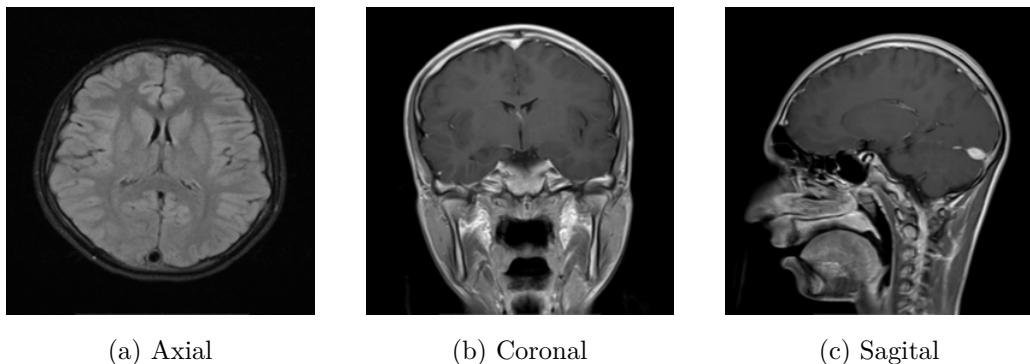
con secuencias T2 desde el plano AXIAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro de una joven adolescente de 13 años con dolor de cabeza constante durante 4 semanas. Presentaba dolor de cabeza en forma de migrañas, aunque su sistema cerebral presenta una apariencia normal.

4. **Caso 4:** compuesto por un total de 83 imágenes. De este total, 9 imágenes se corresponden con secuencias FLAIR capturadas desde el plano AXIAL, 19 con secuencias T1 desde el plano AXIAL, 11 con secuencias T2 del plano AXIAL, 18 con secuencias T1 del plano CORONAL y 26 con secuencias T1 del plano SAGITAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro con apariencia normal de un adolescente de 15 años que tenía dolor de cabeza.
5. **Caso 5:** compuesto por un total de 19 imágenes. De este total, 9 imágenes se corresponden con secuencias FLAIR capturadas desde el plano AXIAL y las 10 restantes con secuencias T2 del plano AXIAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro con apariencia normal de una mujer de 22 años que tenía dolor de cabeza.
6. **Caso 6:** compuesto por un total de 85 imágenes. De este total, 13 imágenes se corresponden con secuencias FLAIR capturadas desde el plano AXIAL, 19 con secuencias T1 desde el plano AXIAL, 17 con secuencias T2 del plano AXIAL, 15 con secuencias T1 del plano CORONAL y 21 con secuencias T1 del plano SAGITAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro con apariencia normal de una persona de 55 años.
7. **Caso 7:** compuesto por un total de 32 imágenes. De este total, 16 imágenes se corresponden con secuencias T2 del plano AXIAL y otras 16 con secuencias FLAIR del plano CORONAL. Concretamente, estas imágenes MRI son visualizaciones del cerebro con apariencia normal de una mujer de 45 años que tenía dolores de cabeza, pero no presentaba lesiones u otras patologías.

CASO	Nº IMÁGENES MRI CEREBROS SANOS							TOTAL	PORCENTAJE
	1	2	3	4	5	6	7		
AXIAL	50	41	31	39	19	49	16	245	65 %
CORONAL	0	0	0	18	0	15	16	49	13 %
SAGITAL	17	21	0	26	0	21	0	85	22 %
TOTAL	67	62	31	83	19	85	32	379	100 %

Tabla 4.2: Distribución de imágenes MRI de cerebros sanos del conjunto de datos 2

En la tabla 4.2, se ilustra la distribución de imágenes MRI de cerebros sanos para cada uno de los casos utilizados en este conjunto de datos. Además, se identifica el tipo de secuencia de las imágenes MRI según el plano empleado en su captura (AXIAL, CORONAL o SAGITAL). Tal y como se puede apreciar, el 65 % de imágenes MRI de cerebros sanos se corresponden con secuencias MRI de tipo AXIAL, el 13 % con secuencias de tipo CORONAL y el 22 % con secuencias de tipo SAGITAL. De este modo, la mayor parte de las imágenes se corresponden con secuencias MRI de tipo AXIAL, al igual que ocurre en el conjunto de datos 1.

Figura 4.2: Imágenes MRI del cerebro sano del caso 4 desde los planos AXIAL, CORONAL y SAGITAL. Fuente: *Radiopaedia* (Gaillard, 2023)

En la figura 4.2 se visualizan imágenes MRI desde los planos AXIAL, CORONAL y SAGITAL del cerebro sano correspondiente al caso 4, cuyo paciente presentaba dolor de cabeza constante durante 4 semanas, además de migrañas.

4.2. Herramientas Hardware utilizadas

La infraestructura hardware que se ha empleado para la realización de este trabajo está formada por un ordenador portátil con las siguientes características:

1. Unidad de Procesamiento Gráfico (GPU): NVIDIA GeForce GTX 1050.
2. Unidad Central de Procesamiento (CPU): Intel(R) Core(TM) i7-7700HQ con una frecuencia de reloj de 2.80GHz. Dispone de 4 núcleos (procesadores principales) y 8 hilos (procesadores lógicos).
3. Memoria principal (RAM) con un almacenamiento de 8,0 GB.
4. Sistema Operativo (SO): Microsoft Windows 10 Home

4.3. Herramientas Software utilizadas

Para implementar los modelos de clasificación, se ha instalado la distribución Anaconda. Esta distribución de código abierto ha permitido la creación y gestión de un entorno virtual con Python instalado. Además de utilizar Python como lenguaje de programación, la principal librería de *Deep Learning* que se ha instalado en el entorno para implementar los modelos de clasificación ha sido TensorFlow. Concretamente, se ha utilizado Keras, una librería integrada en TensorFlow que facilita la implementación de modelos de redes neuronales artificiales por medio de una API (Interfaz de Programación de Aplicaciones) disponible a través de su documentación oficial (Chollet, 2015). En la tabla 4.3 se especifican las versiones utilizadas tanto de Python como de cada una de las librerías necesarias que se han utilizado para el desarrollo del trabajo.

Nombre	Tipo de paquete/librería	Versión
python	Lenguaje de Programación	3.9.16
tensorflow	Librería de <i>Machine Learning y Deep Learning</i>	2.10.0
keras	Librería de <i>Deep Learning</i>	2.10.0
pandas	Librería de estructuras de datos y análisis de datos	1.5.3
numpy	Librería de Almacenamiento y Procesamiento de datos	1.24.3
scikit-learn	Librería de <i>Machine Learning y Data Mining</i>	1.2.2
skimage	Librería de Procesamiento de Imágenes	0.20.0
matplotlib	Librería de Visualización de datos	3.7.1
seaborn	Librería de Visualización de datos estadísticos	0.12.2

Tabla 4.3: Especificación de versiones de los paquetes software instalados

La disposición de la unidad de procesamiento gráfico (GPU) especificada en la sección 4.2, así como su correcta configuración en el entorno virtual de Anaconda, ha permitido acelerar los tiempos de entrenamiento de cada uno de los modelos de clasificación construidos en el presente proyecto. En los siguientes tres capítulos, se mostrará, además de otras métricas, el tiempo que han tardado los principales modelos de CNN en completar su entrenamiento.

4.4. Metodología

La metodología empleada para el desarrollo del presente trabajo consta de las siguientes partes:

En primer lugar, se adquiere y se prepara la muestra de datos, la cual estará compuesta por un conjunto de imágenes MRI correspondientes a 2 clases diferentes: secuencias MRI de cerebros enfermos con encefalitis herpética y secuencias MRI de cerebros sanos, estas últimas correspondientes a pacientes que presentaron dolores de cabeza. Una vez los datos han sido obtenidos y pre-procesados, se procede a

aplicar un proceso cíclico compuesto por las siguientes fases:

1. Ideación y diseño del experimento.

En esta primera fase del proceso cíclico se realiza la construcción del modelo del experimento. Para ello, se selecciona una determinada arquitectura de red neuronal convolucional o se crea una desde cero. Posteriormente, se inicializan una serie de hiper-parámetros del modelo de clasificación.

2. Entrenamiento y evaluación.

Una vez se ha construido el modelo asociado al experimento, se procede a entrenarlo con los datos de entrenamiento y de validación. El modelo se encargará de modificar los parámetros de la red a medida que va detectando patrones en los datos de entrenamiento. Además, los datos del conjunto de validación permitirán comprobar cómo de bien está entrenando el algoritmo y si es capaz de clasificar correctamente las imágenes del conjunto de validación. Una vez entrenado el modelo, éste se evalúa prediciendo las clases de las imágenes pertenecientes al conjunto de pruebas y se obtienen las métricas de clasificación.

3. Análisis de los resultados.

Después de haber evaluado el modelo y haber obtenido los resultados proporcionados por las métricas de clasificación, se procede a la elaboración de gráficas que permitan analizar la capacidad predictiva del modelo, así como visualizar la evolución de los valores de las métricas a lo largo de su entrenamiento. Esta última fase de análisis es determinante para poder conocer qué cambios se deberán realizar sobre la arquitectura, sobre los datos o sobre los hiperparámetros del modelo en las siguientes iteraciones de experimentos con el objetivo de mejorar los resultados obtenidos en las métricas de clasificación.

Este proceso cíclico resulta bastante recurrente en los proyectos de *Deep Learning*, principalmente en aquellos en los que se trabaja con redes neuronales artificiales (ANN). En la figura 4.3 se ilustra este proceso cíclico en forma de diagrama.



Figura 4.3: Proceso cílico en proyectos de Deep Learning

Partiendo de este proceso cílico, resulta importante aclarar que para el desarrollo de este proyecto se ha seguido una metodología incremental. Esto se debe a que cada modelo trata de mejorar los resultados con respecto a los obtenidos en el modelo anterior. De este modo, después de realizar cada uno de los experimentos, se seleccionará aquel que mejores resultados haya proporcionado.

Antes de proceder a describir las particularidades de cada modelo de clasificación implementado, todos ellos comparten los siguientes pasos o fases:

1. **Lectura de imágenes del conjunto de datos:** se leen las imágenes MRI del conjunto de datos en cuestión desde el directorio correspondiente y se cargan en memoria principal (RAM). Este directorio contendrá a su vez tres subdirectorios correspondientes a los conjuntos de entrenamiento, de validación y de pruebas.
2. **Definición, compilación y entrenamiento del modelo de CNN:** se definen cada una de las capas por las que estará formada la arquitectura de la red neuronal convolucional. En caso de emplear *Transfer Learning*, se describirá el modelo base empleado. Por otro lado, cada modelo de CNN ha empleado el optimizador *Adam* con una tasa de aprendizaje o *learning rate* de 0.0001. Además, se han definido 2 callbacks que permitan detener el entrenamiento de cada modelo si el valor de pérdida o de exactitud en el conjunto de validación

no mejora en 10 *epochs* consecutivas. Una vez construido y compilado cada modelo, se procede a su entrenamiento a lo largo de un máximo de 30 épocas (*epochs*), asignando el tamaño de cada *mini-batch* (*batch size*), que variará entre 8 y 16 muestras. Los valores empleados para el número máximo de *epochs*, así como para el tamaño del *mini-batch* en estos modelos son dependientes de las características de hardware que dispone el computador utilizado para la realización del presente proyecto. Sobre un computador con mayores prestaciones se podría probar a incrementar el tamaño del *batch* de 8 a 16 ejemplos, o incluso hasta 32 o 64.

3. **Evaluación del modelo y análisis de los resultados:** se evalúa el modelo mediante las imágenes del conjunto de pruebas. Una vez hecho esto, se analizan los resultados del modelo de clasificación utilizando varias gráficas entre las que destacan la matriz de confusión y la evolución de métricas como la exactitud o la pérdida del modelo a lo largo del entrenamiento. También se evalúa el modelo atendiendo a la exhaustividad como métrica de clasificación.
4. **Exportación del modelo:** se exporta el modelo, así como los resultados y gráficas generadas a partir de los resultados obtenidos. La exportación de cada modelo permite el uso del modelo en el futuro, y así evitar tener que entrenar el modelo de red neuronal convolucional de nuevo.



Figura 4.4: Fases del proceso seguido durante el desarrollo de los experimentos de Red Neuronal Convolutacional

En la figura 4.4 se ilustran las fases más relevantes del proceso de desarrollo de cada uno de los experimentos realizados en el trabajo.

4.4.1. Preprocesamiento de las imágenes

Tal y como se comentó en las secciones 4.1.1 y 4.1.2, las imágenes MRI empleadas presentan dimensiones diferentes. Dado que la capa de entrada de un modelo de red neuronal implementado en keras requiere que cada una de las imágenes presenten las mismas dimensiones, se ha aplicado un fase de preprocesamiento que permita redimensionar cada una de las imágenes para que presenten la forma 256x256x1, es decir, una resolución de 256x256 y un único canal de color (escala de grises). Además, se han normalizado cada una de las imágenes del conjunto de datos para que los valores de intensidad de sus píxeles se encuentren en el rango de valores [0, 1]. Es importante normalizar las imágenes para reducir la variabilidad en los valores de intensidad de los datos y para mejorar el rendimiento de la red neuronal convolucional, ya que las funciones de activación suelen comportarse de manera más estable ante datos normalizados.

En los siguientes tres capítulos, se describen cada uno de los experimentos y modelos de CNN construidos con el objetivo de clasificar los 2 conjuntos de imágenes MRI en sus 2 respectivas clases. Cabe destacar que cada modelo consiste en la modificación o adición de algún nuevo componente sobre el modelo anterior, de tal modo que las características y peculiaridades de cada modelo serán descritas a partir del siguiente capítulo.

Capítulo 5

Iteración 1 de experimentos

En este capítulo se describen los modelos de clasificación implementados en esta primera iteración. Además, se discuten los resultados obtenidos tanto en el primer modelo implementado (modelo base), como en el segundo modelo, que es el que mejores resultados ha proporcionado en esta iteración.

5.1. Modelos implementados

En esta primera iteración de experimentos se ha partido del conjunto inicial de imágenes MRI descrito en las secciones 4.1.1 y 4.1.2. De este modo, se dispone de un total de 745 imágenes de secuencias MRI. De este total, 366 imágenes se corresponden a secuencias MRI de cerebros con la enfermedad HSVE (clase 1), mientras que las 379 restantes se corresponden a secuencias MRI de cerebros sanos (clase 0). Estas imágenes han sido divididas en los 3 conjuntos de entrenamiento, validación y pruebas, tal y como se ilustra en las tablas 5.1 y 5.2.

	Nº IMÁGENES MRI CEREBROS CON HSVE										
	ENTRENAMIENTO						VALIDACIÓN		PRUEBAS		
CASO	2	3	5	7	8	9	4	6	1	10	TOTAL
AXIAL	0	1	36	2	24	37	25	25	25	35	210
CORONAL	1	0	0	2	17	24	0	0	23	0	67
SAGITAL	0	0	19	0	0	17	17	0	19	17	89
TOTAL	180						67		119		366
PORCENTAJE	49 %						18 %		33 %		100 %

Tabla 5.1: Distribución de imágenes MRI de cerebros con HSVE en los 3 conjuntos

	Nº IMÁGENES MRI CEREBROS SANOS								
	ENTRENAMIENTO				VALIDACIÓN		PRUEBAS		
CASO	1	3	4	2	7	5	6	TOTAL	
AXIAL	50	31	39	41	16	19	49	245	
CORONAL	0	0	18	0	16	0	15	49	
SAGITAL	17	0	26	21	0	0	21	85	
TOTAL	181				94		104		379
PORCENTAJE	48 %				25 %		27 %		100 %

Tabla 5.2: Distribución de imágenes MRI de cerebros sanos en los 3 conjuntos

En problemas de *Machine Learning* normalmente se destina alrededor del 15 % o 20 % del total de imágenes para el conjunto de validación y otro 15 % o 20 % para el conjunto de pruebas. Sin embargo, dado que el conjunto de imágenes utilizado para el presente trabajo está formado por un número reducido de casos, se ha decidido destinar al menos 2 casos tanto al conjunto de validación como al conjunto de pruebas. De este modo, a pesar de emplear en estos conjuntos un elevado número de imágenes respecto al total (alrededor del 25 %), se logra que estos conjuntos dispongan de una distribución representativa de imágenes con todos o casi todos los formatos, al igual que las imágenes del conjunto de entrenamiento. Además, se ha aplicado una fase de submuestreo con el objetivo de equilibrar el número de imáge-

nes existentes en de cada clase a lo largo de sus conjuntos. De este modo, cada clase empleará alrededor de 180 imágenes en su respectivo conjunto de entrenamiento, alrededor de 67 imágenes en su conjunto de validación y alrededor de 104 imágenes en su conjunto de pruebas. A partir de esta distribución en los datos, se han implementado 6 modelos que emplean arquitecturas de CNN diferentes, con el objetivo de comprobar cuál de ellas se adapta y generaliza mejor sobre los datos y, por tanto, proporciona valores más altos en las métricas de clasificación.

