



ugr | Universidad
de **Granada**

TRABAJO FIN DE GRADO
INGENIERÍA EN INGENIERÍA INFORMÁTICA

Deep Learning para diagnóstico a partir de imágenes Biomédicas

Autor

Francisco Carrillo Pérez

Directores

Luis Javier Herrera Maldonado
Alberto Guillén Perales



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

—
Granada, 12 de Septiembre de 2017

Deep Learning para diagnóstico a partir de imágenes Biomédicas

Francisco Carrillo Pérez (alumno)

Palabras clave: DeepLearning, CNN, Alzheimer

Resumen

La enfermedad de Alzheimer es una de las enfermedades que más afecta a pacientes de edad avanzada en todo el mundo. Su cura se desconoce, por lo que un diagnóstico precoz puede ayudar a mejorar notablemente la vida del paciente. El problema es que la clasificación de la enfermedad en edad temprana es una tarea complicada, además de que se puede confundir con otros deterioros que acaecen en el mismo rango de edad ~~propios de la edad~~. Con los nuevos avances que se han obtenido en el uso de técnicas de Deep Learning en el área de Visión por Computador y clasificación de imágenes, el uso de estas técnicas para la clasificación de pacientes puede suponer una ayuda notable a la hora de que se pueda diagnosticar correctamente si un paciente está comenzando a desarrollar esta enfermedad, con lo que se podría tratar con suficiente tiempo. Con este trabajo, se intenta responder si con estas técnicas podemos realizar una clasificación correcta de imágenes 2D cerebrales de pacientes y de ser así cuáles serían las capas del cerebro más favorables a la hora de realizar esta clasificación.

Deep Learning for diagnosis using Biomedical images

Francisco Carrillo Pérez (student)

Keywords: DeepLearning, CNN, Alzheimer

Abstract

Alzheimer is one of the diseases that affect the most to the elderly in the whole world. The cure is unknown, so an early diagnosis is very important in order to make improvements in the patients life. With this disease we have the problem that an early diagnosis is really complicated, adding the fact that it could be difficult to distinguish it in respect to other elderly brain damages. In the recent years, there has been an active research on Deep Learning techniques in the area of Computer Vision and Image Classification. The use of this techniques for patient diagnosis could be a useful help when deciding if a patient is developing the disease, allowing the application of an appropriate treatment at an early stage. With this work, we are trying to answer if by the usage of these techniques we can make a good classification of 2D brain images of patients and, if so, which are the best slices of the brain to perform this classification.

Yo, **Francisco Carrillo Pérez**, alumno de la titulación Ingeniería Informática de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con 77140580Y, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: Francisco Carrillo Pérez

Granada a 12 de Septiembre de 2017 .

D. Luis Javier Herrera Maldonado, Profesor del Área de Arquitectura y Tecnología de Computadores del Departamento Arquitectura y Tecnología de Computadores de la Universidad de Granada.

D. Alberto Guillén Perales, Profesor del Área de Arquitectura y Tecnología de Computadores del Departamento Arquitectura y Tecnología de Computadores de la Universidad de Granada.

Informan:

Que el presente trabajo, titulado *Deep Learning para diagnóstico a partir de imágenes Biomédicas*, ha sido realizado bajo su supervisión por **Francisco Carrillo Pérez (alumno)**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a 14 de Septiembre de 2017.

Los directores:

D. Luis J. Herrera Maldonado

D. Alberto Guillén Perales

Agradecimientos

En primer lugar me gustaría agradecer el apoyo de mis tutores Luis Javier y Alberto, sin los cuales este trabajo hubiese resultado mucho más arduo y costoso.

También agradecer a mi familia y pareja todo el apoyo dado durante la realización de este trabajo y a lo largo de esta carrera, tanto en los momentos buenos como en los malos.

Por último, dar las gracias a mis compañeros durante estos 4 años, ya que sin ellos la carrera habría sido mucho más aburrida.

Índice general

Introducción	3
1.1. Motivación	3
1.2. Alzheimer	5
1.2.1. Definiciones	6
1.3. Resumen y estructura del proyecto	8
Objetivos	11
Fundamento del proceso implementado	13
3.1. Redes Neuronales Artificiales y Redes Neuronales Profundas	13
3.2. Redes Neuronales Convolucionales	13
3.2.1. Capa convolucional	14
3.2.2. Capa de Pooling	15
3.2.3. Batch Normalization	16
3.2.4. Funciones de activación	17
3.3. Métodos de validación de modelos	19
3.3.1. Validación cruzada en K segmentos	20
3.3.2. Validación cruzada dejando uno fuera	20
Análisis del diagnóstico de la enfermedad de Alzheimer	23
4.1. Análisis de estudios médicos	23
4.2. Análisis de estudios basados en computación	26
Análisis y Desarrollo Experimental	29
5.1. Máquinas en las que se han realizado los experimentos	29
5.2. Herramientas utilizadas	29
5.2.1. Librerías para trabajar con técnicas de Deep Learning	29
5.2.2. Librería para trabajar con las imágenes en formato NIFTI	31
5.3. Procesamiento de las imágenes NIFTI	32
5.4. Arquitectura de la CNN 3.2	32
5.5. Experimentos	33
5.5.1. Parámetros comunes a todos los experimentos	33
5.5.2. Conjunto de datos reducido y completo	33

5.5.3. Primeros experimentos: 5-Fold CV con la capa 55 en un conjunto reducido y el conjunto completo	33
5.5.4. Selección de las mejores capas 2D: 5-Fold CV para ver cuál es la mejor capa	34
5.5.5. Verificación de resultados en LOO: LOOCV por paciente a las mejores capas	34
Base de datos empleada	37
6.1. Post procesamiento	37
6.2. Software de visualización de las imágenes	38
6.3. Separación de las imágenes por tipos	39
6.4. Conjunto reducido y conjunto completo	39
Resultados	41
7.1. Primeros experimentos: 5-Fold CV con la capa 55 en un conjunto reducido y el conjunto completo	41
7.2. Identificación del rendimiento de las diferentes capas 2D de las imágenes MRI	42
7.3. Verificación de resultados en LOO: LOOCV por paciente a las mejores capas	44
Conclusiones y trabajos futuros	47
8.1. Conclusiones	47
8.2. Trabajos futuros	48
Bibliografía	49

Índice de figuras

1.1. Caracterización de MCI que precede al diagnóstico de la enfermedad de Alzheimer [1]	7
3.2. Ejemplo de la arquitectura de una CNN llamada LeNet por Yan Lecun [2]	14
3.3. Ejemplo de filtros aprendidos por parte de Krizhevsky [3] . .	15
3.4. Ejemplo de capa de pool.	16
3.5. Ejemplo de la operación maxpool.	16
3.6. Ejemplo de evolución del error cunado se toma o no, la capa de batch normalization, extraído de [4]	17
3.7. Función de activación Softmax	18
3.8. Función de activación ReLU	19
3.9. Ejemplo de CV con K=4 [5]	20
3.10. Ejemplo de <i>LOOCV</i> [6]	21
4.11. Ejemplo de imagen obtenida por MRI	24
4.12. Ejemplo de imagen obtenida por CAT	24
4.13. Tomografía del cerebro de un paciente con alzheimer mostrando pérdida de la función en el lóbulo temporal.	25
5.14. Interés de búsqueda de Frameworks de Deep Learning. Fuente: François Chollet	31
5.15. Stars a repositorios de Github. Fuente: François Chollet . .	31
5.16. Arquitectura utilizada basada en la arquitectura VGG [7] del artículo [8]	36
6.17. Ejemplo de una imagen con el post-procesamiento relatado en 6.1	38
6.18. Ejemplo de visualización de una imagen con el software MRIcron [9]	39
7.19. Rendimiento en los distintos experimentos, el orden es el mismo que en la tabla 7.1	42
7.20. Rendimiento de las distintas capas	43
7.21. Rendimiento de en los experimentos en los que se utilizó como método de validación LOO CV	44

Índice de cuadros

1.1.	Incidencias de la enfermedad en España respecto a la edad [10]	6
7.2.	Tabla con los resultados obtenidos en distintos experimentos .	42
7.3.	Tabla con los resultados obtenidos por las distintas capas . .	43
7.4.	Tabla con los resultados obtenidos en LOO con las tres capas	44

Introducción

El objetivo de este proyecto es la predicción temprana de la enfermedad de Alzheimer a partir de imágenes biomédicas de pacientes en distintos estados de la enfermedad. Este proceso de predicción se realizará utilizando técnicas que se engloban dentro del aprendizaje profundo o *Deep Learning*.

1.1. Motivación

Actualmente, la enfermedad de Alzheimer es la principal causa de demencia en el ser humano. Lo que se produce es un deterioro progresivo de las células nerviosas cerebrales, lo que desemboca en demencia senil. El aumento de personas afectadas por esta enfermedad viene dado en parte por el aumento de la esperanza de vida, ya que esta enfermedad se da en personas de edad avanzada. Las consecuencias de esta enfermedad son varias, pero todas limitan la calidad de vida de la persona que la padece, además del de las personas de su entorno, ya que esta enfermedad va minando la autonomía del enfermo sobre su propio cuerpo. Esto provocará que el enfermo necesite un apoyo externo para poder realizar su día a día.

Aún no se sabe la causa de la enfermedad de Alzheimer, aunque si se conocen factores de riesgo que pueden ayudar a su desarrollo [11]. Los principales que se conocen son los siguientes: la edad (como se ha explicado antes con el aumento de la esperanza de vida), el sexo (afectando más a mujeres que a hombres), el nivel de educación (las personas con estudios y mayor actividad intelectual son menos propensas a esta enfermedad), los trastornos metabólicos asociados con la resistencia a la insulina, la obesidad, hipertensión o diabetes y la genética (la presencia del genotípico determinado del gen de la EPOE) entre otros factores.

No existe una cura de la enfermedad, solo se conocen tratamientos que ralentizan su avance. Estos tratamientos deben realizarse en las primeras fases de la enfermedad, ya que después pueden resultar perjudiciales para el paciente.

El diagnóstico del Alzheimer se podría dividir en tres partes [12]:

- **Evaluación de estado de ánimo y estado mental:** El estado mental se evalúa para darle al médico una idea general de qué tan bien está funcionando la mente. Este examen da un sentido general sobre si la persona: está consciente de que tiene síntomas, sabe la fecha, la hora, y dónde está ella, puede recordar una lista corta de palabras, seguir instrucciones y hacer cálculos simples. El doctor puede que le pregunte a la persona su dirección, qué año es y quién es el presidente del país. Puede que al individuo se le pida que deletree una palabra al revés, que dibuje un reloj y que copie un diseño. El doctor también puede evaluar el estado de ánimo y el sentimiento de bienestar de la persona para detectar si hay depresión u otra enfermedad que puede causar pérdida de memoria y confusión.
- **Examen físico y exámenes para el diagnóstico:** El doctor hará ciertos procedimientos para evaluar la salud en general de la persona como su dieta, tomar la presión arterial, o escuchar su corazón. Se harán pruebas de sangre y de orina y posiblemente se ordenen otros exámenes de laboratorio. La información que pueden proporcionar estos exámenes puede ayudar a identificar problemas como anemia, diabetes, problemas de los riñones o del hígado, deficiencias de vitaminas, anomalías de la tiroides y problemas del corazón o de los vasos sanguíneos. Todas estas condiciones pueden causar confusión, problemas de memoria u otros síntomas similares a la Demencia.
- **Examen neurológico:** Un doctor o a veces un Neurólogo que se especializa en problemas del cerebro y del sistema nervioso, evaluará muy cuidadosamente a la persona para determinar si hay señales de otro tipo de problema del cerebro que no es Alzheimer. El doctor también va a evaluar los reflejos de la persona, el equilibrio, movimiento de los ojos, lenguaje y sensibilidad. El doctor está buscando señales de derrames pequeños o grandes, enfermedad de Parkinson, tumores cerebrales, acumulación de líquido en el cerebro y otros males que pueden perjudicar la memoria o la capacidad de pensar. El examen neurológico puede incluir un MRI (Imagen por Resonancia Magnética) o CT (tomografía computarizada). Los MRI y CT pueden revelar tumores, evidencia de derrames pequeños o grandes, daño a causa de lesiones por traumas severos de cabeza o acumulación de líquido. Medicare cubre el PET (tomografía por emisión de positrones) como una ayuda para el diagnóstico en ciertos casos .

El análisis de la imagen de resonancia magnética a través del ojo humano resulta sencillo si las alteraciones estructurales son apreciables. Normalmente se utiliza una escala de grises para poder resaltar las diferencias [13]. El

ÍNDICE DE CUADROS

número de bits con el que trabajan estos sistemas suele ser de 16 bits [13] lo cuál da una escala de grises de hasta 65536 niveles, muy superior a los niveles que el ojo puede diferenciar.

Para tener un estudio más exhaustivo del estado del paciente, se aplican una serie de algoritmos de extracción de características para poder extraer las zonas de mayor interés que son las que observa el evaluador para poder determinar si existen indicios de la enfermedad en el paciente. El principal problema es que si nos encontramos en el inicio de la enfermedad, pueden no existir una alteración suficiente para poder diagnosticar la enfermedad por el ojo humano. La motivación de este trabajo es comprobar si nuevas técnicas de clasificación de imágenes, las cuales son las técnicas de Deep Learning, pueden ser usadas a la hora de la clasificación de pacientes entre alzheimer y normales, y obtener unos resultados robustos que nos puedan indicar el futuro que tiene el uso de estas técnicas.

1.2. Alzheimer

La **enfermedad de Alzheimer (EA)** es una enfermedad neurodegenerativa que se manifiesta como deterioro cognitivo y trastornos conductuales. Se caracteriza, en la mayoría de ocasiones, por una perdida de memoria inmediata y de otras capacidades mentales, a medida que mueren las células nerviosas(neuronas) y se atrofian diferentes zonas del cerebro. La enfermedad suele tener una duración media de aproximada después del diagnóstico al paciente de unos 10 años, todo dependiendo por supuesto de la etapa en la que se diagnostique la enfermedad.

La enfermedad de Alzheimer es la forma más común de demencia, es incurable y terminal, y aparece con mayor frecuencia en personas mayores de 65 años de edad [14], aunque en otros casos más raros y extremos puede ser desarrollada a partir de los 40 años.

Los síntomas de la enfermedad como una enfermedad nosológica fueron definidos por primera vez por **Emil Kraepelin** (Neustrelitz, 15 de febrero de 1856 – Múnich, 7 de octubre de 1926), mientras que la neuropatología característica fue observada por primera vez por **Alois Alzheimer** (Marktbreit, 14 de junio de 1864 - Breslavia, 19 de diciembre de 1915) en 1906. Ambos trabajaban en el mismo laboratorio, por lo que se puede considerar que el descubrimiento fue obra de ambos psiquiatras. Pero dado a que Kraepelin daba mucha importancia a encontrar la base neuropatológica de los desordenes psiquiátricos, decidió nombrar la enfermedad Alzheimer en honor a su compañero.

Normalmente, el síntoma inicial es una perdida de la habilidad para adquirir nuevos recuerdos, y esto suele confundirse con actitudes relacionadas con la vejez o el estrés, ya que es en esta etapa de la vida en la que se suele desarrollar la enfermedad. Ante la sospecha de alzhéimer, el diagnóstico se realiza con evaluaciones de conductas cognitivas, así como neuroimágenes, de estar estas disponibles.

A medida que la enfermedad va progresando, aparecen la confusión mental, irratibilidad y agresión, cambios de humor, trastornos del lenguaje, pérdida de la memoria a corto plazo y una predisposición a aislarse a medida que declinan los sentidos del paciente. Finalmente se pierden funciones biológicas, lo que conlleva la muerte.

La causa de la enfermedad permanece desconocida, aunque las últimas investigaciones parecen indicar que se encuentran implicados procesos de tipo priórico[15]. Las investigaciones suelen asociar la enfermedad a la aparición de placas seniles y ovillos neurofibrilares. Hoy en día, los tratamientos que se ofrecen moderados beneficios sintomáticos, pero no hay tratamiento que retrase o detenga el progreso de la enfermedad.

Según el siguiente estudio [10], en España podemos encontrar la siguiente tabla con respecto a la población que padece esta enfermedad:

Edad	Indidencia (nuevos casos) por cada mil personas
65-69	3
70-74	6
75-79	9
80-84	23
85-89	40
90-	69

Cuadro 1.1: Incidencias de la enfermedad en España respecto a la edad [10]

1.2.1. Definiciones

Dentro de la enfermedad, se pueden distinguir 4 tipos de pacientes:

- **AD:** Acrónimo para un paciente que padece la enfermedad de Alzhéimer.
- **LMCI:** Acrónimo para *Late Mild cognitive impairment*. El **Mild**

ÍNDICE DE CUADROS

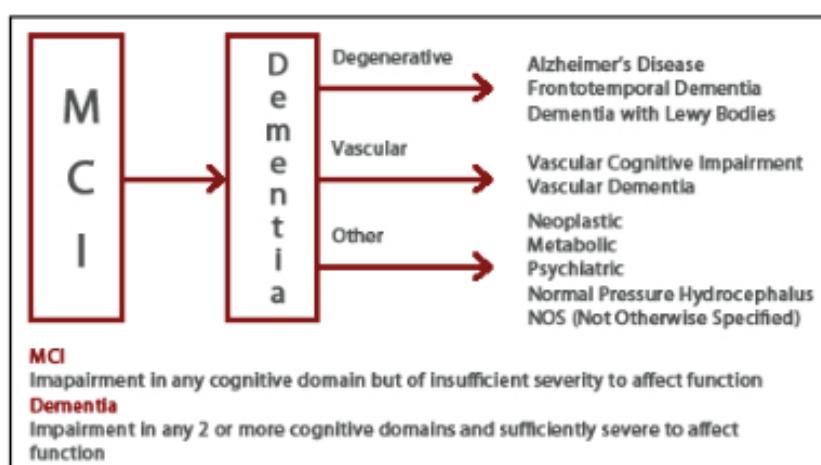
Cognitive Impairment es una entidad nosológica que pretende describir la sintomatología previa al Alzheimer. En este caso, el Late inicial indica que se encuentra en una fase avanzada de este estado previo a poder diagnosticar que el paciente sufre Alzheimer.

- **EMCI:** Acrónimo para *Early Mild cognitive impairment*. En este caso, el Early inicial indica que se encuentra en una fase temprana de este estado previo a poder diagnosticar que el paciente sufre Alzheimer.
- **Normal:** Este término se utiliza para hablar de un paciente el cuál no sufre la enfermedad.

Como se ha explicado previamente, cuanto antes se encuentren señales de un desarrollo de la enfermedad, antes se podrá actuar para el tratamiento de la misma lo cual conlleva una mejor situación de vida para el paciente.

Esto implica que el principal grupo de estudio, es el de los pacientes **MCI**. Las causas del MCI puede ser agrupadas en enfermedades neurológicas, desordenes en el sistema, factores tóxicos o de medicación, factores psiquiátricos o una combinación de los anteriores, como se pueden observar en [16].

Figura 1.1: Caracterización de MCI que precede al diagnóstico de la enfermedad de Alzheimer [1]



Aún así, el grupo de pacientes más difícil de clasificar es este, ya que si se encuentra en la fase de **EMCI**, es difícil diferenciarlo de un paciente normal que presenta síntomas de perdida de tejido cerebral propio de la edad avanzada que de un paciente que está comenzando a padecer la enfermedad de Alzheimer.

1.3. Resumen y estructura del proyecto

Con este proyecto se intenta responder a 2 preguntas principales. La primera pregunta es si con estas técnicas podemos realizar una clasificación correcta de imágenes 2D cerebrales de pacientes y la segunda pregunta es que de ser así, cuáles serían las capas del cerebro más favorables a la hora de realizar esta clasificación.

Para responder a estas preguntas se deben llevar a cabo distintas fases como el procesamiento de las imágenes en 3D para obtener imágenes 2D cerebrales, la elección de una arquitectura de red neuronal, la elección de los distintos métodos de validación y la elección de los experimentos para poder responder a estas dos preguntas.

La estructura del proyecto es la siguiente:

1. **Introducción:** En este caso se ha explicado brevemente la motivación a la hora de la realización de este proyecto además de un breve repaso de la enfermedad de Alzheimer.
2. **Objetivos:** En este capítulo se hablará de los objetivos que se desean conseguir en la realización de este proyecto.
3. **Fundamento del proceso implementado:** En este capítulo se explicará teóricamente la base de las herramientas que se han utilizado para la realización del proyecto y la obtención de sus resultados.
4. **Análisis del diagnóstico de la enfermedad de Alzheimer:** En esta sección se realizará un repaso a la bibliografía existente sobre los factores más notables a la hora de predicción de la enfermedad tanto en el ámbito médico como con el uso de ordenadores.
5. **Análisis y Desarrollo Experimental:** En este capítulo se explicarán cuáles han sido las máquinas en las que se ha realizado el grueso del proyecto, cuáles han sido los elementos comunes a todos los experimentos y cuáles son los experimentos que se han realizado para responder a las preguntas que nos realizábamos al comienzo del mismo.
6. **Base de Datos:** En este capítulo se hablará de la base de datos que ha sido utilizada y cómo está estructurada, además del post procesamiento que han recibido las imágenes de la base de datos que se ha utilizado.
7. **Resultados:** En este capítulo se hablará de los resultados obtenidos en los distintos experimentos.