Resulta conveniente aclarar que, para el entrenamiento de todos los modelos de CNN implementados en el presente trabajo, se han empleado 2 *callbacks* de *Early Stopping*. Estos *callbacks* permiten detener el entrenamiento de cada modelo si la exactitud o la pérdida sobre el conjunto de validación no mejora en 10 *epochs* consecutivas.

5.1.1. Modelo 1 - Arquitectura de CNN construida desde cero

La propiedad que caracteriza este primer modelo es el empleo de una arquitectura de red neuronal convolucional construida desde cero, además de utilizar para su entrenamiento un tamaño del *mini-batch* de 16 ejemplos. Este hiperparámetro se ha definido tras haber probado con valores superiores para el *batch-size* tales como 32 ó 64 ejemplos. Al no poder soportar la unidad gráfica del computador sobre el que se han entrenado los modelos todos estos ejemplos a la vez, se detenía automáticamente su entrenamiento en la primera o segunda *epoch*. Por este motivo, y para que la GPU pudiese procesar los ejemplos correctamente durante el entrenamiento del modelo, se ha empleado un tamaño de *batch* de 16 ejemplos, y de 8 ejemplos para los modelos posteriores. La arquitectura empleada para este modelo está formada por las siguientes 2 partes bien diferenciadas:

1. La parte de aprendizaje o extracción de características espaciales de la imagen está compuesta por la capa de entrada de la red y por 2 bloques de la forma: capa convolucional seguida de una capa de *MaxPooling*. Para ambos tipos de

capas se ha definido un tamaño del filtro o kernel de 3x3. Además, cada capa convolucional emplea la función de activación 'relu', mientras que las capas de *MaxPooling* hacen uso de un valor 3 sobre el parámetro *stride*.

2. La parte correspondiente al clasificador está formada, en primer lugar, por una capa que permite vectorizar o unir en un vector unidimensional los datos bidimensionales generados por los filtros convolucionales y de *pooling* de la parte anterior de aprendizaje. En segundo lugar, se añaden 2 capas completamente conectadas de 256 y 512 neuronas respectivamente, con funciones de activación 'relu'. Finalmente, se añade una capa de salida, también completamente conectada, de 1 sola neurona y función de activación 'sigmoid', al tratarse de un problema de clasificación binaria.

La representación de dicha arquitectura, la cual ha sido implementada en keras, es especificada en el apéndice B.

5.1.2. Modelo 2 - Transfer Learning con Resnet50

En este segundo modelo se ha modificado la arquitectura de red convolucional para aplicar la técnica *Transfer Learning*. Concretamente, se ha empleado el modelo pre-entrenado 'ResNet50' de la librería 'keras' como modelo base, desactivando todos sus parámetros, de tal modo que no serán modificados durante el nuevo entrenamiento. Sobre este modelo base, se ha añadido la parte correspondiente al clasificador del modelo anterior, es decir, una capa que une la salida del bloque de aprendizaje en un vector unidimensional, 2 capas completamente conectadas y la capa de salida. Dado que la arquitectura de este modelo presenta un mayor tamaño (mayor número de parámetros) respecto a la anterior, en gran parte debido al uso de *Transfer Learning*, a partir de este modelo se utilizará un tamaño menor del *mini-batch*: 8 ejemplos.

La representación de esta segunda arquitectura es especificada en el apéndice B.

5.1.3. Modelo 3 - Transfer Learning con VGG16

En este tercer modelo se ha modificado exclusivamente el modelo pre-entrenado 'ResNet50', empleado como modelo base en el modelo anterior, por el modelo pre-

entrenado 'VGG16'. Del mismo modo, se han desactivado todos sus parámetros y se ha añadido la parte correspondiente al clasificador del anterior modelo.

La representación de esta tercera arquitectura es especificada en el apéndice B.

5.1.4. Modelo 4 - Transfer Learning con VGG19

En este cuarto modelo se ha vuelto a modificar exclusivamente el modelo pre-entrenado 'VGG16', empleado como modelo base en el modelo anterior, por el modelo pre-entrenado 'VGG19'.

La representación de esta cuarta arquitectura es especificada en el apéndice B.

5.1.5. Modelo 5 - Transfer Learning con InceptionV3

En este quinto modelo se ha vuelto a modificar exclusivamente el modelo pre-entrenado 'VGG19', empleado como modelo base en el modelo anterior, por el modelo pre-entrenado 'InceptionV3'.

La representación de esta quinta arquitectura es especificada en el apéndice B.

5.1.6. Modelo 6 - Transfer Learning con MobileNet

En este sexto modelo se ha vuelto a modificar exclusivamente el modelo pre-entrenado 'InceptionV3', empleado como modelo base en el modelo anterior, por el modelo pre-entrenado 'MobileNet'.

La representación de esta sexta arquitectura es especificada en el apéndice B.

5.2. Resultados obtenidos

En esta sección, se discuten los resultados obtenidos tanto del primer modelo implementado (modelo base), como del segundo modelo, que es el que mejores resultados de exactitud y exhaustividad ha proporcionado en esta primera iteración.

5.2.1. Modelo 1 - Arquitectura de CNN desde cero

En la figura 5.1 se aprecia la evolución de la exactitud y la pérdida de este modelo.

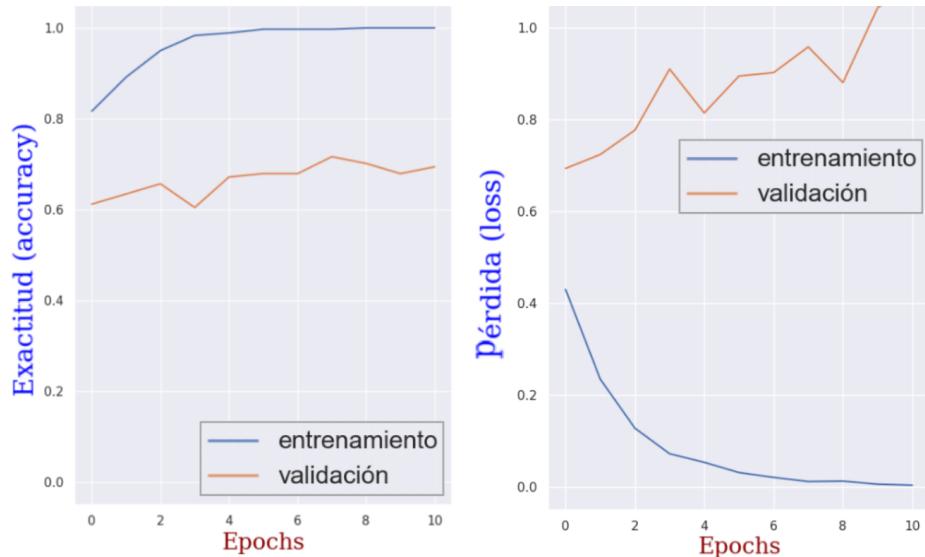


Figura 5.1: Evolución de la exactitud y la pérdida del primer modelo

En la tabla 5.3 se puede ver que, a pesar de que este primer modelo es entrenado en poco tiempo (1 minuto y 19 segundos), su exactitud obtenida sobre el conjunto de pruebas presenta un valor demasiado bajo (0.6233) en comparación con la exactitud obtenida en el entrenamiento (1.00). Esta situación supone un claro sobre-ajuste (*overfitting*) sobre el modelo. Este sobre-ajuste se intenta solventar en los modelos de la presente iteración y de las siguientes iteraciones empleando otras arquitecturas de CNN y mediante la aplicación de técnicas de regularización tales como *Dropout*, *Batch Normalization*, *Early Stopping* o aumento de datos. Tras analizar la exactitud obtenida en pruebas, no es posible determinar que el modelo clasifique correctamente las imágenes MRI en las dos clases empleadas.

	Exactitud (accuracy)			
	Entrenamiento	Validación	Pruebas	Tiempo entrenamiento
CNN construida desde cero	1.0000	0.6940	0.6233	1' 19"

Tabla 5.3: Evaluación del primer modelo mediante la exactitud durante entrenamiento, validación y pruebas

Por otro lado, en la tabla 5.4 se muestran algunas de las métricas de clasificación más importantes, tales como la precisión, la exhaustividad y el F1-Score, obtenidas tras la evaluación del primer modelo. Tal y como se describe en la sección 2.5, los valores de estas métricas de clasificación se pueden calcular conociendo, para cada clase, los verdaderos positivos (VP), los verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN), valores obtenidos tras evaluar el modelo. Considerando que los verdaderos positivos son las muestras correctamente clasificadas de la clase 'Cerebro con HSVE', en la tabla se observa cómo se obtiene una exhaustividad (recall) del 0.7983, siendo así capaz de clasificar correctamente el 79.83 % de las MRI de la clase 'cerebro con HSVE' del conjunto de pruebas, clasificando el 20.17 % restante de las MRI de cerebros con esta enfermedad como cerebros sanos. En los modelos posteriores se tratará de maximizar, además del valor de exactitud del modelo, el valor de exhaustividad obtenido.

Clase	Precisión	Exhaustividad	F1-Score	Nº imágenes test
Cerebro sano	0.6470	0.4231	0.5116	104
Cerebro con HSVE	0.6129	0.7983	0.6934	119

Tabla 5.4: Métricas de clasificación del primer modelo

Adicionalmente, la figura 5.2 presenta la matriz de confusión obtenida tras evaluar el modelo de clasificación inicial. De este modo, se indica sobre la diagonal decreciente de la matriz, la frecuencia relativa respecto a las imágenes de cada clase del conjunto de pruebas clasificadas correctamente, mientras que el resto de posiciones indican la frecuencia relativa respecto a las imágenes clasificadas incorrectamente.

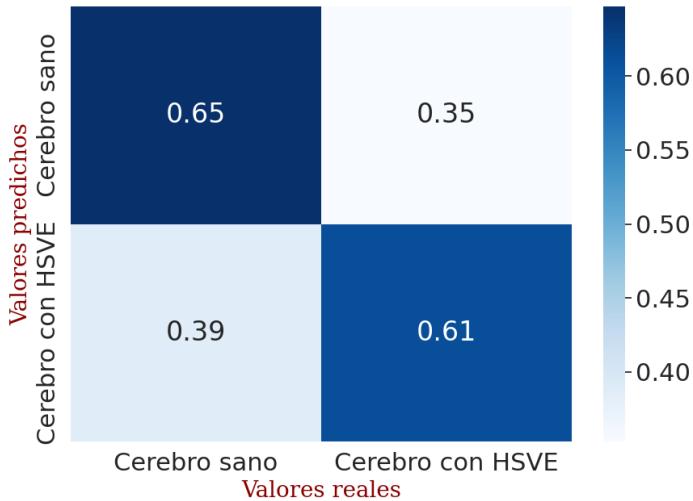


Figura 5.2: Matriz de confusión del primer modelo

5.2.2. Modelo 2 - Transfer Learning con Resnet50

En la figura 5.3 se aprecia la evolución de la exactitud y la pérdida de este segundo modelo.

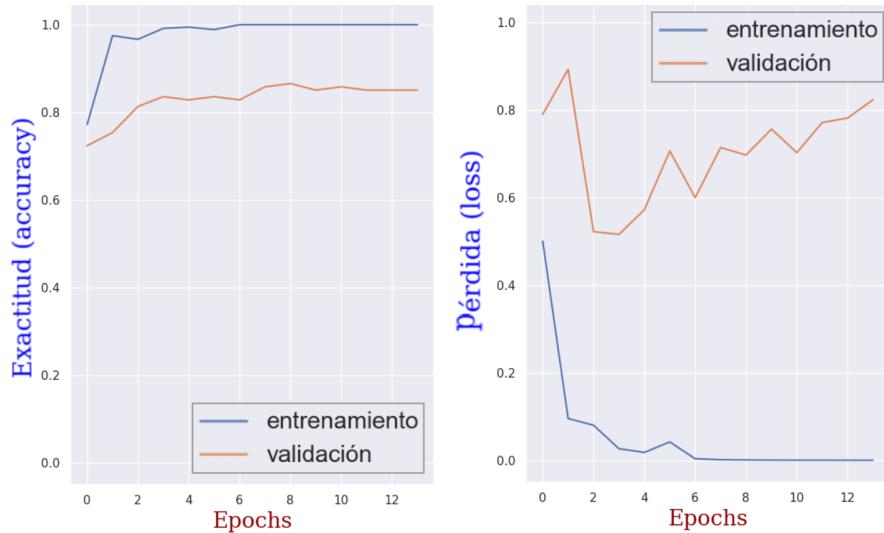


Figura 5.3: Evolución de la exactitud y la pérdida del segundo modelo

En la tabla 5.5 se puede ver que la exactitud obtenida sobre el conjunto de pruebas en el segundo modelo ha proporcionado el mayor valor con respecto al obtenido en el resto de modelos de esta primera iteración. Sin embargo, sigue existiendo un claro sobreajuste en el modelo, al haber obtenido un valor de exactitud más bajo en pruebas (0.7175) respecto al obtenido en validación (0.8507) y en el entrenamiento (1.0000). Este sobre-ajuste, causado probablemente por el uso de un número reducido de imágenes correspondientes a un total de 17 casos de ambas clases, se trata de solventar en los modelos implementados en las siguientes iteraciones mediante la aplicación de técnicas de regularización tales como *Dropout*, *Batch Normalization* o aumento de datos. Por otro lado, el tiempo de entrenamiento del modelo es de 28 minutos y 44 segundos.

Exactitud (accuracy)			
Entrenamiento	Validación	Pruebas	Tiempo entrenamiento
<i>Transfer Learning</i> con 'ResNet50'	1.0000	0.8507	0.7175

Tabla 5.5: Evaluación del segundo modelo mediante la exactitud durante entrenamiento, validación y pruebas

Por otro lado, en la tabla 5.6 se muestran algunas de las métricas de clasificación más importantes, tales como la precisión, la exhaustividad y el F1-Score, obtenidas tras la evaluación del segundo modelo. Se observa cómo el valor de exhaustividad ha bajado desde el valor 0.7983, obtenido en el primer modelo, hasta el valor 0.7479, obtenido con el presente modelo. Dado que el resto de métricas, como es el caso del F1-score, adquieren valores más altos que en el modelo inicial, se considera esta disminución en la exhaustividad como una situación normal, siempre y cuando este valor se estabilice en los próximos modelos y no siga disminuyendo. En las siguientes iteraciones de experimentos se tratará de mejorar el valor de esta métrica de clasificación, así como el de la exactitud, para seguir reduciendo el sobreajuste existente.

Clase	Precisión	Exhaustividad	F1-Score	Nº imágenes test
Cerebro sano	0.7029	0.6827	0.6927	104
Cerebro con HSVE	0.7295	0.7479	0.7386	119

Tabla 5.6: Métricas de clasificación del segundo modelo

Adicionalmente, la figura 5.4 presenta la matriz de confusión obtenida tras evaluar el segundo modelo de clasificación.