ÍNDICE DE CUADROS

8. **Conclusiones y Trabajos Futuros:** En este último capítulo se comentarán las conclusiones que se han podido obtener con los experimentos y posibles continuaciones que se desean realizar a partir de lo obtenido en este proyecto.

Objetivos

El objetivo que se persigue con este proyecto es la clasificación de las imágenes de pacientes entre Alzheimer y Normal utilizando imágenes en 2 dimensiones, y determinar qué capas del eje z son las que ofrecen mejor rendimiento al aplicar las técnicas escogidas.

El estudio que se realiza puede dividirse en los siguientes objetivos específicos:

- **Análisis y tratamiento de las imágenes:** Las imágenes de la base de datos se encuentran en formato NiFTI. Para su tratamiento con los algoritmos, debemos convertirlas a un formato con el que estos puedan tratar. Con esto se creará la base de datos para su uso en los distintos experimentos.
- **Análisis y selección de la arquitectura de la Convolutional Neural Network:** se deberá elegir una arquitectura para la red neuronal basándonos en la literatura actual. Esta arquitectura debe tratar con imágenes en dos dimensiones, debido a la potencia de computación que exigen las redes tridimensionales.
- **Optimización de la red neuronal y selección métodos de validación:** Entrenamiento de la red neuronal con la base de datos que se ha obtenido previamente. Para ello se utilizarán diversos métodos de validación, de forma que los resultados obtenidos puedan considerarse significativos. Además, se deberán realizar una serie de experimentos para determinar qué capas del cerebro son las más significativas a la hora de clasificar entre normal y alzheimer.
- **Análisis de los resultados:** En este punto, se analizarán los resultados obtenidos, donde se podrá observar si las técnicas aplicadas tienen un buen rendimiento con respecto a la literatura que se puede encontrar, y cuál es la capa del cerebro que mejor rendimiento demuestra.

Fundamento del proceso implementado

3.1. Redes Neuronales Artificiales y Redes Neuronales Profundas

Las redes neuronales artificiales son un intento de imitar el comportamiento de las redes neuronales biológicas en nuestro cerebro a la hora de aprender sobre un fenómeno concreto. Cada unidad neuronal está conectada con muchas otras y los enlaces entre ellas pueden incrementar o inhibir el estado de activación de las neuronas adyacentes. Son un modelo computacional, el cuál se puede utilizar para distintas tareas. Estas redes neuronales pueden estar compuestas por distintas capas. Cada neurona tiene asociada una función de activación (normalmente la misma función para toda la capa) que es la que indicará si la neurona se activa dependiendo de la entrada correspondiente y nos dará un valor de salida. El valor de entrada a la neurona viene dado por unos pesos, que son los parámetros que se aprenden cuando entrenamos la red neuronal. Los hiperparámetros de una red se corresponden con todos los demás aspectos del diseño de la misma, como el número de capas, tipo de función de activación, número de neuronas por capa, etc.

Las redes neuronales son la base de las técnicas utilizadas para la realización de este TFG, que son las Redes Neuronales Convolucionales.

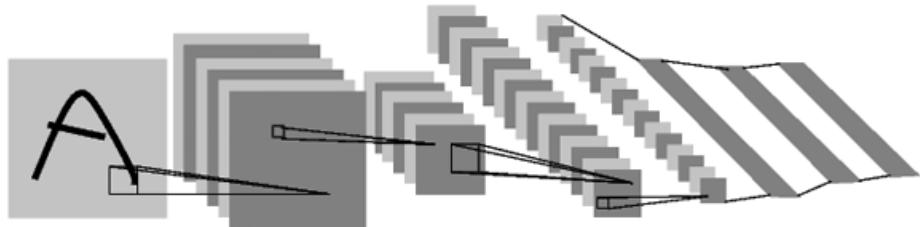
3.2. Redes Neuronales Convolucionales

Las Redes Neuronales Convolucionales, o *Convolutional Neural Networks (CNN)*, son un tipo de redes neuronales profundas que han obtenido resultados más que notables en principalmente en el campo de Visión por Computador, aunque similarmente en otras áreas como el procesamiento de sonido, etc. [17][2]

Las *CNN* se componen de una capa de entrada y una de salida además de múltiples capas ocultas , o *hidden layers*. Dentro de estas capas ocul-

tas podemos encontrar diversos tipos, pero las principales son las capas de Convolución, y las capas de *Pooling*.

Figura 3.2: Ejemplo de la arquitectura de una CNN llamada LeNet por Yan Lecun [2]



LeCun et al. 1998

3.2.1. Capa convolucional

En estas capas se aplica una operación de convolución a la entrada y el resultado es pasado a la siguiente capa. La operación de convolución emula la respuesta de una neurona individual ante un estímulo visual. Los parámetros de una capa convolucional se componen de una serie de filtros que se aprenden. Cada filtro es pequeño, pero se extiende a lo largo de toda la profundidad de la entrada, en este caso una imagen. Esto quiere decir que si el filtro es 5×5 y la imagen es de 50×50 , iremos pasando el filtro por cuadrados de 5×5 de la imagen hasta haberla recorrido en su totalidad.

Entonces, la red aprenderá filtros que se activarán cuando vean cierto tipo de característica, como por ejemplo una esquina..

ÍNDICE DE CUADROS

Figura 3.3: Ejemplo de filtros aprendidos por parte de Krizhevsky [3]



3.2.2. Capa de Pooling

Es común insertar una capa de *pooling* entre sucesivas capas convolucionales. La función de una capa de *pooling* es reducir progresivamente el tamaño espacial de la representación para reducir la cantidad de parámetros y la computación en la red, y ayuda también para el control del sobreajuste, o overfitting. El sobreajuste se produce cuando se aprenden unos parámetros en el entrenamiento que son tan específicos del conjunto de entrenamiento que la red falla en la clasificación al llegar nuevos ejemplos. Se tienen por tanto resultados muy dispares entre los conjuntos de entrenamiento y el de *test*. Una operación común que se selecciona para la capa de *Pooling* es la de MAX, con esta lo que nos hacemos es quedar con el valor máximo de los que se encuentran en el tamaño de la capa de *Pooling* (si por ejemplo la ventana de la capa de *Pooling* es de 2×2 nos quedaremos con el máximo valor de los 4 que acapara la ventana y esto acaba causando la reducción del tamaño de la información que se pasa a la siguiente capa de la red).

Un ejemplo. Si tenemos una imagen de 224×224 y utilizamos una capa de *pooling* de 2×2 , a la salida de esa capa, nuestra imagen, o mejor dicho la representación de la misma dentro de la red, tendrá un tamaño de 112×112 , ya que la hemos reducido a la mitad con la operación de *Pooling*. Con la operación de MAX, lo que hacemos es que nos quedamos con el valor del máximo dentro de un grupo de píxeles.

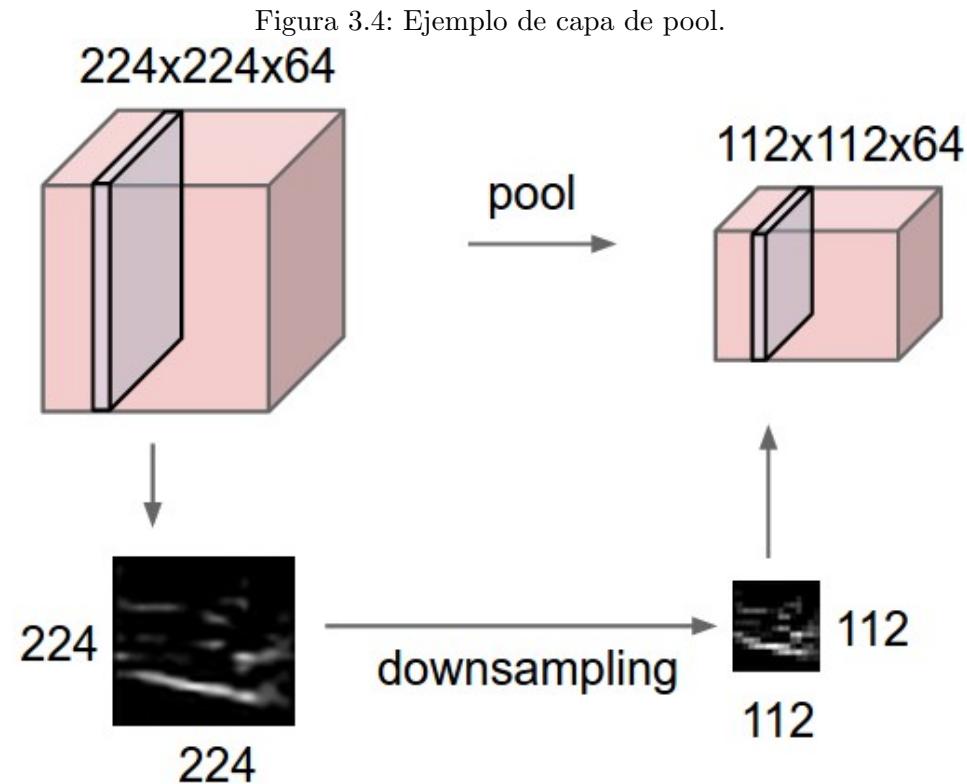
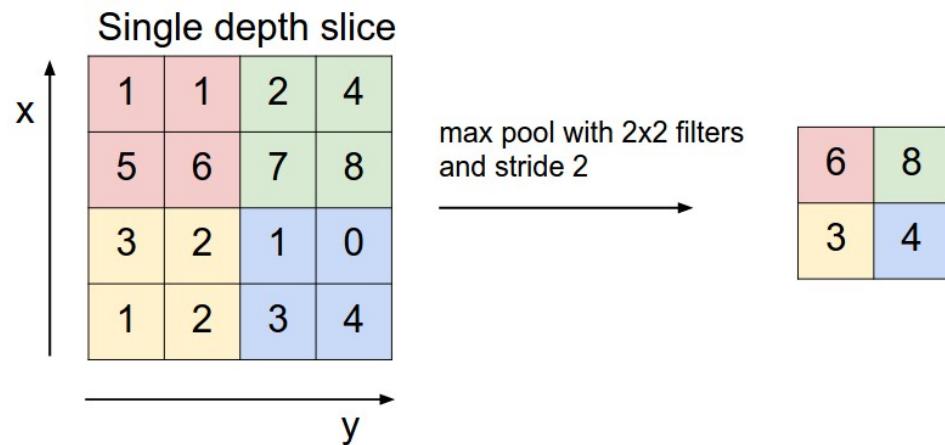


Figura 3.5: Ejemplo de la operación maxpool.