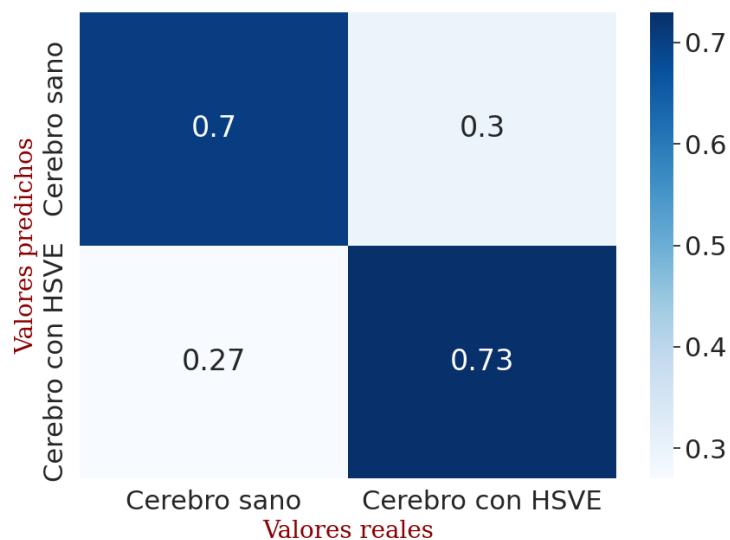


Figura 5.4: Matriz de confusión del segundo modelo

5.3. Resultados globales iteración 1

En la tabla 5.7 se muestra una comparativa de la exactitud y la exhaustividad obtenidas sobre cada uno de los modelos implementados en la iteración 1 de experimentos. Debido a que cada modelo ha empleado una arquitectura de red neuronal convolucional diferente, se puede apreciar el rendimiento obtenido con cada arquitectura. Cabe destacar que en estos 6 modelos de la primera iteración se ha empleado el mismo conjunto de datos para entrenar y evaluar los modelos, así como los mismos hiperparámetros, a excepción del tamaño del *batch*, al emplear el primer modelo un tamaño de 16 ejemplos y el resto de modelos un tamaño de 8. Tras analizar estos resultados globales, se puede apreciar cómo, a pesar de que el segundo modelo sigue presentando una situación de *over-fitting* o sobreajuste, este modelo es capaz de clasificar correctamente el mayor número de imágenes del conjunto de pruebas (el 71.75 % de dichas imágenes). Debido a que el segundo modelo es el único que presenta un valor superior a 0.70 tanto en la exactitud en pruebas como en la exhaustividad, se ha seleccionado este modelo para continuar con los experimentos de la siguiente iteración.

Modelo Implementado	Entrenamiento	Validación	Pruebas	Exhaustividad
Arquitectura CNN construida desde cero	1.0000	0.6940	0.6233	0.7983
<i>Transfer Learning</i> con 'ResNet50'	1.0000	0.8507	0.7175	0.7479
<i>Transfer Learning</i> con 'VGG16'	1.0000	0.8209	0.6547	0.7479
<i>Transfer Learning</i> con 'VGG19'	1.0000	0.7313	0.5336	0.6555
<i>Transfer Learning</i> con 'InceptionV3'	1.0000	0.6791	0.6143	0.5378
<i>Transfer Learning</i> con 'MobileNet'	1.0000	0.8507	0.6457	0.7815

Tabla 5.7: Comparativa de exactitud y exhaustividad en modelos de CNN implementados en la iteración 1. Resaltado en azul más intenso se muestran los modelos de CNN con valores de exactitud más altos y en azul menos intenso aquellos con valores más bajos.

Capítulo 6

Iteración 2 de experimentos

En este capítulo se describen los modelos de clasificación implementados en esta segunda iteración. Además, se discuten los resultados obtenidos sobre el séptimo modelo, siendo éste el que mejores resultados ha proporcionado de los dos nuevos experimentos realizados en esta iteración. Además, se comparan sus resultados a los obtenidos con el segundo modelo de la primera iteración, el cual sigue proporcionando los valores más altos en relación a las métricas de clasificación de exactitud y exhaustividad.

6.1. Modelos implementados

En esta segunda iteración de experimentos se ha seleccionado la arquitectura de CNN de la primera iteración de experimentos que ha superado el valor 0.70 tanto en la exhaustividad como en la exactitud sobre el conjunto de pruebas. De este modo, la arquitectura seleccionada ha sido la utilizada en el modelo 2, la cual emplea el modelo pre-entrenado 'ResNet50'.

Partiendo de esta arquitectura, en esta nueva iteración se ha evaluado su rendimiento sobre 3 subconjuntos diferentes del conjunto de datos original. Esto se debe a que la mayor parte de las imágenes MRI del conjunto de datos inicial fueron capturadas desde el plano AXIAL (alrededor del 60 %), por lo que existe la posibilidad de que el modelo obtenga un mayor rendimiento si se emplea exclusivamente este tipo

de imágenes y se descartan las imágenes MRI capturadas desde los planos CORONAL y SAGITAL. Además, dentro de las imágenes MRI capturadas desde el plano AXIAL existe una gran variedad de formatos, entre los que se encuentran el formato FLAIR, el T1, T2, T1C+ o T2C+. Dado que la mayor parte de las imágenes MRI del plano AXIAL son de alguno de los formatos FLAIR, T1 o T2, también se implementará un modelo en el que se evalúe su rendimiento con imágenes exclusivamente de estos formatos.

6.1.1. Modelo 2 - Transfer Learning con ResNet50 y MRI conjunto inicial

Se trata exactamente del modelo 2 de la primera iteración, en el que se emplea la técnica *Transfer Learning* mediante el modelo pre-entrenado 'ResNet50' sobre el conjunto de datos original, el cual es especificado en las tablas 5.1 y 5.2. Como se pudo observar en dichas tablas del capítulo 5, se han empleado imágenes MRI de cerebros desde los 3 planos AXIAL, SAGITAL y CORONAL. Este modelo se ha utilizado de referencia para comparar los resultados de los modelos de esta iteración, los cuales emplean diferentes subconjuntos de este conjunto de datos inicial.

6.1.2. Modelo 7 - Transfer Learning con ResNet50 y MRI subconjunto sólo plano AXIAL

En este nuevo modelo, se emplea el modelo pre-entrenado 'ResNet50' mediante *Transfer Learning*, al igual que en el segundo modelo. Sin embargo, en este caso se ha utilizado el subconjunto del conjunto de datos original correspondiente a las imágenes MRI capturadas sólo desde el plano AXIAL. Este subconjunto de datos empleado para este modelo, el cual está formado por un total de 455 imágenes, es especificado en las tablas 6.1 y 6.2.

	Nº IMÁGENES MRI CEREBROS CON HSVE (SÓLO AXIAL)										
	ENTRENAMIENTO					VALIDACIÓN		PRUEBAS			
	CASO	3	5	7	8	9	4	6	1	10	TOTAL
AXIAL	1	36	2	24	37	25	25	25	35	210	
TOTAL		100				50		60		210	
PORCENTAJE		48 %				24 %		28 %		100 %	

Tabla 6.1: Subconjunto de imágenes MRI de cerebros con HSVE sólo del plano AXIAL a lo largo de los 3 conjuntos

	Nº IMÁGENES MRI CEREBROS SANOS (SÓLO DESDE EL PLANO AXIAL)								
	ENTRENAMIENTO				VALIDACIÓN		PRUEBAS		
	CASO	1	3	4	2	7	5	6	TOTAL
AXIAL	50	31		39	41	16	19	49	245
TOTAL		120				57	68		245
PORCENTAJE		49 %				23 %	28 %		100 %

Tabla 6.2: Subconjunto de imágenes MRI de cerebros sanos sólo del plano AXIAL a lo largo de los 3 conjuntos

6.1.3. Modelo 8 - Transfer Learning con ResNet50 y MRI subconjunto formatos FLAIR, T1 y T2 sólo plano AXIAL

En este nuevo modelo, se emplea el modelo pre-entrenado 'ResNet50' mediante *Transfer Learning*, al igual que en el modelo anterior. Sin embargo, en este caso se ha utilizado el subconjunto del conjunto de datos original correspondiente a las imágenes MRI exclusivamente con los formatos FLAIR, T1 y T2 desde el plano AXIAL, descartando aquellas que presenten otros formatos tales como T1C+ o

T2C+, o que hayan sido capturadas desde los planos CORONAL y SAGITAL. Este subconjunto de datos empleado para este modelo, el cual está formado por un total de 399 imágenes, es especificado en las tablas 6.3 y 6.4.

	Nº IMÁGENES MRI CEREBROS CON HSVE (SÓLO AXIAL, FORMATOS FLAIR, T1 y T2)									
	ENTRENAMIENTO					VALIDACIÓN			PRUEBAS	
CASO	3	5	7	9	12	4	6	1	10	TOTAL
AXIAL	1	34	2	23	35	26	25	25	23	194
FLAIR	1	17	2	11	12	12	13	13	0	81
T1	0	0	0	0	11	14	0	0	12	37
T2	0	17	0	12	12	0	12	12	11	76
TOTAL	95					51			48	
PORCENTAJE	49 %					26 %			25 %	
									100 %	

Tabla 6.3: Subconjunto de imágenes MRI de cerebros con HSVE sólo de los formatos FLAIR, T1 y T2 del plano AXIAL a lo largo de los 3 conjuntos

	Nº IMÁGENES MRI CEREBROS SANOS (SÓLO AXIAL, FORMATOS FLAIR, T1 y T2)							
	ENTRENAMIENTO				VALIDACIÓN			PRUEBAS
CASO	1	3	4	2	7	5	6	TOTAL
AXIAL	38	31	30	41	16	19	30	205
FLAIR	12	15	9	19	0	9	13	77
T1	12	0	10	0	0	0	0	22
T2	14	16	11	22	16	10	17	106
TOTAL	99				57			49
PORCENTAJE	48 %				28 %			100 %

Tabla 6.4: Subconjunto de imágenes MRI de cerebros sanos sólo de los formatos FLAIR, T1 y T2 del plano AXIAL a lo largo de los 3 conjuntos

6.2. Resultados obtenidos

En esta sección se discuten los resultados obtenidos sobre el séptimo modelo, siendo éste el que mejores resultados ha proporcionado de los dos nuevos experimentos realizados en esta iteración. Además, se comparan sus resultados con los obtenidos mediante el segundo modelo de la primera iteración, el cual sigue manteniendo los valores más altos en las métricas de clasificación, destacando los valores obtenidos sobre la exactitud (0.7175) y la exhaustividad (0.7479).

6.2.1. Modelo 2 - Transfer Learning ResNet50 y conjunto inicial

Al tratarse del modelo 2, los resultados son expuestos en la sección 5.2.2.

6.2.2. Modelo 7 - Modelo 2 + MRI subconjunto sólo plano AXIAL

En la figura 6.1 se aprecia la evolución de la exactitud y la pérdida de este modelo.

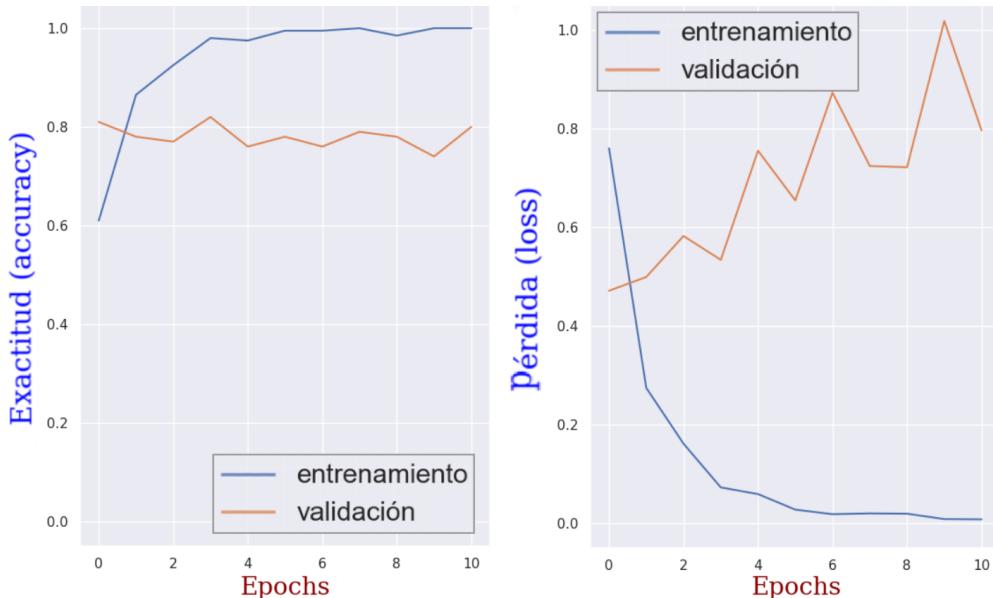


Figura 6.1: Evolución de la exactitud y la pérdida del séptimo modelo

En la tabla 6.5 se puede ver que en este séptimo modelo se obtiene un valor de exactitud inferior con respecto a la obtenida en el segundo modelo, tanto sobre el conjunto de validación (0.80 en el séptimo modelo y 0.85 en el segundo), como sobre el conjunto de pruebas (0.575 en el séptimo modelo y 0.7175 en el segundo). De este modo, al clasificar correctamente este séptimo modelo un menor porcentaje de imágenes sobre ambos conjuntos, se determina que el uso de imágenes MRI capturadas sólo desde el plano AXIAL no mejora los resultados en las métricas de clasificación con respecto a las obtenidas con el modelo 2, donde se emplea el conjunto de datos original compuesto por MRI capturadas desde los 3 planos. Adicionalmente, se especifica que el tiempo de entrenamiento obtenido al entrenar este séptimo modelo es de 10 minutos.

	Exactitud (accuracy)			Tiempo entrenamiento
	Entrenamiento	Validación	Pruebas	
<i>Modelo 7: Modelo 2 + sólo MRI AXIAL</i>	1.0000	0.8000	0.575	10' 00"

Tabla 6.5: Evaluación del séptimo modelo mediante la exactitud durante entrenamiento, validación y pruebas

Por otro lado, en la tabla 6.6 se muestra cómo las medidas estadísticas presentan valores más bajos respecto a las obtenidas en el segundo modelo. Además, dado que la métrica *F1-score* se calcula como la media armónica entre la precisión y la exhaustividad y los valores de *F1-score* en ambas clases son más bajos sobre el modelo actual (0.5486 para la clase 'cerebro sano' y 0.5984 para la clase 'cerebro con HSVE') que sobre el segundo modelo (0.6927 para 'cerebro sano' y 0.7386 para 'cerebro con HSVE'), se determina que el uso exclusivamente de imágenes MRI capturadas desde el plano AXIAL no permite mejorar los resultados de clasificación de las imágenes del conjunto de pruebas. Además, la exhaustividad obtenida en este séptimo modelo sobre la clase 'Cerebro con HSVE' adquiere un valor de 0.6333, obteniendo así un valor inferior con respecto al valor 0.7479, obtenido con el segundo

modelo. De este modo, el segundo modelo sigue manteniendo los valores más altos en lo relativo a métricas de clasificación.

Clase	Precisión	Exhaustividad	F1-Score	Nº imágenes test
Cerebro sano	0.5849	0.5166	0.5486	60
Cerebro con HSVE	0.5671	0.6333	0.5984	60

Tabla 6.6: Métricas de clasificación del séptimo modelo

Adicionalmente, la figura 6.2 presenta la matriz de confusión obtenida tras evaluar el séptimo modelo de clasificación.

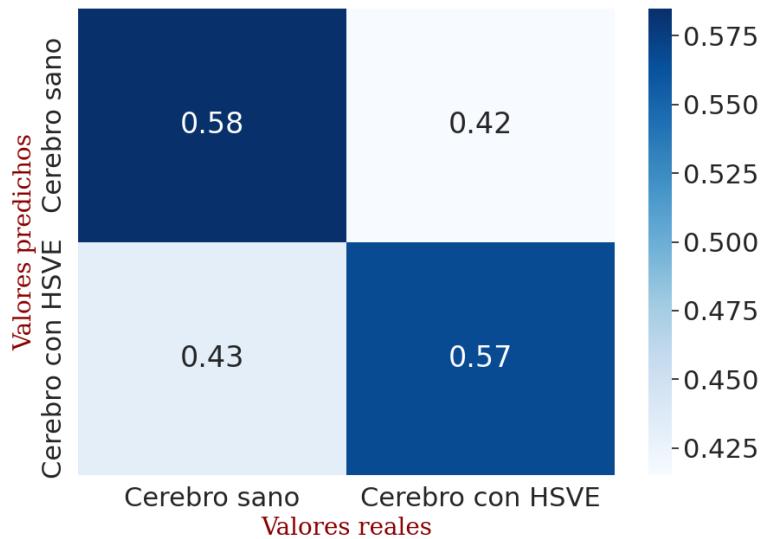


Figura 6.2: Matriz de confusión del séptimo modelo

6.3. Resultados globales iteración 2

En la tabla 6.7 se muestra una comparativa de la exactitud y la exhaustividad obtenidas sobre cada uno de los modelos implementados en la iteración 2. Debido a que cada modelo ha empleado un subconjunto diferente de imágenes del conjunto de datos original, se puede apreciar la capacidad predictiva del modelo pre-entrenado 'ResNet50' sobre los diferentes subconjuntos. Se observa cómo el segundo modelo de la primera iteración, el cual emplea la totalidad de las imágenes del conjunto de datos (745 imágenes), sigue proporcionando los valores más altos en las métricas de clasificación, a diferencia del séptimo y octavo modelo, los cuales emplean exclusivamente imágenes desde el plano AXIAL. De este modo, el segundo modelo de la iteración 1 es capaz de clasificar correctamente el mayor número de imágenes del conjunto de pruebas (el 71.75 %). Además, este modelo presenta el valor de exhaustividad más alto entre los modelos de esta iteración (0.7479), logrando así minimizar en gran medida el número de falsos negativos obtenidos tras evaluar el modelo.