3.2.3. Batch Normalization

Hay un tipo de capa que se utiliza ampliamente en la bibliografía y es la de *batch normalization* [18], que se podría traducir como normalización por lotes, pero se suele conocer exclusivamente por su término en inglés. Con

ÍNDICE DE CUADROS

esta capa lo que se consigue es forzar las activaciones a través de la red a tomar una distribución gaussiana al comienzo del entrenamiento. Normalmente, esta capa se sitúa después de la capa completamente conectada.

Se ha observado que las redes que utilizan este tipo de capas son más robustas a la hora de malas inicializaciones de los pesos.

Figura 3.6: Ejemplo de evolución del error cuando se toma o no la capa de batch normalization, extraído de [4]

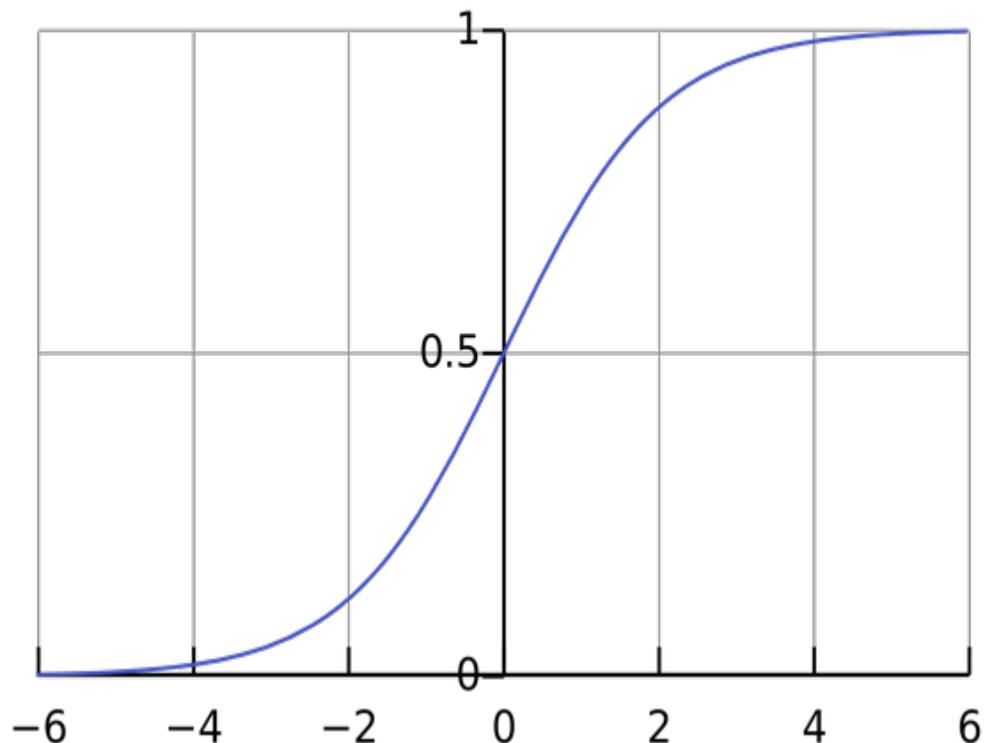


3.2.4. Funciones de activación

Como se ha explicado en la sección de las redes neuronales 3.1, podemos tener distintas funciones de activación. Estas, definen la salida de la neurona dependiendo de cuál ha sido la entrada.

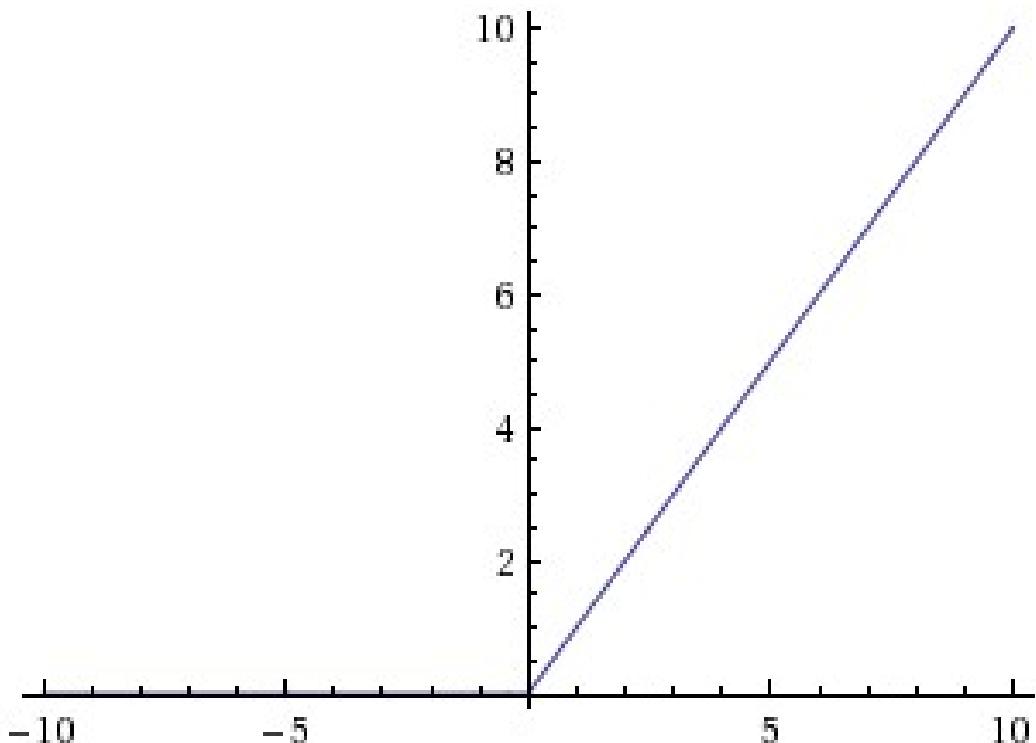
En nuestro trabajo usaremos dos funciones de activación. Una es la función conocida como ReLU(*Rectified Linear Unit*) y la otra es la función conocida como Softmax.

Figura 3.7: Función de activación Softmax



ÍNDICE DE CUADROS

Figura 3.8: Función de activación ReLU



La función de activación Softmax es la que se utilizará a la salida de la CNN para que nos de un valor de predicción. En el caso de que este valor sea 0, se clasificará la imagen como AD, y en el caso de que sea 1 se clasificará como Normal.

Y la función de ReLU la usaremos como función de activación de las capas convolucionales.

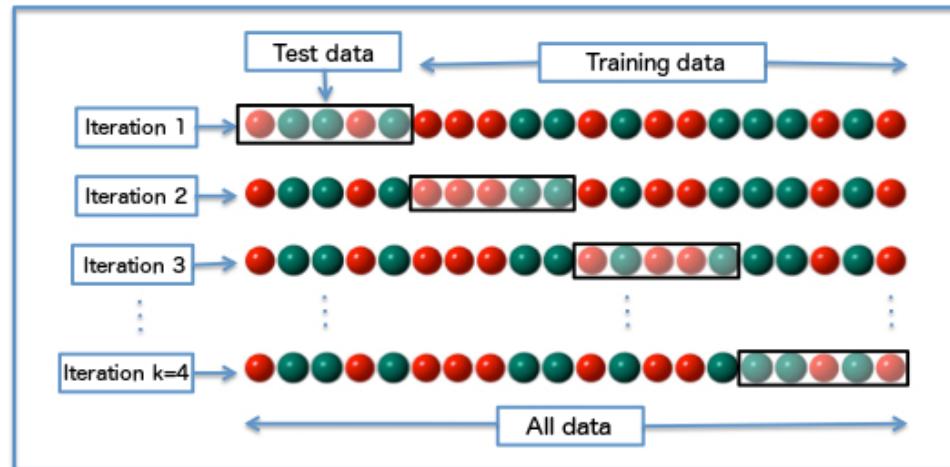
3.3. Métodos de validación de modelos

Para la validación del modelo empleado se han utilizado dos métodos, los cuales son Validación cruzada en K segmentos, o *K-Fold cross-validation*, y la validación cruzada dejando uno fuera, o *Leave-one-out cross-validation*, que a continuación serán explicados brevemente.

3.3.1. Validación cruzada en K segmentos

En la validación cruzada en K segmentos, o *K-Fold Cross-Validation* en su traducción al inglés, se realiza una partición aleatoria del conjunto de datos en K partes. Entonces, se utilizará uno de estos subconjuntos para *test* y los $k-1$ subconjuntos restantes se utilizarán para entrenamiento. Esto se repite K veces, tomando cada vez uno de los subconjuntos como *test*, y el resto como *entrenamiento*. Entonces, se obtiene una media de los resultados en *test* y con esto una medida del funcionamiento esperado del modelo para datos no vistos. Esto incrementa el tiempo de entrenamiento y comprobación, por lo que no es un método muy usado en los conjuntos de datos cuyo tamaño es muy grande, pero es de gran ayuda en conjuntos de datos más pequeños, pues como se ha dicho nos permite validar mejor como será la capacidad de generalización del modelo.

Figura 3.9: Ejemplo de CV con $K=4$ [5]

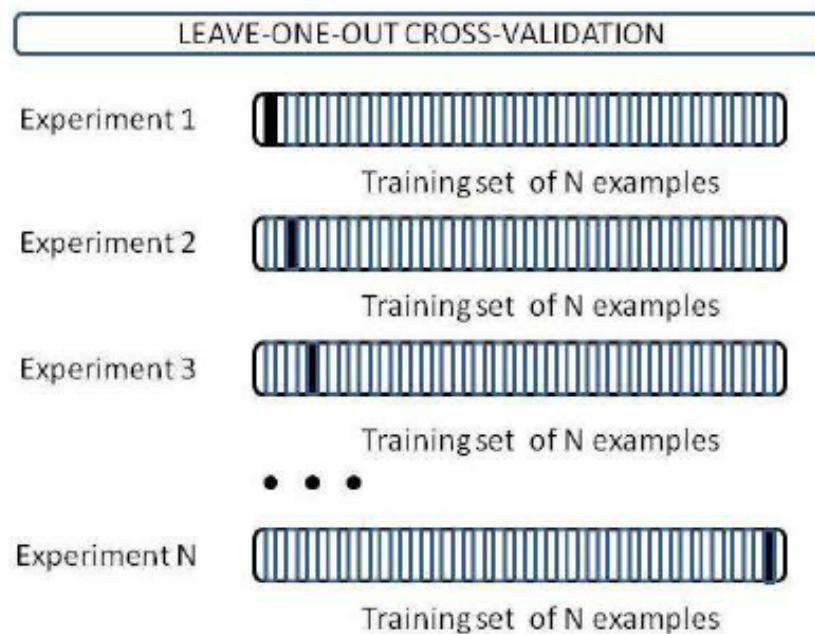


3.3.2. Validación cruzada dejando uno fuera

La validación cruzada dejando uno fuera, o *Leave-one-out Cross Validation (LOOCV)* en su traducción al inglés, podría ser un tipo de *K-Fold CV* en el que K es igual al número de elementos dentro del conjunto de N datos. Esto implica que el tiempo de validación sea muy alto, ya que vamos a realizar N entrenamientos y N evaluaciones, pero es un método muy robusto para la validación del modelo.

ÍNDICE DE CUADROS

Figura 3.10: Ejemplo de *LOOCV* [6]



Análisis del diagnóstico de la enfermedad de Alzheimer

4.1. Análisis de estudios médicos

En la última década, el uso de neuroimágenes para el diagnóstico de un paciente ha crecido exponencialmente. Hace solo 10 años los parámetros de la *American Academy of Neurology* indicaban que las tomografías computarizada (TC/CT) y las resonancias magnéticas (MR) eran opciones de examinación optionales. Sin embargo, se tienen bastantes pruebas para afirmar que los análisis de cambios estructurales y funcionales pueden tener un impacto en el proceso de tratamiento de un paciente. [19].

Actualmente, las recomendaciones de la *American Acafemy of Neurology* para el estudio de la demencia incluyen una examinación por imagen, una tomografía axial computarizada (TAC/CAT) o un MRI craneal. Aunque el TAC es una prueba más rápida, barata y accesible, el MRI tiene unas ventajas las cuáles son una mejor resolución y contraste entre los tejidos y puede detectar anomalías focales existentes y no utilizada radiaciones ionizantes. Además, el MRI puede llegar a ser la técnica elegida en los estudios relegando al TAC en casos en los que las causas secundarias de la demencia deben ser descartadas rápido. En las Figura 4.1 y Figura 4.1 podemos observar un ejemplo de una imagen obtenida por MRI y otra por TAC.

La perdida neuronal y consecuentemente la atrofia del cerebro causa un ensanchamiento de los surcos, un estrechamiento de los giros cerebrales y una dilatación de los ventrículos con una reducción significativa del peso del cerebro. En años recientes, se han identificado diferentes distribuciones de atrofia dependiendo del tipo de demencia. Para el Alzheimer, los efectos de la atrofia afectan al promedio del lóbulo temporal. De todas formas, las medidas de la atrofia en general no parecen que sean útiles para el diagnóstico del Alzheimer ya que la mayoría de enfermedades neurodegenerativas causan una atrofia global similar.

Figura 4.11: Ejemplo de imagen obtenida por MRI

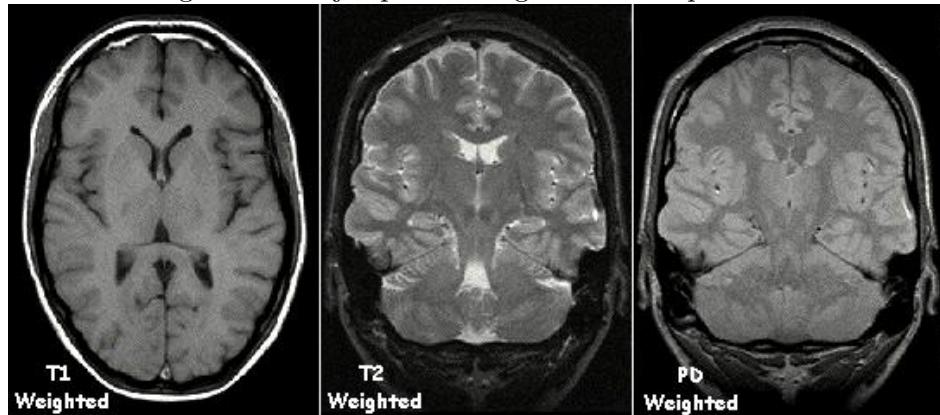
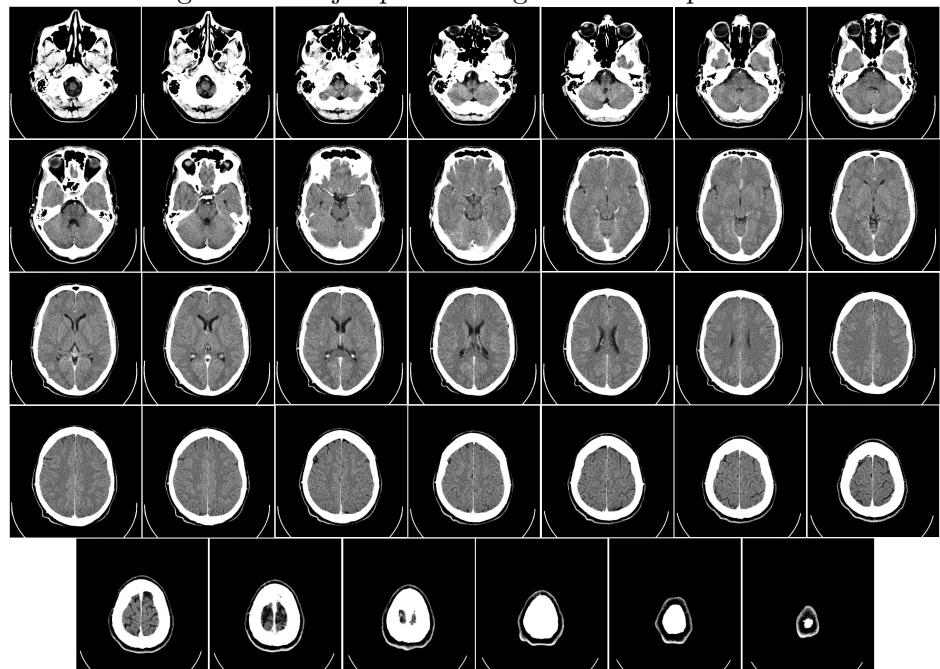


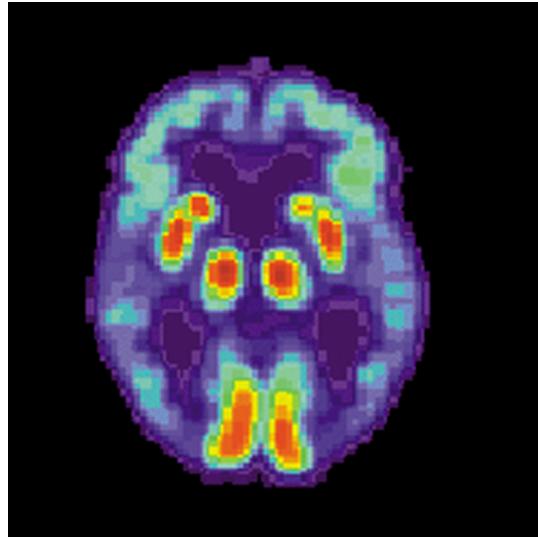
Figura 4.12: Ejemplo de imagen obtenida por CAT



ÍNDICE DE CUADROS

En la siguiente tomografía del cerebro de un paciente se puede observar la perdida de la función del lóbulo temporal:

Figura 4.13: Tomografía del cerebro de un paciente con alzheimer mostrando pérdida de la función en el lóbulo temporal.



Muchos estudios se han enfocado en cuantificar cuál es la cantidad de atrofia en el lóbulo temporal, e incluso existen escalas para cuantificar cuál es el grado de atrofia, las cuales son rápidas y fáciles de usar. En la última década, también se han desarrollado métodos automáticos para medir el grado medio de atrofia en el lóbulo temporal [20], y esto se ha utilizado para la predicción de pacientes que padecen la enfermedad [21].

En paralelo, también se ha observado que los pacientes de Alzheimer una notable atrofia en regiones posteriores conocida como atrofia cortical. Esto parece ser característico, pero no exclusivo, de la enfermedad de Alzheimer. Es por ello, que en la última época, se han intentado proponer escalas visuales para cuantificar el grado de atrofia cortical [22].

Recientemente también se han realizado estudios sobre la diferencia de la atrofia en el hipocampo y cortical hipo-metabolismo de personas que no padecen la enfermedad y las que sí, pudiendo llevar a mejores diagnósticos. [23].

También se ha podido observar que la atrofia cerebral comienza años antes del diagnóstico del Alzheimer. De hecho, la atrofia del hipocampo aparece en pacientes con riesgo de padecer Alzheimer años antes de que los síntomas comiencen, como se puede observar en [24].

4.2. Análisis de estudios basados en computación

Por otra parte, el estudio de la enfermedad basándose en el uso del análisis de las imágenes MRI mediante algoritmos de aprendizaje automático, o *Machine Learning* es un campo que en la última década ha cobrado mucha fuerza a la hora del diagnóstico de la enfermedad. En [25] podemos observar el uso de la actividad eléctrica producida por el cerebro y características de las imágenes para entrenar un algoritmo SVM y también el uso de redes neuronales con mapas auto organizados de características (SOM) con el fin de clasificar entre cerebros normales o anormales, obteniendo unos buenos resultados de rendimiento, o *accuracy*.

Estudios siguientes continuaban la misma línea de trabajo, usando como características actividad eléctrica producida por el cerebro y luego utilizando algún algoritmo de Aprendizaje Automático como clasificador. Además, también se incluyeron métodos como el Análisis del Componente Principal, o PCA (*Principal Component Analysis*) para reducir el tamaño dimensional de los datos. [26]

Dentro de nuestra propia universidad, tenemos trabajos que demuestran que el uso de SVM con las apropiadas características de ondas y de las propias imágenes además del uso de PCA para la reducción de la dimensión, permite obtener unos buenos resultados a la hora de la clasificación, como se puede comprobar en [27].

Hoy en día, con los avances que se han producido con la introducción de nuevos algoritmos de Deep Learning además del avance en las técnicas de adquisición de imágenes, muchos estudios han intentado utilizar estas técnicas para mejorar la clasificación de los pacientes.

La mayoría de estos estudios hacen uso de redes convolucionales con muchas capas para el entrenamiento y aprendizaje de filtros que puedan ser útiles a la hora de clasificar a los pacientes. Entre ellos cabe destacar el siguiente estudio [8], en el que se utilizaban imágenes MRI en 3D junto con redes convolucionales tridimensionales para el aprendizaje y clasificación de pacientes. En este se obtienen unos resultados muy destacables, partiendo de una base de datos muy reducida.

También se pueden encontrar estudios en los cuáles el uso de redes convolucionales junto con el uso de *autoencoders*, se obtienen buenos resultados en el *accuracy* a la hora de la clasificación de los pacientes, como se puede observar en [28].