Modelo Implementado	Entrenamiento	Validación	Pruebas	Exhaustividad
<i>Modelo 2</i>	1.0000	0.8507	0.7175	0.7479
<i>Modelo 7:</i> <i>Modelo 2 + sólo MRI AXIAL</i>	1.0000	0.8000	0.575	0.6333
<i>Modelo 8:</i> <i>Modelo 2 + sólo MRI formatos FLAIR, T1 y T2, plano AXIAL</i>	1.0000	0.6275	0.5104	0.5

Tabla 6.7: Comparativa de exactitud y exhaustividad en modelos de CNN implementados en la iteración 2. Resaltado en azul más intenso se muestran los modelos de CNN con valores de exactitud más altos y en azul menos intenso aquellos con valores más bajos.

Capítulo 7

Iteración 3 de experimentos

En este capítulo se describen los modelos de clasificación implementados en esta tercera y última iteración de experimentos. Además, se discuten los resultados obtenidos sobre el último modelo, el cual aplica todas las técnicas de regularización descritas en la sección 2.4.8 y es el modelo que proporciona los valores más altos en las métricas de clasificación de todo el trabajo.

7.1. Modelos implementados

Partiendo del modelo 2, siendo éste el que mejores resultados de clasificación ha proporcionado hasta ahora, con el objetivo de reducir el sobre-ajuste de los modelos anteriores, en esta tercera y última iteración se procede a aplicar las siguientes técnicas de regularización, descritas en la sección 2.4.8 del capítulo 2:

1. Capas de *Dropout*
2. Capas de *Batch Normalization*
3. Aumento de datos (*Data Augmentation* en inglés)
4. Los *callbacks* de *Early Stopping* que se llevan aplicando desde el primer modelo de la iteración 1 y que permiten detener el entrenamiento del modelo si la exactitud o la pérdida en el conjunto de validación no mejora en 10 *epochs* consecutivas

7.1.1. Modelo 9 - Modelo 2 + *Dropout* + *Batch Normalization*

Partiendo del modelo 2, siendo éste el que mejores resultados de clasificación proporcionó tanto en la primera como en la segunda iteración, en este nuevo modelo se ha modificado la parte del clasificador de la arquitectura de la red. Este clasificador ahora está compuesto, además de por la capa que vectoriza la salida del bloque de aprendizaje (capa 'Flatten' en keras), por 3 bloques de la forma:

1. Capa completamente conectada con 256, 128 y 64 neuronas en el primer, segundo y tercer bloque respectivamente.
2. Capa de *Batch Normalization* antes de aplicar la función de activación de cada capa completamente conectada.
3. Función de activación de tipo 'relu' después de la capa de *Batch Normalization*.
4. Capa de *Dropout* con valores de 0.5, 0.5 y 0.2 después de cada función de activación, desactivando así de manera aleatoria aproximadamente el 50 % o el 20 % de las neuronas de cada capa en cada iteración durante el entrenamiento de la red.

La representación de dicha arquitectura, implementada en keras al igual que el resto de modelos de CNN, es especificada en el apéndice B.

7.1.2. Modelo 10 - Modelo 9 + aumento de datos

Partiendo del modelo anterior, en este último modelo se ha aplicado la técnica de regularización aumento de datos o *Data Augmentation*. Dado que las imágenes sintéticas generadas a partir de las imágenes originales no son reales, se ha decidido aplicar esta técnica exclusivamente sobre el conjunto de entrenamiento, de tal modo que el conjunto de validación y el conjunto de pruebas puedan validar el entrenamiento y evaluar el modelo respectivamente, con imágenes correspondientes a casos reales.

Tras analizar la naturaleza y características de la distribución de imágenes del conjunto de datos original, se ha observado que todas ellas presentan una pequeña variación de la posición del cerebro sobre los ejes X e Y. Además, dado que no se han encontrado indicios de que la enfermedad HSVE afecte con mayor frecuencia a los lóbulos de un hemisferio respecto al contrario, se ha asumido que es indiferente el hemisferio sobre el que se identifique la enfermedad en cuestión. Segundo la descripción realizada por el portal de enfermedades raras Orphanet, la enfermedad afecta de forma asimétrica a los lóbulos frontal y temporal, por lo que no se determina que efecte en un mayor número de los casos a un hemisferio respecto al otro (Orphanet, 2015).

En este experimento en concreto, se ha decidido generar 2 nuevas imágenes a partir de cada imagen de entrenamiento, de tal modo que estas imágenes sintéticas sean creadas tras la aplicación de combinaciones de las modificaciones especificadas a continuación:

1. Volteo horizontal de la imagen: al asumir la hipótesis que determina que es indiferente el hemisferio sobre el que haya afectado la enfermedad al paciente.
2. Desplazamiento de la imagen en el eje X de hasta 9 píxeles hacia la izquierda o hacia la derecha.
3. Desplazamiento de la imagen en el eje Y de hasta 9 píxeles hacia arriba o hacia abajo.

De este modo, se logra generar imágenes donde el cerebro queda desplazado como máximo 9 píxeles hacia izquierda o hacia la derecha, y hacia arriba o hacia abajo, y donde se encuentra posicionado desde el lado derecho o izquierdo (volteo horizontal de la imagen). De este modo, se logra construir un modelo más preciso respecto al anterior al poder detectar un mayor número de patrones sobre una mayor cantidad de imágenes. La modificación del brillo de la imagen original se podría haber tenido también en cuenta dado que las imágenes MRI originales del conjunto de datos presentan variaciones en su brillo, probablemente debido al momento en el que las imágenes fueron tomadas o por diferencias en el escáner que capturó las imágenes. Sin embargo, dado que, para poder aplicar esta transformación, el módulo de keras precisa que las imágenes se encuentren en el rango de valores de intensidad [0, 255], y las imágenes ya han sido normalizadas para que estén en el rango [0, 1], no ha sido posible aplicarlo. Se ha probado a aplicar esta transformación sobre imágenes normalizadas y se produce la saturación de las mismas, generando así imágenes sólo con un fondo negro. En el futuro, se podría evaluar el rendimiento de entrenar un modelo con imágenes sin normalizar y donde se haya aplicado la técnica aumento de datos incluyendo esta modificación del brillo.

En la figura 7.1 se muestran las 2 nuevas imágenes generadas a partir de una imagen MRI con formato T2 y capturada desde el plano AXIAL de un cerebro con HSVE del conjunto de entrenamiento. Se puede apreciar cómo la imagen sintética 1 se ha generado tras aplicar un leve desplazamiento de la imagen original sobre el eje X e Y, de tal modo que el cerebro queda desplazado varios píxeles hacia la derecha y hacia abajo. Por otro lado, se puede apreciar cómo la imagen sintética 2 ha sido generada a partir de un claro volteo de la imagen original sobre el eje horizontal.

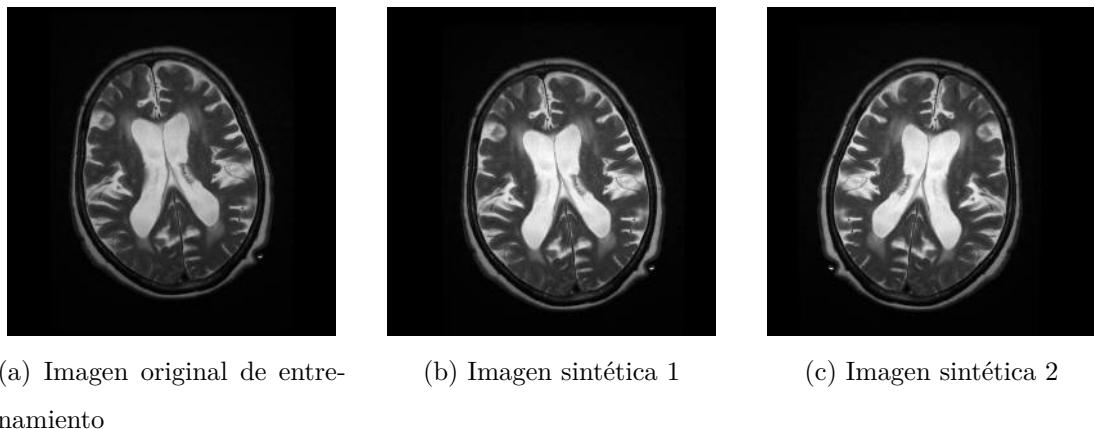


Figura 7.1: Imágenes sintéticas generadas con aumento de datos

Tras generar las nuevas imágenes con aumento de datos, la distribución de imágenes a lo largo de los 3 diferentes conjuntos es la siguiente:

1. Conjunto de entrenamiento: formado por 1083 imágenes (74 % del total).
2. Conjunto de validación: formado por 161 imágenes (11 % del total).
3. Conjunto de pruebas: formado por 223 imágenes (15 % del total).

7.2. Resultados obtenidos

En esta sección se discuten los resultados obtenidos sobre el modelo final, al ser éste el que emplea todas las técnicas de regularización mencionadas previamente y aquel que proporciona valores más altos en las métricas de clasificación de todos los experimentos realizados en el trabajo.

7.2.1. Modelo 10 - Modelo 9 + aumento de datos

En la figura 7.2 se aprecia la evolución de la exactitud y la pérdida del último modelo.

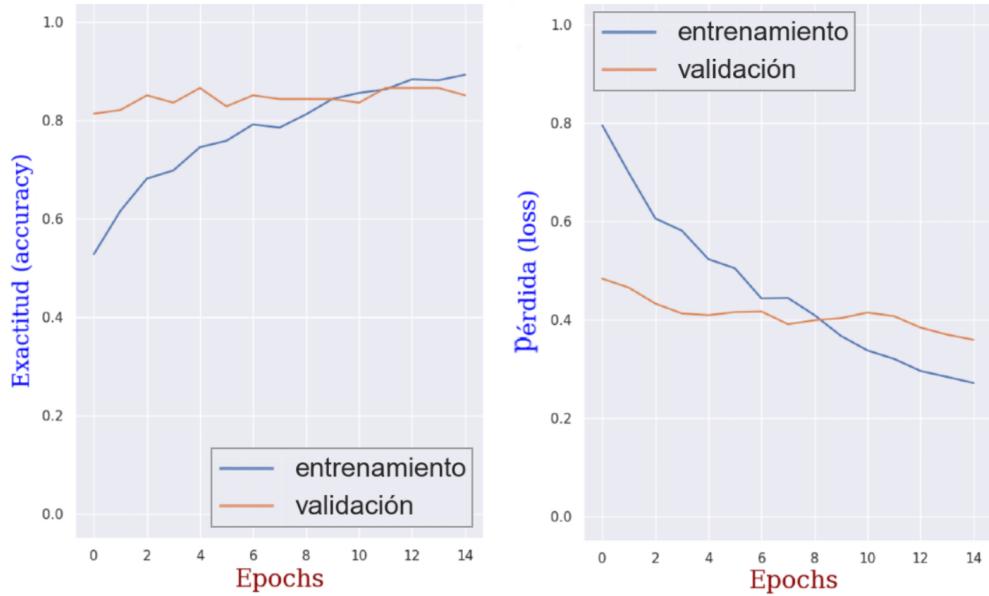


Figura 7.2: Evolución de exactitud y pérdida del modelo final

En la tabla 7.1 se puede ver que este décimo y último modelo obtiene un valor de exactitud sobre el conjunto de validación de 0.8507, exactamente el mismo valor obtenido, tanto en el modelo 2 como en el modelo 9. Sin embargo, la exactitud obtenida en este modelo final sobre el conjunto de pruebas es de 0.8295, logrando así clasificar correctamente el 82.95 % de las imágenes de este conjunto. Tras haber obtenido un valor de exactitud en pruebas superior a 0.80, y superando en 0.112 al valor de exactitud obtenido en pruebas en el modelo 2, se determina que la aplicación de las 4 técnicas de regularización en el décimo modelo mejora los resultados sobre las métricas de clasificación respecto a los modelos anteriores. Adicionalmente, se especifica que el tiempo de entrenamiento en este último modelo es de 90 minutos y 12 segundos, necesitando así más tiempo el modelo para completar su entrenamiento en relación al tiempo de entrenamiento de los modelos anteriores. Este incremento del tiempo de entrenamiento se debe principalmente al uso de un mayor número de imágenes en el conjunto de entrenamiento tras la generación de imágenes sintéticas mediante aumento de datos.

	Exactitud (accuracy)			Tiempo entrenamiento
	Entrenamiento	Validación	Pruebas	
Modelo 10: Modelo 9 + aumento de datos	0.8926	0.8507	0.8295	90' 12"

Tabla 7.1: Evaluación del décimo modelo mediante la exactitud durante entrenamiento, validación y pruebas

Por otro lado, en la tabla 7.2 se muestra cómo las medidas estadísticas presentan valores más altos con respecto a las obtenidas en el segundo y noveno modelo. Además, dado que la métrica *F1-score* se calcula como la media armónica entre la precisión y la exhaustividad y los valores de *F1-score* en cada clase son más altos sobre este modelo final (alrededor de 0.83 en ambas clases) que sobre el segundo modelo (alrededor de 0.69 para la clase de cerebros sanos y alrededor de 0.74 para la clase de cerebros con HSVE), se determina que la aplicación de las técnicas de regularización descritas en la sección 2.4.8 permite alcanzar valores más altos en las métricas de clasificación con respecto a los modelos anteriores, que no aplican todas estas técnicas. De la misma forma, la exhaustividad obtenida en este décimo y último modelo, sobre la clase 'Cerebro con HSVE', adquiere un valor de 0.7731, obteniendo así un valor superior con respecto al valor 0.7479, obtenido con el segundo modelo.

Clase	Precisión	Exhaustividad	F1-Score	Nº imágenes test
Cerebro sano	0.775	0.8942	0.8303	104
Cerebro con HSVE	0.8932	0.7731	0.8288	119

Tabla 7.2: Métricas de clasificación del décimo modelo

Adicionalmente, la figura 7.3 presenta la matriz de confusión obtenida tras evaluar el décimo modelo de clasificación.

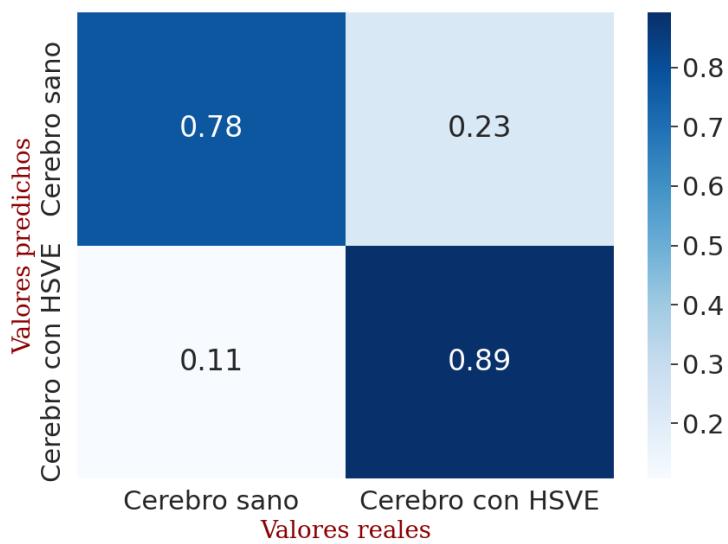


Figura 7.3: Matriz de confusión del modelo final

Por último, con el objetivo de comprobar e identificar aquellas imágenes que le cuesta más clasificar al modelo final, en la figura 7.4 se muestran las imágenes del conjunto de pruebas que este modelo ha clasificado incorrectamente.

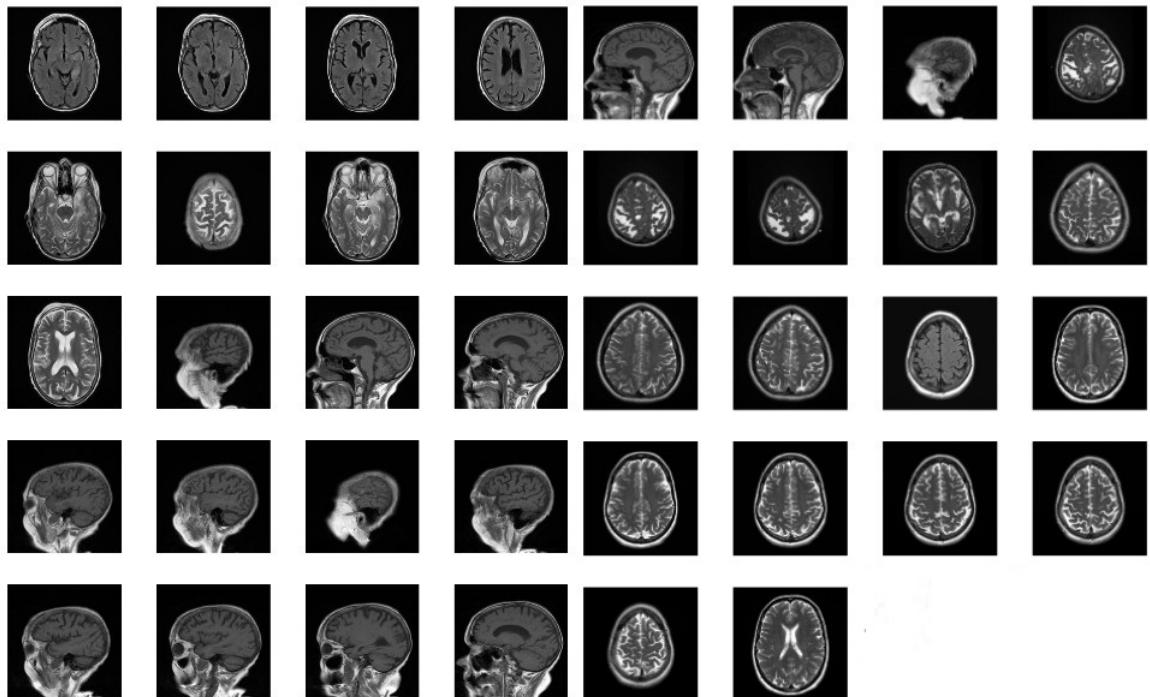


Figura 7.4: Imágenes clasificadas incorrectamente por el modelo final

Del total de 223 imágenes MRI que conforman el conjunto de pruebas del conjunto de datos empleado para el modelo final, las 38 imágenes que se presentan en la figura anterior son aquellas que dicho modelo ha clasificado incorrectamente. Se pueden distinguir algunas imágenes, tales como las presentes en la fila 2 y columna 6 o en la fila 4 y columna 3, en las que se ha capturado una sección muy pequeña del cerebro y correspondiente a la parte más superficial del encéfalo. Es muy probable que el modelo haya fallado al predecir la clase de algunas de estas imágenes debido a la escasa superficie del cerebro escaneada. No obstante, existen otras imágenes que presentan una mayor superficie del cerebro con respecto al fondo y que el modelo ha fallado a la hora de clasificarlas. Se considera que las dos principales causas por las que puede fallar el modelo final son las siguientes:

1. Disponer de casos de la enfermedad en su fase más temprana (en los que la enfermedad no se haya desarrollado del todo), lo cual deriva en una mayor dificultad para la red de encontrar patrones visuales que clasifiquen el ejemplo como un cerebro con HSVE.
2. Haber empleado para el entrenamiento de la red ejemplos formados por una sola sección del cerebro (imagen o dato bidimensional), por lo que puede que en esa sección del cerebro no se pueda identificar la enfermedad, en caso de tratarse de un cerebro con HSVE. Para resolver esta situación, se propone para trabajos futuros probar a entrenar modelos de red convolucional que soporten ejemplos formados por varias imágenes MRI del mismo caso (cubos de imágenes o datos tridimensionales). De este modo, la red dispondrá de mayor información del cerebro en cada ejemplo (mayor número de secciones del cerebro en forma de una secuencia de imágenes MRI) para decidir si éste presenta la enfermedad o si, por el contrario, se trata de un cerebro sano.

No obstante, además de las posibles causas expuestas anteriormente, uno de los trabajos futuros puede consistir en investigar, mediante la ayuda de un experto del campo médico, por qué el modelo falla más al clasificar ciertas imágenes del cerebro respecto a otras que aparentemente son similares.

7.3. Resultados globales iteración 3

En la tabla 7.3 se muestra una comparativa de la exactitud y la exhaustividad obtenidas sobre los dos nuevos modelos implementados en la iteración 3. En esta última iteración, se han implementado y evaluado dos modelos de forma incremental, de tal forma que el noveno modelo aplica las técnicas de *Dropout*, *Batch Normalization* y *Early Stopping*, y el décimo y último modelo, además de emplear estas técnicas, aplica aumento de datos sobre las imágenes del conjunto de entrenamiento. Se puede apreciar cómo el modelo final, tras aplicar cada una de estas técnicas de regularización, logra alcanzar un valor de exactitud sobre el conjunto de pruebas de 0.8295, siendo así capaz de clasificar correctamente el 82.95 % de las imágenes por las que está formado el conjunto de pruebas. Además, este último modelo presenta una exhaustividad de 0.7731, logrando así minimizar más aún el número de falsos negativos obtenidos tras evaluar el modelo.

Modelo Implementado	Entrenamiento	Validación	Pruebas	Exhaustividad
Modelo 9: Modelo 2 + <i>Dropout</i> + <i>Batch Normalization</i>	0.8833	0.8507	0.7399	0.7395
Modelo 10: Modelo 9 + aumento de datos	0.8926	0.8507	0.8295	0.7731

Tabla 7.3: Comparativa de exactitud y exhaustividad en modelos de CNN implementados en la iteración 3. Resaltado en azul más intenso se muestra el modelo de CNN con el valor de exactitud más alto y en azul menos intenso aquel con el valor más bajo.

Capítulo 8

Conclusiones y Trabajos Futuros

8.1. Conclusiones

El presente trabajo ha implementado una serie de modelos de Red Neuronal Convolucional en los que se han aplicado diferentes técnicas de *Deep Learning* con el objetivo de clasificar imágenes de secuencias MRI, capturadas sobre cerebros humanos, en 2 clases diferentes: cerebros sanos (clase negativa) y cerebros con la enfermedad rara HSVE (clase positiva). Concretamente, se han construido los 10 modelos que se listan a continuación a lo largo de 3 iteraciones: 6 modelos en una primera iteración dedicados a evaluar el rendimiento de diferentes arquitecturas de CNN sobre el problema de clasificación binaria, 2 modelos en la segunda iteración dedicados a identificar el subconjunto de imágenes del conjunto de datos inicial que mejores resultados de clasificación proporciona y 2 modelos en la tercera iteración dedicados a la aplicación de forma incremental de las siguientes 4 técnicas de regularización: *Dropout*, *Batch Normalization*, *Early Stopping* y aumento de datos (*Data Augmentation*).

Una vez entrenados y evaluados cada uno de estos modelos de carácter experimental, se determina que el problema de clasificación se resuelve con resultados satisfactorios, alcanzando el último modelo de la tercera iteración una exactitud en pruebas de 0.8295, es decir, es capaz de clasificar correctamente el 82.95 % de las imágenes de secuencias *MRI* en sus 2 respectivas clases. Además, este último modelo alcanza

una exhaustividad de 0.7731, siendo éste capaz de clasificar correctamente el 77.31 % de las imágenes MRI de cerebros con la enfermedad HSVE, habiendo clasificado incorrectamente como cerebros sanos el 22.69 % de las imágenes MRI de cerebros con HSVE.

En la figura 8.1 se muestran, de forma esquemática y resumida, los diferentes modelos implementados en cada iteración de experimentos junto a sus resultados de exactitud obtenidos sobre el conjunto de pruebas.

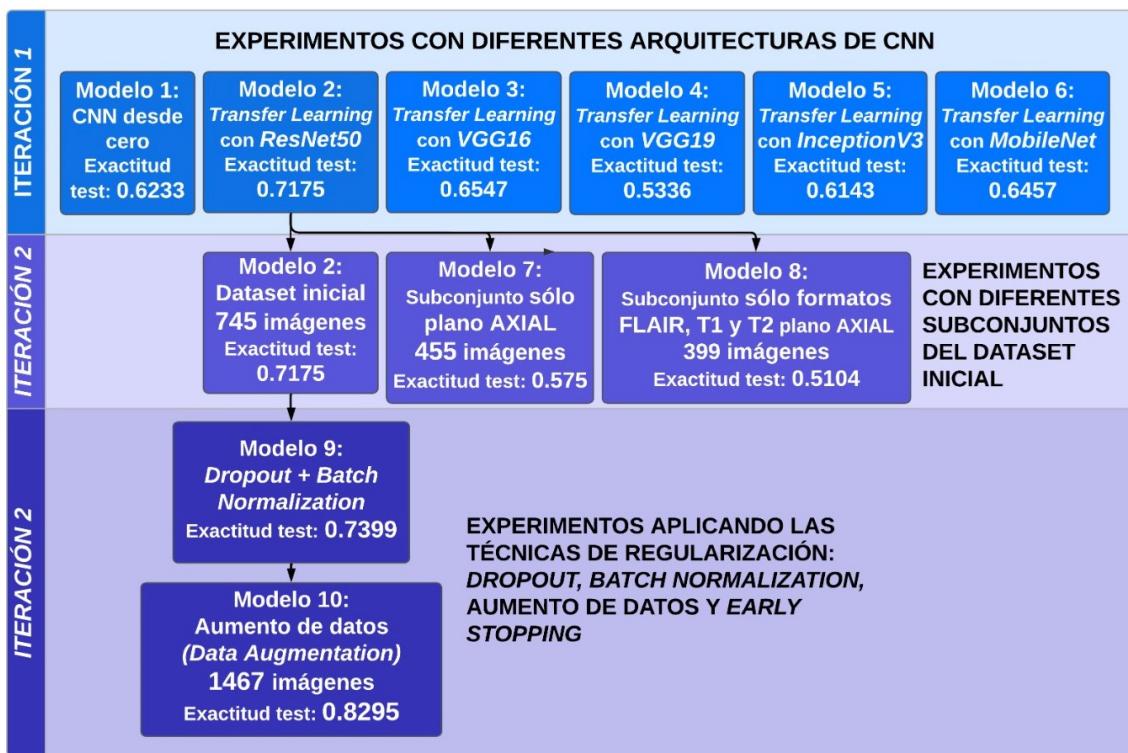


Figura 8.1: Descripción y resultados de experimentos por iteración

8.2. Líneas de trabajos futuros

Con el objetivo de mejorar los resultados obtenidos sobre el problema de clasificación binaria de imágenes MRI de cerebros en función de la presencia de la enfermedad rara HSVE, se propone realizar en el futuro alguna de las siguientes tareas:

1. Adquirir imágenes de secuencias MRI pertenecientes a nuevos casos de cerebros con encefalitis herpética, así como de cerebros con un funcionamiento normal, que hayan sido diagnosticados por expertos. Esto será fundamental para aumentar el número de imágenes del conjunto inicial de datos, lo cual permitirá reducir el sobreajuste detectado en los modelos implementados en el presente proyecto.
2. Evaluar el rendimiento de aplicar los mismos experimentos de este trabajo (con posibles variaciones en la arquitectura, hiperparámetros o técnicas aplicadas) sobre las nuevas imágenes obtenidas. En caso de disponer GPU's con mayor capacidad de cómputo con respecto a la empleada en este trabajo, se propone emplear arquitecturas de mayor escala o con mayor número de parámetros para permitir al modelo extraer un mayor número de patrones en las imágenes.
3. Aplicar *Transfer Learning*, seleccionando modelos pre-entrenados sobre imágenes MRI de naturaleza similar a las empleadas en este trabajo y aprovechar así las características extraídas por estos modelos. Además de *Transfer Learning* se propone aplicar *Fine-Tuning* sobre una mayor cantidad de imágenes con el objetivo de mejorar la detección de características de alto nivel o encontrar características nuevas sobre las imágenes MRI del conjunto de datos.
4. Con la ayuda de un experto en el ámbito médico, analizar e investigar las posibles causas por las que el modelo final implementado en el presente trabajo clasifica incorrectamente ciertas imágenes, las cuales aparentemente son muy similares a otras que sí son clasificadas correctamente. Verificar si las dos posibles causas que se han descrito en los resultados del modelo final de este trabajo son ciertas.

5. Probar a entrenar modelos de CNN que soporten ejemplos formados por varias imágenes de secuencias MRI del mismo caso para dotar a la red de mayor información del cerebro en cada ejemplo, permitiéndole discriminar mejor si éste presenta la enfermedad o no.
6. Además, se propone re-entrenar los modelos implementados en el presente trabajo sobre imágenes MRI de cerebros que presenten otras enfermedades o trastornos raros. De este modo, se evaluará la capacidad de adaptación de estos modelos a imágenes de naturaleza y formato similar. También puede resultar interesante construir un conjunto de datos más grande y diverso en el que, además de las clases 'Cerebros sanos' y 'Cerebros con HSVE', se empleen más clases relativas a otras enfermedades o trastornos similares.

De este modo, las propuestas anteriores pretenden servir de guía para mejorar en el futuro los resultados de clasificación relativos al presente problema de visión por computador, así como servir de ayuda en el diagnóstico de la enfermedad HSVE en el ámbito médico.

Bibliografía

- Andrew Ng, Younes Bensouda Mourri, K. K. (2021). *Deep Learning specialization program.* <https://www.coursera.org/specializations/deep-learning>.
- BOE (2015). *BOE-A-2015-14083 Real Decreto 1091/2015, de 4 de diciembre, por el que se crea y regula el Registro Estatal de Enfermedades Raras.* <https://www.boe.es/buscar/act.php?id=BOE-A-2015-14083>.
- Chollet, F. (2015). Keras api. <https://keras.io/api/>.
- Chollet, F. (2018). *Deep Learning with Python*. Manning Publications, 2^a edition.
- Cortés, D. F. (2015). *LAS ENFERMEDADES RARAS*. revista médica clínica las condes. <https://t.ly/bXhX0>.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. pages 248–255.
- Gaillard, F. (2023). *Normal brain imaging examples*. radiopaedia.org [data set]. <https://radiopaedia.org/articles/normal-brain-imaging-examples-1?lang=us>. [Accessed 14-May-2023].
- Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow*. O'Reilly Media, Inc, USA, 2^a edition.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications.

Khairandish, M. O., Sharma, M., Jain, V., Chatterjee, J. M., and Jhanjhi, N. Z. (2022). A hybrid cnn-svm threshold segmentation approach for tumor detection and classification of mri brain images. *IRBM*, 43:290–299.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition.

Marcocci, M. E., Napoletani, G., Protto, V., Kolesova, O., Piacentini, R., Puma, D. D. L., Lomonte, P., Grassi, C., Palamara, A. T., and Chiara, G. D. (2020). Herpes simplex virus-1 in the brain: The dark side of a sneaky infection. *Trends in Microbiology*.

Maria Elena Marcocci, Giorgia Napoletani, V. P. O. K. R. P. D. D. L. P. P. L. C. G. A. T. P. and Chiara, G. D. (2020). Herpes simplex virus-1 in the brain: The dark side of a sneaky infection. *Trends in Microbiology*, 28:808–820.

MayoClinic (2023). *Encefalitis - Síntomas y causas - Mayo Clinic*. <https://www.mayoclinic.org/es-es/diseases-conditions/encephalitis/symptoms-causes/syc-20356136>.

Muñoz Garrido, J. (2022). *Neurociencia Cognitiva. Tema 4: Técnicas de estudio del sistema nervioso [Material del aula]. Texto académico del máster. Universidad Internacional de La Rioja (UNIR)*.

Orphanet (2015). *Encefalitis por el virus del herpes simple*. [https://www.orpha.net/consor/cgi-bin/Disease_Search.php?lng=ES&data_id=3275&Disease_Disease_Search_diseaseGroup=Encefalitis-Herpetica&Disease_Disease_Search_diseaseType=Pat&Enfermedad\(es\)/grupo%20de%20enfermedades=Encefalitis-por-el-virus-del-herpes-simple&title=Encefalitis%20por%20el%20virus\%20del%20herpes%20simple&search=Disease_Search_Simple](https://www.orpha.net/consor/cgi-bin/Disease_Search.php?lng=ES&data_id=3275&Disease_Disease_Search_diseaseGroup=Encefalitis-Herpetica&Disease_Disease_Search_diseaseType=Pat&Enfermedad(es)/grupo%20de%20enfermedades=Encefalitis-por-el-virus-del-herpes-simple&title=Encefalitis%20por%20el%20virus\%20del%20herpes%20simple&search=Disease_Search_Simple).