Como se puede observar, la mayoría de estudios realizados hacen usos

ÍNDICE DE CUADROS

de redes convolucionales en tres dimensiones, lo que produce que se necesite mucha capacidad de computación y tiempo para obtener los resultados del entrenamiento. Este es un campo en el que aún queda mucho desarrollo, con el objetivo de mejorar el rendimiento de la clasificación, en la implantación de sistemas de ayuda a la decisión para los médicos, y el *Deep Learning* es la metodología más prometedora para el análisis de este tipo de datos hoy en día.

Análisis y Desarrollo Experimental

En este capítulo voy a hacer una recapitulación de la estructura y desarrollo de los experimentos que se han llevado a cabo para la realización de este TFG, además de una breve explicación de las distintas herramientas que se han utilizado para ello.

5.1. Máquinas en las que se han realizado los experimentos

Las dos máquinas principales en las que se han realizado los experimentos han sido las siguientes:

- **MacBook Air 2016:** *Procesador, 1,6 GHz Intel Core i5. Memoria, 8 GB 1600 MHz DDR3.* Principalmente utilizado para el procesamiento de las imágenes NIFTI.
- **Servidor:** *Procesadores, 2 Intel(R) Xeon(R) CPU E5645 @ 2.40GHz. GPU, 2 Tesla C2050 / C2075.* Utilizada para la realización de los experimentos.

5.2. Herramientas utilizadas

En primer lugar, quiero citar cuáles han sido las herramientas utilizadas para la realización de este trabajo.

5.2.1. Librerías para trabajar con técnicas de Deep Learning

Actualmente debido a los grandes avances que se han conseguido en ciertos campos con el uso de técnicas de Deep Learning, el número de librerías dedicadas a facilitar la realización de experimentos con este tipo de técnicas

ha crecido a una gran velocidad, dejándonos muchas opciones donde elegir.

Existe un abanico muy amplio de librerías donde elegir, entre ellas hay librerías que llevan más tiempo como por ejemplo **Caffe**[29] o **Theano**[30] y otras cuyo tiempo de vida es menor. En este último grupo encontramos una librería que en los últimos años ha tenido mucha repercusión, la cuál es **Tensorflow**[31]. Su gran crecimiento puede deberse a varios factores, entre ellos su rapidez (dentro de lo posible) en tiempos de computación, y que viene de la mano de Google.

La mayoría de estas librerías requiere mucho conocimiento técnico de lo que se está haciendo y controlar a mano muchos parámetros que pueden suponernos un retraso a la hora de la realización de los experimentos, debido a necesitar tiempo para arreglar fallos dentro del código.

Es por estos motivos por los que François Chollet, un trabajador de Google, creó la *high-level API Keras*[32]. Keras, es una librería que nos sirve como capa para la utilización tanto de Tensorflow como Theano. Nos permite, por ejemplo, métodos automáticos para el entrenamiento utilizando un número fijado de *batch*, mientras que en Tensorflow o Theano tendríamos que programar esto a mano. Este tipo de ayudas permite que el tiempo que se requiere desde el momento en el que comienzas a programar un experimento hasta el fin de su ejecución sea el más rápido posible y lo más ausente de fallos.

Es por estas, y más razones, que se ha decidido utilizar **Keras** como librería para trabajar con técnicas de Deep Learning.

ÍNDICE DE CUADROS

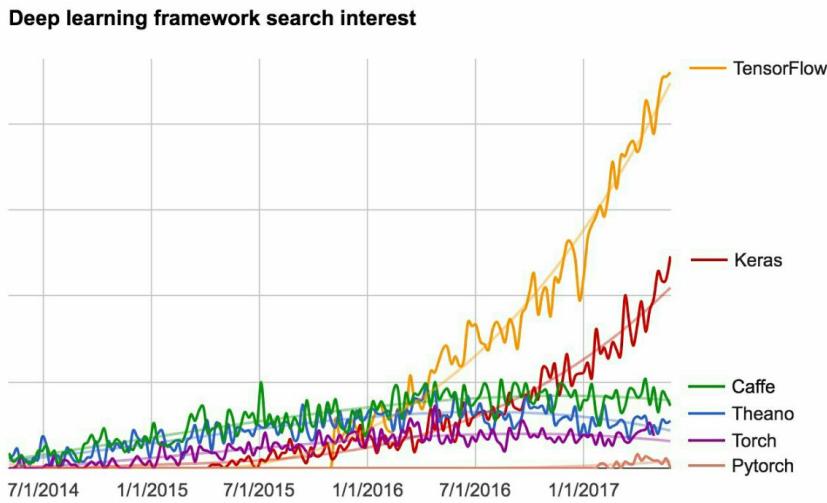


Figura 5.14: Interés de búsqueda de Frameworks de Deep Learning. Fuente: François Chollet

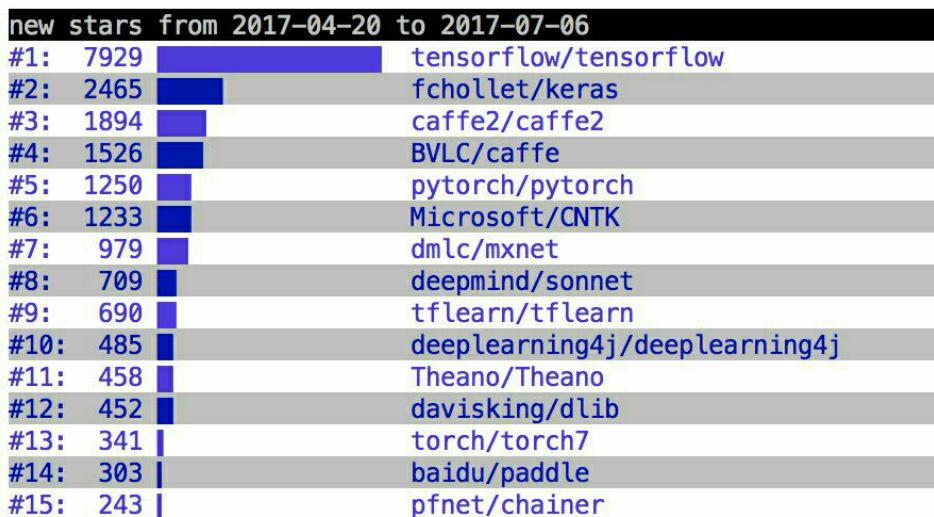


Figura 5.15: Stars a repositorios de Github. Fuente: François Chollet

5.2.2. Librería para trabajar con las imágenes en formato NIFTI

Para poder trabajar con redes convolucionales en 2 dimensiones, se requería una librería con la que poder obtener las capas del cerebro que nosotros quisiésemos y luego guardarlas en un formato con el que Keras pudiese

tratar.

Para ello, se utilizó la librería para Python **Nibabel**[33]. Esta librería nos permite obtener toda la información de un archivo NIFTI y poder coger el **slice** que queramos.

5.3. Procesamiento de las imágenes NIFTI

En nuestro caso, se requería previamente obtener imágenes 2D a partir de los archivos NIFTI. Esto era necesario a la hora de poder tratar estas imágenes con Keras, para la realización de los experimentos. Para ello, se creó un script llamado *create_slices.py*.

Con este script lo que se consigue es la lectura de todos los archivos NIFTI, obteniéndose las capas desde la 10 a la 100 de 5 en 5, y guardándose ordenadamente en una nueva carpeta correspondiente a cada capa, obteniendo 19 carpetas distintas.

Se guardan en esta estructura para luego poder leerlos ordenadamente y obtener resultados sobre cada capa del cerebro. Con esto se nos permitirá determinar cuál es la capa del cerebro que mejor resultados nos da.

5.4. Arquitectura de la CNN 3.2

Esta es una decisión muy importante a la hora de la realización del TFG. Para ello se revisó la bibliografía más reciente, para partir de una base sólida a la hora de la elección del modelo. En la revisión bibliográfica podemos encontrar estudios como [8] y [34].

Por parecernos más concluyentes los resultados de [8] debido a la metodología de separación de pacientes que ellos utilizaban, se decidió utilizar la primera arquitectura que ellos introducen, basada en la arquitectura **VGG** [7].

La arquitectura podemos observarla en la Figura 5.16. En nuestro caso, se ha modificado para que sea una red que acepte entradas 2D en lugar de 3D.

En el artículo [8] también se prueba otro tipo de arquitectura, un tipo de arquitectura residual. Unas pruebas iniciales descartaron dicha arquitectura

ÍNDICE DE CUADROS

debido a que no mejoraban los resultados obtenidos por la primera, aumentando sin embargo el tiempo de cómputo considerablemente. Por reducir densidad de resultados en el trabajo, no se incluyeron dichos resultados en la memoria.

5.5. Experimentos

En esta sección, se irán explicando los distintos experimentos que se han llevado a cabo para el estudio de métodos para la clasificación de las imágenes en 2D y cuál sería la mejor estrategia para el análisis de las mismas.

5.5.1. Parámetros comunes a todos los experimentos

Hay unos parámetros comunes que se han utilizado para todos los experimentos, los cuales se relatan a continuación.

Con respecto a la arquitectura, se ha utilizado como función de pérdida, o *loss* la función **categorical cross-entropy** y un *learning rate* de 27×10^{-6} .

Además, en cuanto al entrenamiento se refiere, se ha utilizado un número de *epochs* igual a 70 y un *batch size* de 10.

5.5.2. Conjunto de datos reducido y completo

Con el conjunto de datos reducido se refiere a que solo se utilizó una imagen por paciente. Esto es debido a que de un paciente hay múltiples imágenes de distintos años.

En cuanto a conjunto completo se refiere a que se utilizaron todas las imágenes de un paciente.

5.5.3. Primeros experimentos: 5-Fold CV con la capa 55 en un conjunto reducido y el conjunto completo

Al inicio comenzamos trabajando con una sola capa de todas las imágenes y se decidió utilizar la capa media del cerebro en el eje Z. En nuestro caso, al ser las imágenes de **110x110** en 2D y **110x110x110** en 3D, la capa media del cerebro en el eje Z es la capa 55. Como tenemos un conjunto de datos reducido, ya que solo disponemos de y este número en términos de usar técnicas de Deep Learning es tener un conjunto de datos muy pequeño,

decidimos usar técnicas de validación para poder determinar si los resultados obtenidos eran significativos o no. Para intentar emular los resultados del artículo [8] previamente citado, se construyó un conjunto de imágenes reducido al que llamaremos conjunto reducido. Para ello se implementó un *5-Fold CV*, ya que es lo que se utilizó en [8] y además nos da una robustez en los resultados.