Preston, M. D. C. (2016). *Magnetic Resonance Imaging (MRI) of the Brain and Spine*. basics. <https://case.edu/med/neurology/NR/MRI%20Basics.htm>.

Radiopaedia (2023). *HSV Cases search*. radiopaedia.org [data set]. https://radiopaedia.org/search?diagnostic_certainty=confirmed_substantiated&lang=us&modality=MRI&page=1&q=HSV&scope=cases&sort=completeness.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.

Srinivas, C., Prasad, N., Zakariah, M., Alothaibi, Y. A., Shaukat, K., Partibane, B., and Awal, H. (2022). Deep transfer learning approaches in performance analysis of brain tumor classification using mri images.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2015). Rethinking the inception architecture for computer vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December:2818–2826.

Wang, W., Li, Y., Zou, T., Wang, X., You, J., and Luo, Y. (2020). A novel image classification approach via dense-mobilenet models. *Mobile Information Systems*, 2020.

WHO (2023). *Herpex Simplex Virus*. the world health organization. <https://www.who.int/es/news-room/fact-sheets/detail/herpes-simplex-virus>. [Accessed 29-Apr-2023].

Xiang, Y., Zeng, C., Liu, B., Tan, W., Wu, J., Hu, X., Han, Y., Luo, Q., Gong, J., Liu, J., and Li, Y. (2021). *Deep Learning-Enabled Identification of Autoimmune Encephalitis on 3D Multi-Sequence MRI*. <https://pubmed.ncbi.nlm.nih.gov/34478565/>.

Yamanakkanavar, N., Choi, J. Y., and Lee, B. (2020). Mri segmentation and classification of human brain using deep learning for diagnosis of alzheimer's disease: A survey. *Sensors 2020, Vol. 20, Page 3243*, 20:3243.

Deep Learning para la detección de Encefalitis Herpética (HSVE)

Carlos Breuer Carrasco



Universidad Internacional de la Rioja, Logroño (España)

16 de julio de 2023

RESUMEN

Con el propósito de desarrollar herramientas complementarias al diagnóstico médico de la enfermedad rara *encefalitis herpética* (HSVE), el presente trabajo realiza el diseño, implementación y evaluación de modelos de red neuronal convolucional (CNN) para clasificar imágenes de resonancia magnética (MRI) del cerebro humano según esté sano o presente dicha enfermedad. Se ha seguido una metodología incremental, de tal modo que la exactitud de clasificación obtenida en cada modelo trata de mejorar la obtenida en los modelos anteriores, alcanzando, finalmente, una exactitud en pruebas de 0.83 y una exhaustividad de 0.77. Se han empleado modelos pre-entrenados mediante *Transfer Learning* y aplicado técnicas de regularización para reducir el sobreajuste. Los experimentos realizados en este trabajo sientan las bases de futuros trabajos a mayor escala, en los que se entrenen y perfeccionen modelos de CNN empleando un mayor número de imágenes MRI, permitiendo obtener resultados más fiables.

I. INTRODUCCIÓN

El objetivo del presente trabajo consiste en implementar y evaluar diversos modelos de red neuronal convolucional (CNN) para la clasificación de imágenes de secuencias de resonancia magnética (MRI) del cerebro humano según se encuentre sano o afectado por la enfermedad rara encefalitis herpética (HSVE). Con el fin de obtener resultados precisos y significativos, se ha seguido un procedimiento compuesto por varias fases.

En primer lugar, se obtuvo un total de 745 imágenes MRI del repositorio *Radio-pedia* (<https://radiopaedia.org/?lang=us>), las cuales fueron sometidas a un proceso de pre-procesamiento. A continuación, se llevó a cabo una primera fase de experimentos

consistente en evaluar el rendimiento de modelos pre-entrenados de CNN con diferentes arquitecturas sobre el problema de clasificación. Posteriormente, se realizó una segunda fase de experimentos para determinar cuál de 3 subconjuntos, seleccionados a partir del conjunto inicial de imágenes, se adaptaba mejor a un modelo de CNN y, finalmente, se realizó una tercera fase consistente en evaluar el impacto sobre las métricas de clasificación tras aplicar técnicas de regularización.

Al finalizar todos los experimentos, se determinó que el problema de clasificación es resuelto alcanzando resultados satisfactorios. Concretamente, el último modelo de la tercera iteración alcanzó una exactitud (*accuracy*) sobre el conjunto de pruebas (test) del 0.8295, siendo capaz de clasificar correctamente el

PALABRAS CLAVE

Deep Learning, Encefalitis herpética (HSVE), Imágenes de Resonancia magnética (MRI), Red neuronal convolucional (CNN), *Transfer Learning*

82.95 % de las imágenes MRI en las dos clases correspondientes. Este último modelo empleó mediante *Transfer Learning* el modelo pre-entrenado 'ResNet50' (He et al., 2015), al ser el único en superar el valor 0.70 tanto en exhaustividad como en exactitud en pruebas en la primera iteración de experimentos. Además, se utilizó el conjunto inicial compuesto por 745 imágenes, al tratarse del subconjunto de datos que mejor rendimiento presentó en la segunda iteración, junto a las imágenes sintéticas generadas mediante aumento de datos en la tercera iteración. Por otro lado, se aplicaron varias técnicas de regularización, mejorando así los resultados sobre las métricas de clasificación del modelo y reduciendo el sobreajuste inicial. Estos resultados, los cuales se presentan de forma esquemática en la figura 1, demuestran la eficacia del uso de CNN, así como de otras técnicas de *Deep Learning*, en la clasificación de imágenes MRI de cerebros sanos y afectados por HSVE.

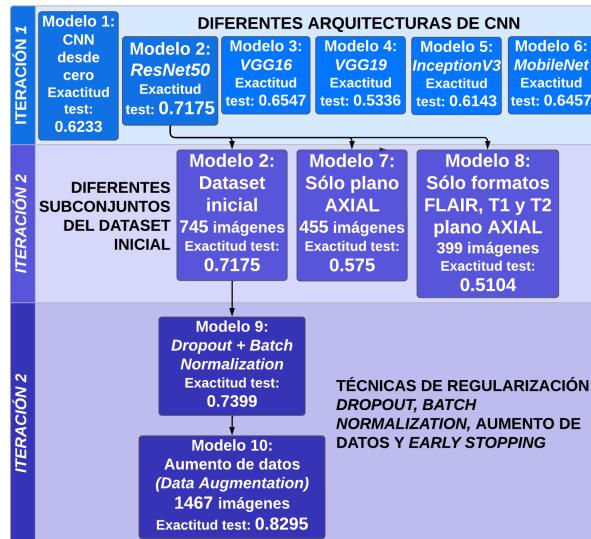


Figura 1: Resultados de experimentos

II. ESTADO DEL ARTE

A. Encefalitis herpética simple

La HSVE es una enfermedad rara causada por el virus del herpes simple que provoca inflamación en el cerebro (MayoClinic, 2023).

Se estima que afecta a 1 de cada 250.000 o 500.000 personas y representa el 20 % de todos los casos de encefalitis. La HSVE suele afectar de manera asimétrica a los lóbulos frontal y temporal del cerebro, lo que puede provocar disfunciones cognitivas en el lenguaje, en la memoria, en el comportamiento y en la personalidad (Orphanet, 2015).

Existen dos tipos de HSVE, dependiendo del tipo de virus del herpes involucrado. El virus del herpes simple tipo 1 (HSV-1) representa el 90 % de los casos, mientras que el virus del herpes simple tipo 2 (HSV-2) constituye el 10 % restante. El HSV-1 se transmite principalmente a través del contacto del virus en los ojos, los labios o la saliva, mientras que el HSV-2 se transmite principalmente por contacto sexual (WHO, 2023).

B. Resonancia Magnética Nuclear

La resonancia magnética nuclear (RMN) es una técnica ampliamente utilizada para estudiar el sistema nervioso humano. Se basa en la aplicación de una onda electromagnética sobre los núcleos de hidrógeno presentes en el cerebro, permitiendo obtener imágenes detalladas de su estructura y densidad para la detección de distintas enfermedades, entre ellas, HSVE. Además, en el campo de la Inteligencia Artificial, estas imágenes se utilizan para construir sistemas de segmentación y/o clasificación, sirviendo de gran ayuda para el diagnóstico médico del paciente (Muñoz Garrido, 2022).

C. Red Neuronal Convolutacional

El aprendizaje profundo, también conocido como *Deep Learning*, es un subcampo de la Inteligencia Artificial y del *Machine Learning* que se centra principalmente en desarrollar sistemas predictivos capaces de encontrar patrones complejos y abstractos sobre grandes cantidades de datos utilizando redes neuronales artificiales con múltiples capas. Esto ha llevado al desarrollo de sistemas exitosos dedicados al reconocimiento de imágenes, reconocimiento facial, traducción automática y segmentación de imágenes, entre otras tareas.

En el presente proyecto se ha empleado como arquitectura algorítmica una red neuronal convolucional. Las redes que emplean este tipo de arquitectura aplican operaciones de convolución y de *pooling*, utilizando sus respectivos filtros para detectar patrones en las imágenes a partir de diferentes representaciones de las mismas.

En el ámbito médico, varios estudios, como el realizado por Yayun Xiang junto a otros investigadores (Xiang et al., 2021), han utilizado CNN junto a MRI para extraer características discriminativas con el objetivo de analizar, clasificar e identificar diferentes tipos de encefalitis, entre ellos la herpética simple.

D. Investigaciones recientes relativas al tema del proyecto

Tal y como se comentó en la sección anterior, en el ámbito de la clasificación de encefalitis, destaca el trabajo realizado por Yayun Xiang junto a otros investigadores (Xiang et al., 2021). En dicho trabajo se implementaron varios modelos de *Deep Learning* en los que, mediante el uso de imágenes MRI multiselencia, se logró su identificación y clasificación, obteniendo una exactitud en pruebas de 0.831 sobre las imágenes MRI de encefalitis autoinmune, de 0.882 sobre MRI de encefalitis herpética simple y de 0.892 sobre MRI de sujetos de control sanos.

Otras investigaciones recientes han construido modelos de CNN con el propósito de clasificar imágenes MRI de cerebros humanos según diferentes tipos de tumores cerebrales. Un ejemplo de ello es el estudio realizado por Chetana Srinivas junto a otros investigadores (Srinivas et al., 2022), en el que se emplean modelos pre-entrenados de red convolucional de diferentes arquitecturas, con el objetivo de evaluar su rendimiento en la clasificación binaria de imágenes MRI de cerebros sanos o con tumores. Tras entrenar cada uno de los modelos de CNN, aquellos que proporcionaron un mejor rendimiento

en la clasificación de las MRI fueron ResNet50 y VGG-16, con valores de exactitud de 0.96 y 0.95 respectivamente. Dados estos resultados tan favorables, el presente trabajo trata de evaluar el rendimiento de estos modelos pre-entrenados, entre otros, sobre el problema de clasificación de imágenes MRI para la detección de la enfermedad rara HSVE.

Sin embargo, se debe tener en cuenta que el rendimiento puede variar dependiendo de los conjuntos de datos utilizados, el problema de clasificación planteado sobre una enfermedad dada, la arquitectura de la red neuronal, los métodos de pre-procesamiento y otros factores. Además, se deberán concebir los sistemas basados en visión por computador como herramientas de apoyo para el diagnóstico realizado por el experto médico, en ningún caso como un sistema para automatizar el diagnóstico del paciente.

III. OBJETIVOS Y METODOLOGÍA

A. Objetivo general

El principal objetivo del proyecto consiste en emplear técnicas de *Deep Learning* aplicadas a la clasificación de imágenes de resonancia magnética del cerebro para la detección de la encefalitis herpética simple. La clasificación consiste en identificar si una imagen MRI es de un cerebro con HSVE o si, por el contrario, es de un cerebro sano. Para alcanzar este objetivo, a continuación se definen un conjunto de objetivos específicos.

B. Objetivos específicos

1. Implementar modelos de CNN capaces de clasificar imágenes MRI del cerebro humano según presente encefalitis herpética simple o no (cerebro sano).
2. Obtener conclusiones del trabajo realizado y proponer posibles líneas de trabajo futuro.

C. Metodología

La metodología empleada para el desarrollo del trabajo consta de las siguientes partes. En primer lugar, se adquiere y se prepara la muestra de datos, la cual está compuesta por un total de 745 de imágenes MRI correspondientes a 2 clases diferentes: secuencias MRI de cerebros enfermos con HSVE y secuencias MRI de cerebros sanos, éstas últimas correspondientes a pacientes que presentaron dolores de cabeza. Concretamente, se han empleado 366 imágenes correspondientes a 10 casos de cerebros con HSVE (Radiopaedia, 2023), obtenidas desde el repositorio público *Radiopaedia* tras aplicar una búsqueda de casos por el nombre “HSV” y cuyos diagnósticos habían finalizado. Por otro lado, también se han empleado 379 imágenes correspondientes a 7 casos de cerebros sanos (Gaillard, 2023), obtenidas a través del listado de casos de cerebros normales, elaborado y recopilado por Frank Gaillard en el repositorio *Radiopaedia*. Una vez han sido obtenidas las imágenes, se procede a su pre-procesamiento, redimensionándolas para que la resolución en todas ellas sea ‘256x256x1’ y normalizándolas para que todos sus píxeles se encuentren en el rango [0, 1].

Resulta importante aclarar que para el desarrollo de este proyecto se ha seguido una metodología incremental, tratando en cada modelo de mejorar los resultados obtenidos en los modelos previos. De este modo, después de realizar cada uno de los experimentos en cada iteración, se ha seleccionado aquel que proporciona valores más altos de exactitud y exhaustividad. Cabe destacar que en todos los modelos implementados se ha empleado el optimizador ‘Adam’ con un *learning rate* de 0.0001. Además, en todos ellos se han definido 2 *callbacks* que permiten detener el entrenamiento del modelo si el valor de pérdida o de exactitud en el conjunto de validación no mejora en 10 *epochs* consecutivas. Por otro lado, los modelos son entrenados a lo largo de un máximo de 30 épocas (*epochs*), especificando en el modelo inicial un tamaño del *mini-batch* de 16 muestras y en los modelos

sucesivos de 8 muestras. Los valores empleados para el número de *epochs*, así como para el tamaño del *mini-batch* en estos modelos son dependientes de las características de hardware que dispone el computador utilizado para la realización del presente trabajo. Sobre un computador con mayores prestaciones se podría probar a incrementar el tamaño del *batch* de 8 a 16 ejemplos, o incluso hasta 32 o 64. A continuación, se describen los detalles y peculiaridades de cada modelo implementado:

1. Iteración 1 de experimentos

Modelo 1: en este primer modelo se ha empleado una CNN construida desde cero. Además, para el entrenamiento del modelo se ha utilizado un tamaño del *mini-batch* de 16 ejemplos. La arquitectura empleada para este modelo está formada, en primer lugar, por una capa de entrada de la red y por 2 bloques de la forma: capa convolucional con función de activación ‘relu’ seguida de una capa de *MaxPooling* con un *stride* de 3. En segundo lugar, se añadió una capa que permite vectorizar los datos bidimensionales, seguida por 2 capas completamente conectadas de 256 y 512 neuronas respectivamente, con funciones de activación ‘relu’. Finalmente, se añadió una capa de salida, también completamente conectada, de 1 sola neurona y con función de activación ‘sigmoid’, al tratarse de un problema de clasificación binaria.

Modelo 2: en este segundo modelo se ha modificado la arquitectura anterior de red convolucional para importar los pesos del modelo pre-entrenado ‘ResNet50’, aplicando así la técnica *Transfer Learning*. Dado que la arquitectura de este modelo presenta un mayor tamaño respecto a la anterior, a partir de este modelo se ha utilizado un tamaño menor del mini-batch: 8 ejemplos.

Modelo 3: en este tercer modelo se ha modificado exclusivamente el modelo

pre-entrenado 'ResNet50' por el modelo pre-entrenado 'VGG16'.

Modelo 4: en este cuarto modelo se ha modificado el modelo pre-entrenado anterior por el modelo 'VGG19'.

Modelo 5: en este quinto modelo se ha modificado el modelo pre-entrenado anterior por el modelo 'InceptionV3'.

Modelo 6: en este sexto modelo se ha modificado el modelo pre-entrenado anterior por el modelo 'MobileNet'.

2. Iteración 2 de experimentos

Modelo 7: en este nuevo modelo, se emplea el modelo pre-entrenado 'ResNet50', al tratarse de la única arquitectura que en la iteración 1 superó el valor 0.70 tanto en la exhaustividad como en la exactitud. Sin embargo, se ha utilizado el subconjunto de datos correspondiente a las imágenes MRI capturadas sólo desde el plano AXIAL. Este subconjunto de imágenes está formado por un total de 455 imágenes.