Después, se realizó el mismo experimento utilizando todas las imágenes de un mismo paciente, pero siempre respetando que no se repitiese un mismo paciente en *training* y en *test*.

Por último, se realizó también un *5-Fold CV* pero repitiendo imágenes del mismo paciente en distintos conjuntos, para comprobar que se producía filtración de información entre imágenes de un mismo paciente.

5.5.4. Selección de las mejores capas 2D: 5-Fold CV para ver cuál es la mejor capa

Una vez realizados los experimentos previos decidimos responder a la pregunta, ¿cuál será la capa del cerebro con la que se obtienen mejores resultados?. Por lo tanto, se realizó un *5-Fold Cross Validation* para comprobar cuál es la capa que mayor rendimiento obtiene en cuanto a *accuracy* en el conjunto de *test*.

La separación de los pacientes se realizó manualmente, para asegurarnos de que no hubiese imágenes del mismo paciente en distintos conjuntos. Esto nos dejó con 4 conjuntos compuestos pro 77 imágenes y el último compuesto por 76 imágenes.

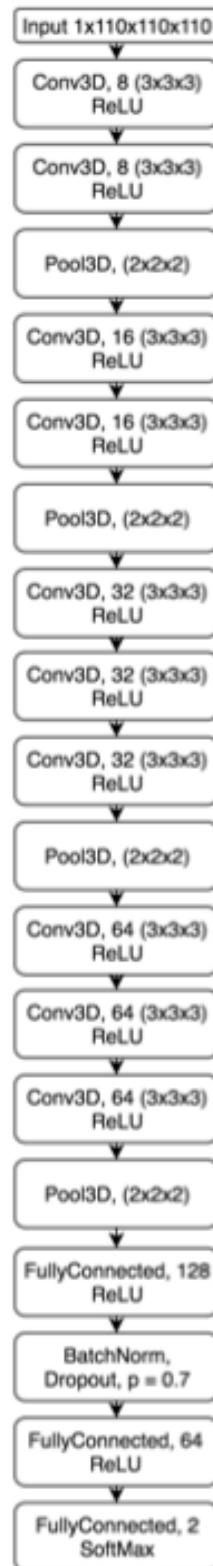
5.5.5. Verificación de resultados en LOO: LOOCV por paciente a las mejores capas

En este caso se realizó fue una implementación especial del algoritmo de *LOOCV*. En lugar de sacar en cada iteración un elemento al conjunto de *test*, se meten en *test* todas las imágenes que pertenece al mismo paciente, gracias a un vector que contiene el número de identificación del paciente. Además, se lleva un registro de los pacientes que ya se han introducido en *test* para que cuando se continúe con el entrenamiento y nos encontremos con otra imagen del mismo paciente no se vuelva a realizar lo mismo que se ha realizado previamente.

ÍNDICE DE CUADROS

Este método de validación se lo aplicamos a las capas que mejores resultados habían obtenido en el experimento anterior 5.5.4 para obtener una referencia muy robusta sobre cuál sería el desempeño de la red a la hora de que llegasen nuevos ejemplos.

Figura 5.16: Arquitectura utilizada basada en la arquitectura VGG [7] del artículo [8]



Base de datos empleada

La base de datos empleada ha sido una de las bases de datos más utilizadas y trabajadas para el estudio del Alzheimer, que es la base de datos conocida como **ADNI (Alzheimer's Disease Neuroimaging Initiative)**. La base de datos pone a disposición datos clínicos de pacientes tales como: imágenes MRI y PET, informes genéticos, test cognitivos y biomarcadores de la sangre. Estos pacientes pertenecen a un grupo de estudio voluntario realizado en Norte América dentro de los cuales se incluye pacientes que padecen de Alzheimer, pacientes que muestran un deterioro cognitivo leve y grupos de control, es decir, sujetos sanos.

El estudio de ADNI ha tenido tres fases: ADNI1, ADNI GO y ADNI2. En cada una de ellas se incorporaban pacientes nuevos para su evaluación médica. Esta evaluación y seguimiento médico se usaban para detectar cualquier patología de Alzheimer en los pacientes.

En nuestro caso se ha utilizado un subconjunto dentro de ADNI en el cuál se había realizado cierto post-procesamiento de las imágenes obtenidas que será explicado a continuación.

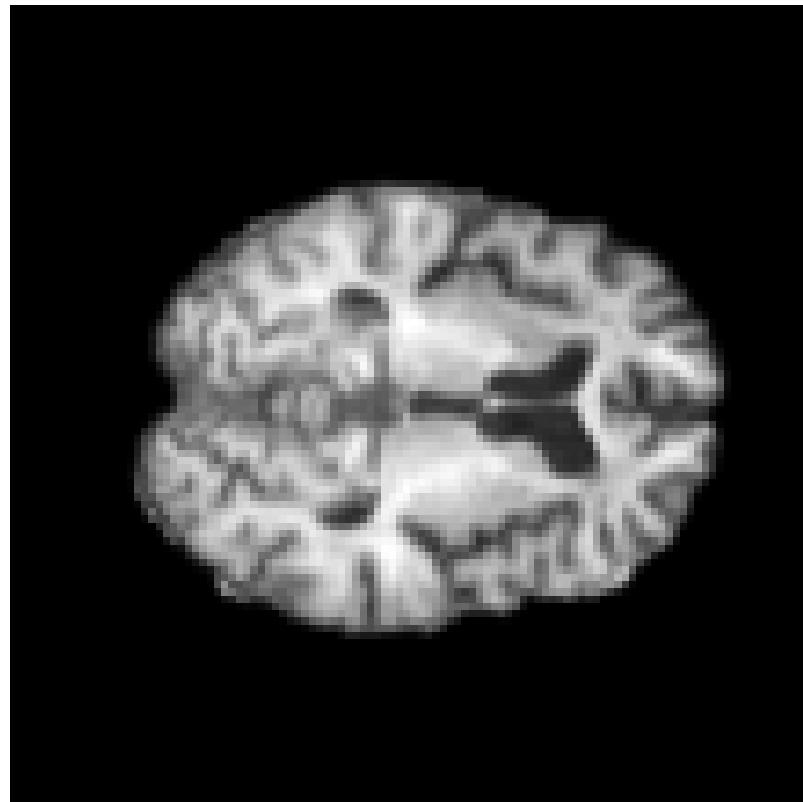
6.1. Post procesamiento

El post-procesamiento está compuesto por los siguientes pasos:

- ***Spatially Normalized:*** Con este paso lo que se consigue es que un mismo pixel corresponda a la misma posición en dos imágenes cerebrales distintas.
- ***Masked:*** Con este paso conseguimos aislar la masa cerebral eliminando el cráneo.
- ***N3 Correction:*** Con este paso realizamos una corrección de la intensidad en la imagen.

Esto permite que tengamos unas imágenes escaladas, centradas y con solo la masa cerebral, lo que nos puede ayudar a que la red aprenda mejores filtros para realizar la clasificación de las imágenes.

Figura 6.17: Ejemplo de una imagen con el post-procesamiento relatado en 6.1

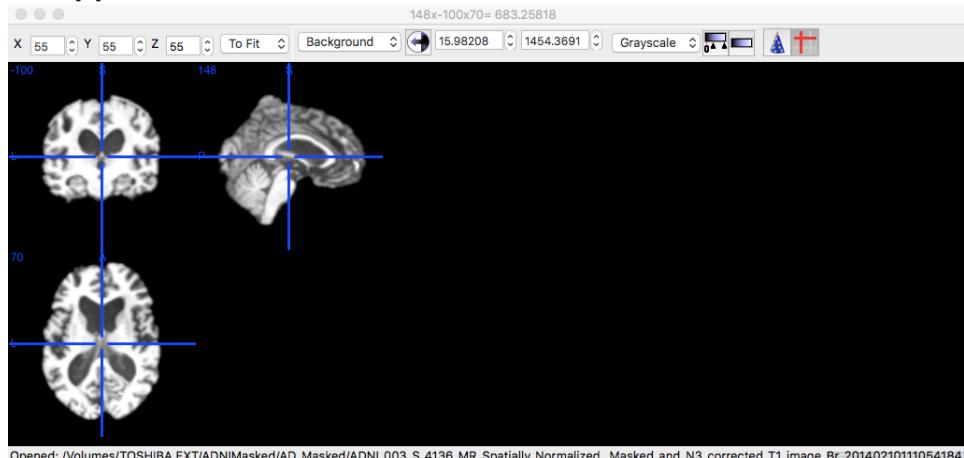


6.2. Software de visualización de las imágenes

Para la visualización de estas imágenes podemos hacer uso de varias herramientas. La utilizada para la visualización de las imágenes en formato *NIFTI* ha sido el software **MRICron**[9]. Este software nos permite visualizar los diferentes cortes cerebrales (coronal, sagital y axial) que están contenidos en el archivo NIFTI.

ÍNDICE DE CUADROS

Figura 6.18: Ejemplo de visualización de una imagen con el software MRIcron [9]



6.3. Separación de las imágenes por tipos

En cuanto a la clasificación de las imágenes, anteriormente cada imagen tenía asociado un archivo *XML* donde venía resumida la información del paciente y a qué tipo de grupo pertenecía. Pero durante este año se ha llevado a cabo una reorganización de la base de datos por lo que ahora, para la separación de los pacientes según el tipo, se utiliza la búsqueda avanzada señalando en una casilla que tipo de pacientes quieras buscar. Así también se han obtenido los imágenes de los pacientes a las que se le había aplicado el post-procesamiento previamente explicado.

6.4. Conjunto reducido y conjunto completo

En el conjunto completo nos encontramos con **389 imágenes**, esto es añadiendo todas las imágenes relacionadas con un mismo paciente. Estas imágenes han sido tomadas en distintas fases. Puede ser con distancias de 6 meses o más tiempo.

En el conjunto reducido nos encontramos con **131 imágenes**, esto es dejando solo la primera imagen de cada paciente en el espacio temporal, como se realizó en [8].

Resultados

7.1. Primeros experimentos: 5-Fold CV con la capa 55 en un conjunto reducido y el conjunto completo

En cuanto a los resultados podemos observar diferencias significativas entre los 3 realizados: el primero, en el cuál se utilizaba el conjunto de datos reducido con la capa 55 en el eje Z, el segundo, en el cuál se utilizaba el conjunto de datos completo con la capa 55 en el eje Z y el tercero, en el cuál utilizábamos otra vez el conjunto de datos completo pero sin asegurar que imágenes de un mismo paciente no se encontrasen en entrenamiento y en test a la vez.

Primero se intentaron repetir los resultados obtenidos en [8], dónde se utilizaron redes convolucionales en 3D para la clasificación de pacientes usando nuestro denominado conjunto reducido 6.4 usando una sola imagen por paciente, y se obtuvo un rendimiento muy inferior, con un **43'5043 %** de *accuracy* de media en los conjuntos de test.