Modelo 8: en este nuevo modelo, se emplea el modelo pre-entrenado 'ResNet50', al igual que en el modelo anterior. Sin embargo, se ha utilizado el subconjunto de datos correspondiente a las imágenes MRI que presentan sólo los formatos FLAIR, T1 y T2 desde el plano AXIAL, descartando aquellas que presenten otros formatos tales como T1C+ o T2C+, o que hayan sido capturadas desde los planos CORONAL y SAGITAL. Este subconjunto de datos está formado por un total de 399 imágenes.

3. Iteración 3 de experimentos

Modelo 9: partiendo del modelo 2, siendo éste el que proporcionó valores de exactitud más altos tanto en la primera como en la segunda iteración, en este nuevo modelo se ha modificado la parte del

clasificador de la arquitectura de la red. Este clasificador ahora está compuesto, además de por la capa que vectoriza la salida del bloque de aprendizaje, por 3 bloques de la forma:

- a) Capa completamente conectada con 256, 128 y 64 neuronas en el primer, segundo y tercer bloque respectivamente.
- b) Capa de *Batch Normalization* después de cada capa completamente conectada.
- c) Función de activación de tipo 'relu' después de la capa de *Batch Normalization*.
- d) Capa de *Dropout* con valores de 0.5, 0.5 y 0.2 después de cada función de activación, desactivando así de manera aleatoria el 50% o el 20% de las neuronas de cada capa en cada iteración durante el entrenamiento.

De este modo, se han aplicado las técnicas de regularización *Batch Normalization* y *Dropout* tanto en este noveno modelo como en el décimo y último modelo, permitiendo reducir así el sobreajuste.

Modelo 10: partiendo del modelo anterior, en este último modelo se ha aplicado la técnica de regularización aumento de datos. De este modo, se generan imágenes sintéticas en el conjunto de entrenamiento a partir de las imágenes originales de dicho conjunto.

Tras analizar la naturaleza y características de la distribución de imágenes del conjunto de datos original, se ha observado que todas ellas presentan una pequeña variación de la posición del cerebro sobre los ejes X e Y. Además, dado que no se han encontrado indicios de que la enfermedad HSVE afecte con mayor frecuencia a los lóbulos de un hemisferio respecto al contrario, se ha asumido que es indiferente el hemisferio sobre el que se identifique la enfermedad en cuestión.

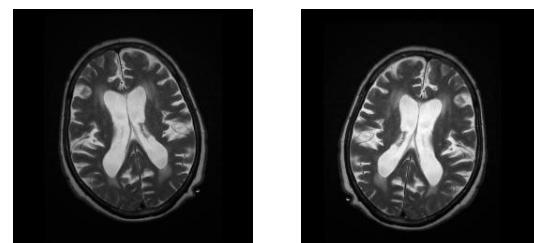
Según la descripción realizada por el portal de enfermedades raras Orphanet, la enfermedad afecta de forma asimétrica a los lóbulos frontal y temporal, por lo que no se determina que efecte en un mayor número de casos a un hemisferio respecto al otro (Orphanet, 2015).

En este experimento en concreto, se ha decidido generar 2 nuevas imágenes por cada imagen de entrenamiento, de tal modo que estas imágenes sintéticas son creadas tras la aplicación de combinaciones de las modificaciones especificadas a continuación:

- a) Volteo horizontal de la imagen: al asumir que resulta indiferente el hemisferio sobre el que haya afectado la enfermedad al paciente.
- b) Desplazamiento de la imagen en los ejes X e Y de hasta 9 píxeles hacia la izquierda/derecha o arriba/abajo.

De este modo, se logra generar imágenes donde el cerebro queda desplazado como máximo 9 píxeles hacia izquierda o hacia la derecha, y hacia arriba o hacia abajo, y donde se encuentra posicionado desde el lado derecho o izquierdo (volteo horizontal de la imagen). De esta forma, se logra entrenar un modelo más robusto capaz de detectar patrones más generales sobre una mayor cantidad de imágenes.

En la figura 2 se muestra 1 de las 2 nuevas imágenes generadas a partir de una imagen MRI con formato T2 y capturada desde el plano AXIAL de un cerebro con HSVE del conjunto de entrenamiento. Se puede apreciar cómo la imagen sintética 1 se ha generado tras aplicar un leve desplazamiento de la imagen original sobre el eje X e Y. Además, se aprecia la aplicación del volteo sobre el eje horizontal de la imagen original.



(a) Imagen original (b) Imagen sintética 1

Figura 2: Imagen original e imagen sintética generada con aumento de datos

Tras generar las nuevas imágenes sintéticas con aumento de datos, la distribución de imágenes a lo largo de los 3 diferentes conjuntos es la siguiente:

1. Conjunto de entrenamiento: formado por 1083 imágenes (74 % del total).
2. Conjunto de validación: formado por 161 imágenes (11 % del total).
3. Conjunto de pruebas: formado por 223 imágenes (15 % del total).

IV. CONTRIBUCIÓN

El trabajo realizado ofrece una comparativa de diferentes soluciones de inteligencia artificial basadas en *Deep Learning* para la clasificación de imágenes MRI de cerebros humanos según se encuentren sanos o presenten la enfermedad HSVE. Además, se sientan las bases para el desarrollo de un sistema de visión artificial que permita identificar, en base a una imagen MRI del cerebro, si un paciente presenta encefalitis herpética o no, sirviendo este sistema de apoyo para el diagnóstico médico del paciente.

V. RESULTADOS

A continuación, se describen los resultados obtenidos en cada uno de los modelos implementados en cada iteración.

A. Iteración 1

Tal y como se puede apreciar en la tabla 1, el modelo de la primera iteración que proporciona el valor de exactitud más alto sobre el conjunto de pruebas es el segundo modelo, con un valor de 0.7175. Por otro lado, el modelo que proporciona un valor de exhaustividad más alto es el primero, con un valor de 0.7983.

Modelo	Exactitud test	Exhaustividad
1	0.6233	0.7983
2	0.7175	0.7479
3	0.6547	0.7479
4	0.5336	0.6555
5	0.6143	0.5378
6	0.6457	0.7815

Tabla 1: Resultados de la iteración 1

B. Iteración 2

Tal y como se observa en la tabla 2, los 2 nuevos modelos implementados en la segunda iteración proporcionan valores de exhaustividad y exactitud en pruebas más bajos con respecto a los obtenidos en el segundo modelo de la primera iteración. De este modo, se obtienen los valores 0.575 y 0.5104 de exactitud en el séptimo y octavo modelo respectivamente.

Modelo	Exactitud test	Exhaustividad
2	0.7175	0.7479
7	0.575	0.6333
8	0.5104	0.5

Tabla 2: Resultados de la iteración 2

C. Iteración 3

Tal y como se puede apreciar en la tabla 3, el décimo modelo, implementado en la tercera iteración y que aplica las cuatro técnicas de regularización, es el que proporciona el valor más alto de exactitud en pruebas, con un valor de 0.8295. Si se atiende a la exhaustividad, el décimo modelo también es el que proporciona el valor más alto: 0.7731.

Modelo	Exactitud test	Exhaustividad
9	0.7399	0.7395
10	0.8295	0.7731

Tabla 3: Resultados de la iteración 3

VI. DISCUSIÓN O ANÁLISIS DE RESULTADOS

A. Iteración 1

Tras analizar los resultados de esta iteración, se aprecia cómo el segundo modelo clasifica correctamente el mayor número de imágenes del conjunto de pruebas (el 71.75 % de dichas imágenes) entre los 6 modelos de esta iteración. Por otro lado, aunque el segundo modelo no es el que presenta el valor más alto de exhaustividad, se ha seleccionado éste debido a que es el único que supera el valor 0.70 tanto en exhaustividad como en exactitud en pruebas.

B. Iteración 2

Debido a que cada modelo de la segunda iteración ha empleado un subconjunto diferente del conjunto original de imágenes, en la tabla 2 se puede apreciar la capacidad predictiva del modelo pre-entrenado 'ResNet50' sobre los diferentes subconjuntos de imágenes utilizados. Se observa cómo el segundo modelo, que emplea la totalidad de las imágenes del conjunto de datos original (745), proporciona el valor más alto en la exactitud en pruebas, a diferencia del séptimo y octavo modelo que obtienen valores más bajos de exactitud al utilizar exclusivamente imágenes desde el plano AXIAL. Además, el segundo modelo presenta un valor más alto de exhaustividad, en comparación con el obtenido en el séptimo y octavo modelo.

C. Iteración 3

Por último, tal y como se puede observar en la tabla 3, el décimo modelo logra alcanzar un valor de exactitud en pruebas de 0.8295, siendo así capaz de clasificar correctamente el 82.95 %

de las imágenes de pruebas. Además, proporciona una exhaustividad de 0.7731, logrando así minimizar aún más el número de falsos negativos tras evaluar el modelo.

VII. CONCLUSIONES

A. Conclusiones

Después de entrenar, evaluar y analizar cada uno de los modelos implementados de carácter experimental, se determina que el problema de clasificación se resuelve con resultados satisfactorios, alcanzando el último modelo de la tercera iteración una exhaustividad de 0.7731 y una exactitud en pruebas de 0.8295. De esta forma, el modelo final logra clasificar correctamente el 82.95 % de las imágenes de secuencias *MRI* del conjunto de pruebas en sus 2 respectivas clases.

B. Líneas de trabajos futuros

Para mejorar los resultados obtenidos en el problema de clasificación, se propone realizar en el futuro alguna de las siguientes tareas:

1. Adquirir más imágenes de secuencias *MRI* pertenecientes a nuevos casos de HSVE u otras enfermedades diagnosticados por expertos. Entrenar estos modelos empleando las arquitecturas de este proyecto u otras más grandes, en caso de que se disponga de mayor capacidad de cómputo.
2. Con la ayuda de un experto en el ámbito médico, analizar e investigar las posibles causas por las que el modelo final implementado en el presente trabajo clasifica incorrectamente ciertas imágenes.
3. Probar a entrenar modelos de CNN que soporten ejemplos formados por varias imágenes de secuencias *MRI* del mismo caso para dotar a la red de mayor información del cerebro en cada ejemplo, permitiéndole discriminar mejor si éste presenta la enfermedad o no.

Referencias

- Gaillard, F. (2023). *Normal brain imaging examples*. radiopaedia.org [data set]. <https://radiopaedia.org/articles/normal-brain-imaging-examples-1?lang=us>. [Accessed 14-May-2023].
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition.
- MayoClinic (2023). *Encefalitis - Síntomas y causas - Mayo Clinic*. <https://www.mayoclinic.org/es-es/diseases-conditions/encephalitis/symptoms-causes/syc-20356136>.
- Muñoz Garrido, J. (2022). *Neurociencia Cognitiva. Tema 4: Técnicas de estudio del sistema nervioso [Material del aula]*. Texto académico del máster. Universidad Internacional de La Rioja (UNIR).
- Orphanet (2015). *Encefalitis por el virus del herpes simple*. <https://t.ly/BeWi9>.
- Radiopaedia (2023). *HSV Cases search*. radiopaedia.org [data set]. https://radiopaedia.org/search?diagnostic_certainty=confirmed_substantiated&lang=us&modality=MRI&page=1&q=HSV&scope=cases&sort=completeness.
- Srinivas, C., Prasad, N., Zakariah, M., Alothaiabi, Y. A., Shaukat, K., Partibane, B., and Awal, H. (2022). Deep transfer learning approaches in performance analysis of brain tumor classification using mri images.
- WHO (2023). *Herpex Simplex Virus*. the world health organization. <https://www.who.int/es/news-room/fact-sheets/detail/herpes-simplex-virus>. [Accessed 29-Apr-2023].
- Xiang, Y., Zeng, C., Liu, B., Tan, W., Wu, J., Hu, X., Han, Y., Luo, Q., Gong, J., Liu, J., and Li, Y. (2021). *Deep Learning-Enabled Identification of Autoimmune Encephalitis on 3D Multi-Sequence MRI*. <https://pubmed.ncbi.nlm.nih.gov/34478565/>.

Apéndice A

Relación de casos: cerebros sanos

A continuación, se especifica la correspondencia directa de las imágenes MRI de los 7 casos de cerebros sanos, descritos y empleados como conjunto de datos en el presente proyecto, con los ejemplos de cerebros normales del listado de casos elaborado y recopilado por Frank Gaillard en el repositorio *Radiopaedia* (Gaillard, 2023).

1. **Caso 1:** las imágenes de este caso fueron capturadas con un escáner de tipo 1.5T. Además, se corresponden con las imágenes publicadas en el ejemplo 1 (example 1) de la sección “routine brain screen protocol” del apartado MRI del listado de cerebros normales recopilados por Frank Gaillard.
2. **Caso 2:** las imágenes MRI de este caso se corresponden con las publicadas en el ejemplo 2 (example 2) de la sección “routine brain screen protocol” del apartado MRI del listado de cerebros normales.
3. **Caso 2:** las imágenes MRI de este caso se corresponden con las publicadas en el ejemplo 2 (example 2) de la sección “routine brain screen protocol” del apartado MRI del listado de cerebros normales.
4. **Caso 3:** las imágenes MRI de este caso se corresponden con las publicadas en el ejemplo 5 (example 5) de la sección “routine brain screen protocol” del apartado MRI del listado de cerebros normales.

5. **Caso 4:** las imágenes MRI de este caso se corresponden con las publicadas en el ejemplo 7 (example 7) de la sección “routine brain screen protocol” del apartado MRI del listado de cerebros normales.
6. **Caso 5:** las imágenes MRI de este caso se corresponden con las publicadas en el ejemplo 1 (example 1) de la sección “MR venography ->contrast enhanced” del apartado MRI del listado de cerebros normales.
7. **Caso 6:** las imágenes MRI de este caso se corresponden con las publicadas en el ejemplo 2 (example 2) de la sección “MR venography ->time-resolved MRV” del apartado MRI del listado de cerebros normales.
8. **Caso 7:** las imágenes MRI de este caso se corresponden con las publicadas en el ejemplo 3 (example 3) de la sección “MR venography ->time-resolved MRV” del apartado MRI del listado de cerebros normales.

Apéndice B

Representación computacional de arquitecturas con keras

A continuación, se proporciona información complementaria para cada una de las arquitecturas utilizadas en el presente trabajo. Para ilustrar la estructura de estas arquitecturas, se hace uso de la representación ofrecida por la librería 'keras', al ser la librería empleada para la implementación de cada modelo.

B.1. Modelo 1 - Arquitectura de CNN construida desde cero

La estructura de esta arquitectura, la cual se puede visualizar en la figura B.1, presenta un total de 12,095,105 parámetros (de los cuales todos son entrenables). Además, se puede apreciar que la capa que vectoriza o une los datos en un vector unidimensional viene identificada por el nombre 'Flatten', mientras que las capas completamente conectadas se identifican con el nombre 'Dense'. Estos son los nombres que ha dotado la librería de 'Keras', integrada en 'TensorFlow', a dichas capas.

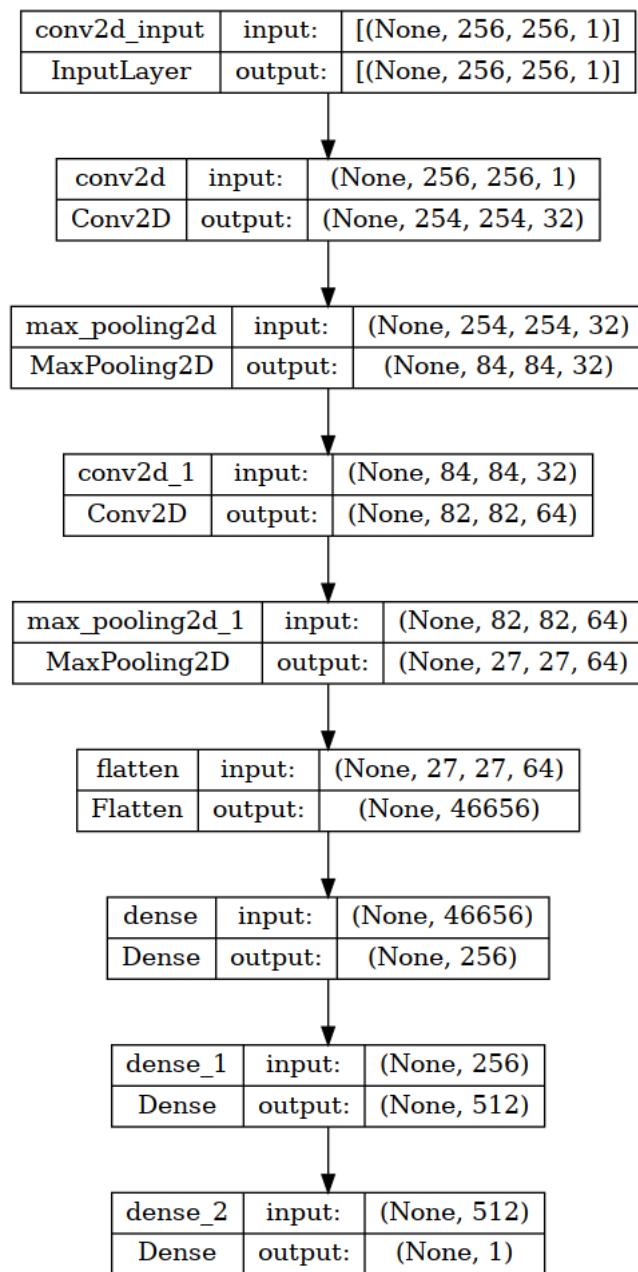


Figura B.1: Arquitectura de modelo de CNN construida desde cero

B.2. Modelo 2 - Transfer Learning con Resnet50

En la segunda arquitectura utilizada, el modelo completo presenta un total de 57,274,527 parámetros, de los cuales 33,686,815 son entrenables, siendo el resto de parámetros no entrenables al tratarse de los parámetros importados que permiten la transferencia de aprendizaje mediante el modelo pre-entrenado 'ResNet50'. En la figura B.2, se ilustra la estructura por capas de este segundo modelo.

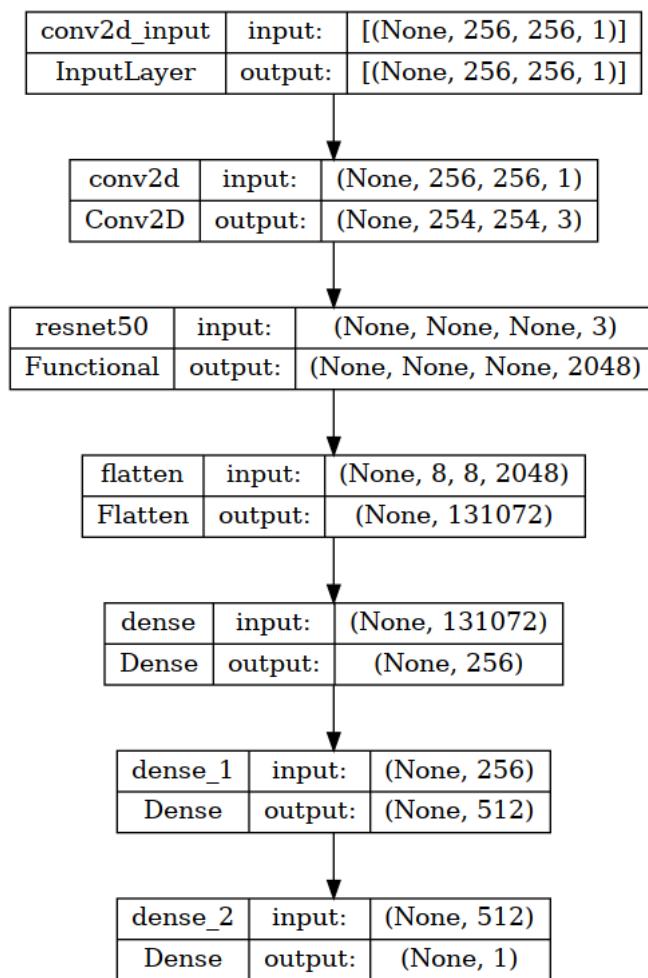


Figura B.2: Arquitectura de modelo de CNN con Transfer Learning de 'ResNet50'

B.3. Modelo 3 - Transfer Learning con VGG16

En esta tercera arquitectura, el modelo completo presenta un total de 21,269,599 parámetros, de los cuales 6,554,911 son entrenables, siendo el resto de parámetros no entrenables al tratarse de los parámetros correspondientes a la transferencia de aprendizaje proporcionada por el modelo pre-entrenado 'VGG16'. En la figura B.3, se ilustra la estructura por capas de este tercer modelo.

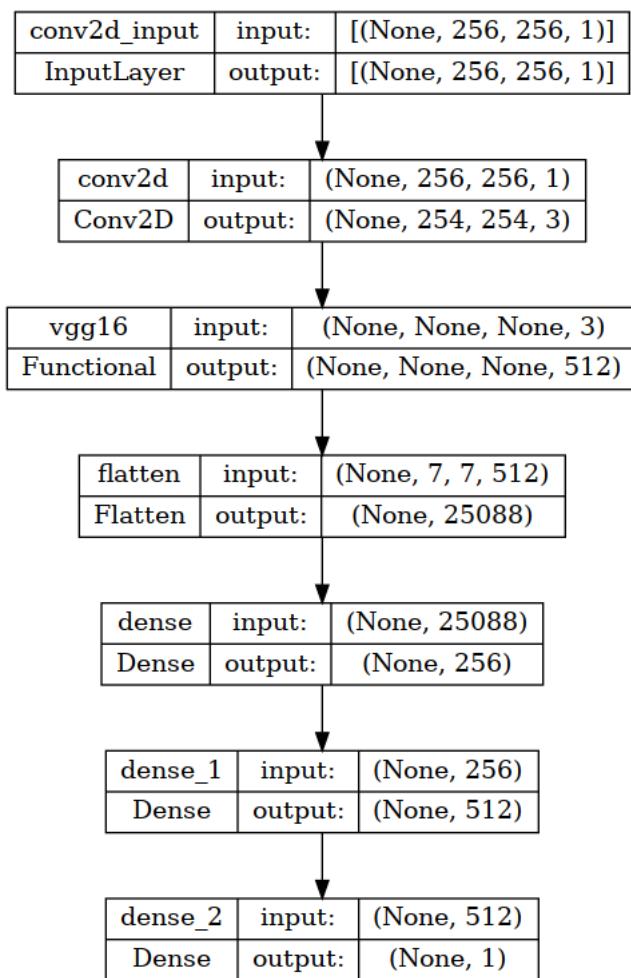


Figura B.3: Arquitectura de modelo de CNN con Transfer Learning de 'VGG16'

B.4. Modelo 4 - Transfer Learning con VGG19

En esta cuarta arquitectura, el modelo completo presenta un total de 26,579,295 parámetros, de los cuales 6,554,911 son entrenables, siendo el resto de parámetros no entrenables al tratarse de los parámetros obtenidos mediante *Transfer Learning* del modelo pre-entrenado 'VGG19'. En la figura B.4, se ilustra la estructura por capas de este cuarto modelo.

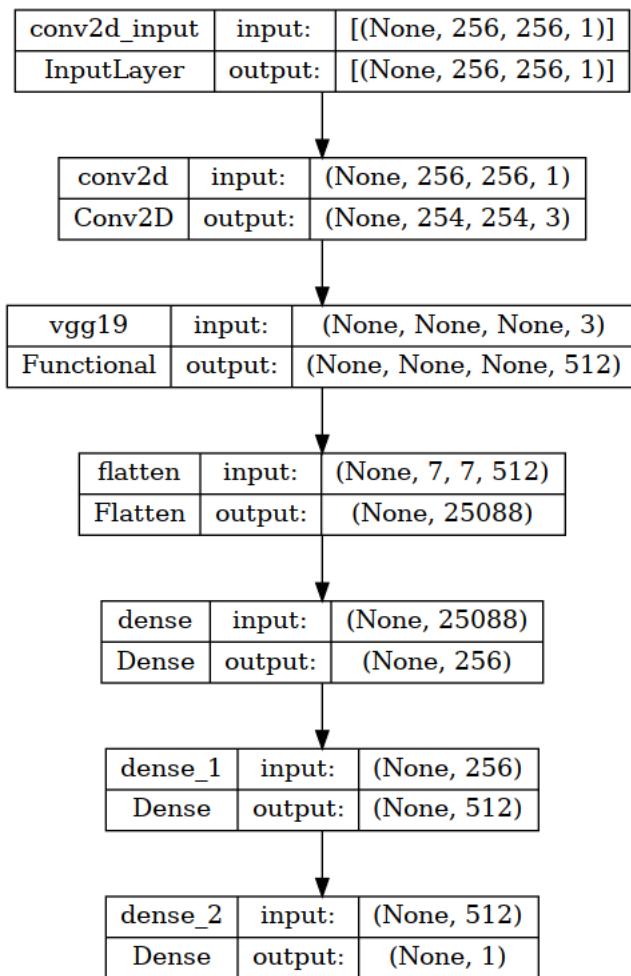


Figura B.4: Arquitectura de modelo de CNN con Transfer Learning de 'VGG19'

B.5. Modelo 5 - Transfer Learning con InceptionV3

En esta quinta arquitectura, el modelo completo presenta un total de 40,809,535 parámetros, de los cuales 19,006,751 son entrenables, siendo el resto de parámetros no entrenables al tratarse de los parámetros obtenidos mediante *Transfer Learning* del modelo pre-entrenado 'InceptionV3'. En la figura B.5, se ilustra la estructura por capas de este quinto modelo.

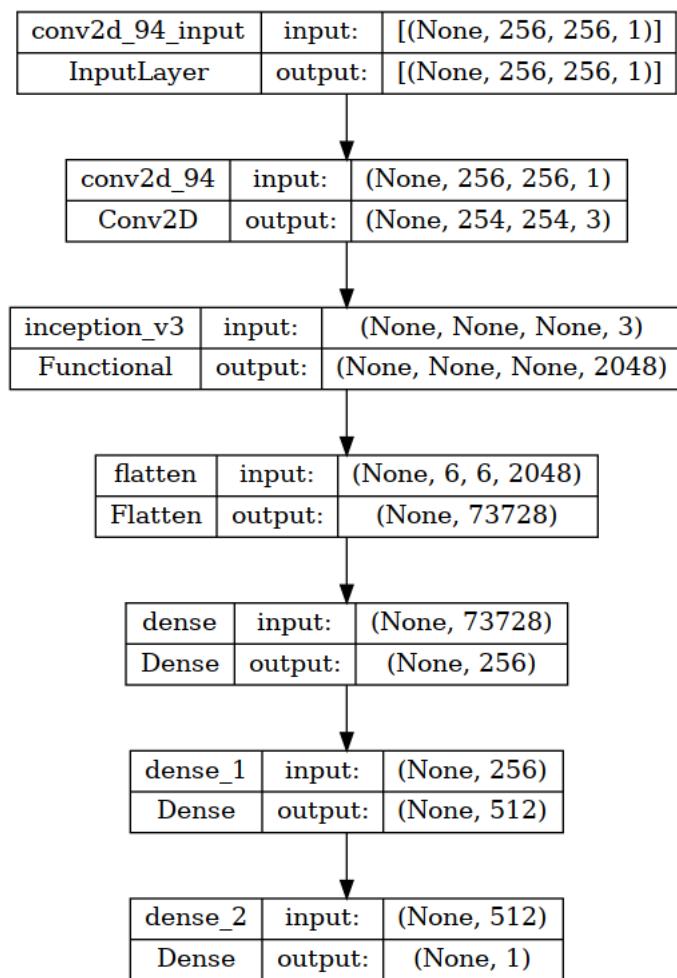


Figura B.5: Arquitectura de modelo de CNN con Transfer Learning de 'InceptionV3'

B.6. Modelo 6 - Transfer Learning con MobileNet

En esta sexta arquitectura, el modelo completo presenta un total de 16,206,303 parámetros, de los cuales 12,977,439 son entrenables, siendo el resto de parámetros no entrenables al tratarse de los parámetros obtenidos mediante *Transfer Learning* del modelo pre-entrenado 'MobileNet'. En la figura B.6, se ilustra la estructura por capas de este sexto modelo.

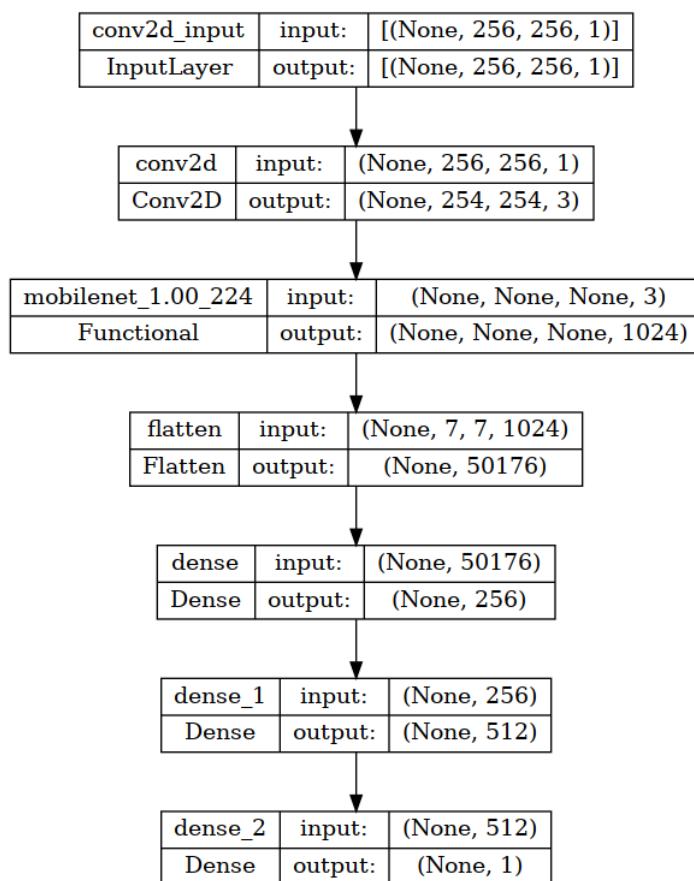


Figura B.6: Arquitectura de modelo de CNN con Transfer Learning de 'MobileNet'

B.7. Modelo 9: *Dropout + Batch Normalization*

En esta última arquitectura, el modelo completo presenta un total de 57,185,439 parámetros, de los cuales 33,596,831 son entrenables, siendo el resto de parámetros no entrenables. En la figura B.7, se ilustra la nueva y última arquitectura de CNN construida en este proyecto.

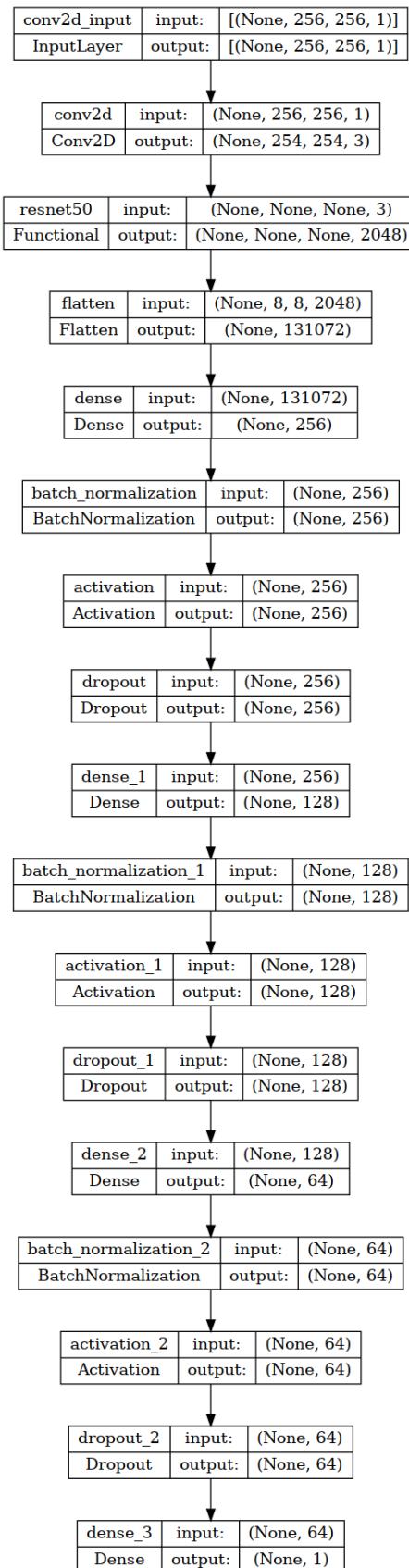


Figura B.7: Arquitectura de CNN con capas de 'Dropout' y 'Batch Normalization'