Cuando se sumaron todas las imágenes de los pacientes al conjunto de datos, es decir usando el conjunto completo definido en 6.4, siempre asegurando que no se repetían imágenes de un mismo paciente en *train* y en *test* se observó una mejora, obteniendo un **50'2358 %** de *accuracy* de media en los conjuntos de test. Esto nos puede llevar a pensar que a cuánto mayor el número de imágenes que tenemos, mejores resultados podemos obtener.

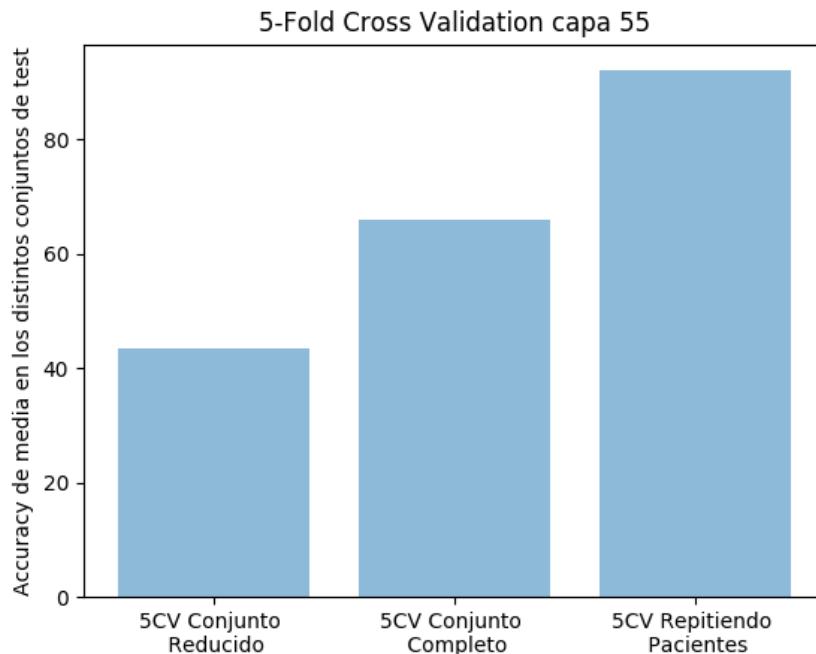
Por último, se realizó un último experimento para comprobar si existía filtración de información entre imágenes de un mismo paciente, es decir, que no aseguramos que no se repitiesen imágenes de un mismo paciente en distintos espacios temporales (como pueden ser en distintos meses) en los conjuntos de entrenamiento y *test*. De esta forma, obtuvimos un resultado de **92 %** de *accuracy* de media en los conjuntos de test, un resultado muy superior a los anteriormente obtenidos, por lo que se puede afirmar que existe filtración de información.

A continuación se puede observar una tabla con un compendio de los resultados previos y una gráfica con los mismos:

Cuadro 7.2: Tabla con los resultados obtenidos en distintos experimentos

Tipo Experimento	Accuracy(%)
5-Fold CV Conjunto Reducido Capa 55	43'5043 %
5-Fold CV Conjunto Completo Capa 55	65'9467 %
5-Fold CV Conjunto Completo Repitiendo pacientes capa 55	92 %

Figura 7.19: Rendimiento en los distintos experimentos, el orden es el mismo que en la tabla 7.1



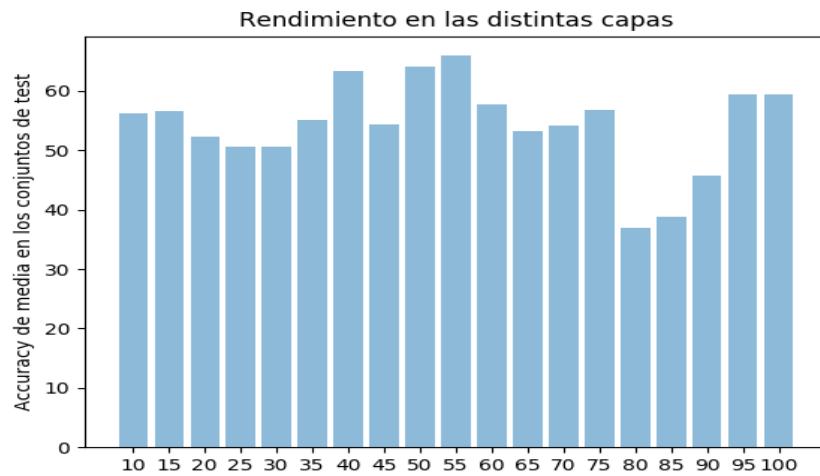
7.2. Identificación del rendimiento de las diferentes capas 2D de las imágenes MRI

Para la realización de este experimento, se tomaron las capas desde la **10 a la 100** de 5 en 5. Los resultados de *accuracy* de media en los conjun-

ÍNDICE DE CUADROS

tos de *test* en las distintas capas se pueden observar en la siguiente tabla 7.2:

Figura 7.20: Rendimiento de las distintas capas



Cuadro 7.3: Tabla con los resultados obtenidos por las distintas capas

Nº Capa	Accuracy (%)
Capa 10	56'326 %
Capa 15	56'5619 %
Capa 20	52'3479 %
Capa 25	50'6972 %
Capa 30	50'6972 %
Capa 35	55'1948 %
Capa 40	63'2912 %
Capa 45	54'3643 %
Capa 50	64'1046 %
Capa 55	65'9467 %
Capa 60	57'7204 %
Capa 65	53'3151 %
Capa 70	54'149 %
Capa 75	56'7601 %
Capa 80	36'9139 %
Capa 85	38'7526 %
Capa 90	45'7758 %
Capa 95	59'4532 %
Capa 100	59'4532 %

7.3. Verificación de resultados en LOO: LOOCV por paciente a las mejores capas

Como se puede observar en la gráfica y en la tabla, las capas con las que mejores resultados se obtienen son la **capa 55, la 50 y la 40**. Es por esto que se decidió utilizar estas capas para obtener un techo realista utilizando un método aún más robusto como es el *Leave-One-Out Cross Validation (LOOCV)*.

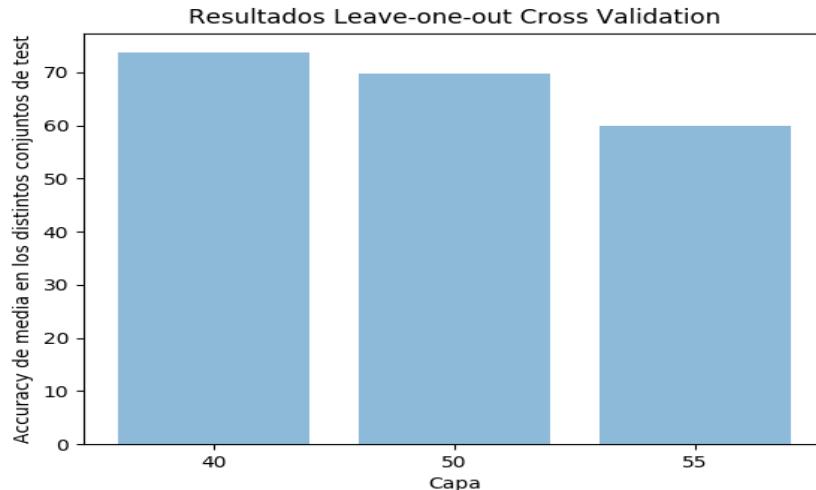
7.3. Verificación de resultados en LOO: LOOCV por paciente a las mejores capas

Los resultados obtenidos se pueden observar a continuación en la siguiente tabla y gráfica, dejando un aumento en los obtenidos en el experimento anterior (en la capa 40 y la 50) y un rendimiento inferior en la capa 55:

Cuadro 7.4: Tabla con los resultados obtenidos en LOO con las tres capas

Tipo Experimento	Accuracy(%)
LOO CV Capa 40	73'7139 %
LOO CV Capa 50	69'6875 %
LOO CV Capa 55	59'9851 %

Figura 7.21: Rendimiento de en los experimentos en los que se utilizó como método de validación LOO CV



Por último, se va a realizar una comparación de los resultados que se ofrecen en [8], ya que es el principal estudio en el que nos hemos basado,

ÍNDICE DE CUADROS

siempre guardando las diferencias en que nosotros realizamos una clasificación con imágenes 2D y un mayor número de imágenes. Se debe recordar que siempre se ha respetado que no se encontrasen imágenes de un mismo paciente en entrenamiento y en *test*. Los mejores resultados obtenidos en [8] usando la misma arquitectura que nosotros pero usando las imágenes cerebrales 3D completas fueron de un **79 %** con un margen de ± 0.8 . Nosotros en su lugar, obtuvimos con la capa 40 del eje Z un **73'7139 %**.

Conclusiones y trabajos futuros

En este capítulo se va a hablar de las conclusiones que se pueden observar a partir de los resultados obtenidos en el capítulo anterior y trabajos futuros que se desearían realizar a partir de lo ya concluido con este trabajo.

8.1. Conclusiones

Con este trabajo se ha procesido a implementar una solución basada en *Deep Learning* para la clasificación de imágenes MRI en 2D, a partir de una base de datos de imágenes en 3D, para diferenciar pacientes enfermos con Alzheimer y pacientes sanos. Hasta el presente trabajo, todas las soluciones basadas en *Deep Learning* para el mismo problema, se habían basado en imágenes en 3D como se puede observar en [8] [34], obteniendo muy buenos resultados con un conjunto no muy extenso de datos en [8].

Los primeros resultados obtenidos, apuntan a que un mayor número de imágenes disponibles mejoran los resultados obtenidos. Así, el conjunto reducido de imágenes con 131 imágenes, no llega a pasar de una clasificación a ciegas, mientras que una clasificación con el conjunto ampliado mejora dicho valor obteniéndose en un 5-CV un resultado de **65'9467 %**.

Se debe resaltar que las comparativas propuestas en este trabajo, evitan la clasificación de imágenes de un mismo paciente en el conjunto de entrenamiento y de test. Sin tener esto en cuenta, se llega a obtener un **92 %**, que se podría entender como un umbral de rendimiento esperado para un conjunto más amplio de imágenes.

Pudiendo llegar a ser una herramienta de utilidad a la hora de que un médico pudiese hacer un diagnóstico a un paciente. Además, al tratarse de imágenes 2D en lugar de 3D, el tiempo para el entrenamiento y la capacidad de computación necesaria para el mismo o la predicción se reducen notable-

mente.

Posteriormente se realizó un estudio para evaluar el rendimiento de las diferentes capas de las imágenes MRI en el eje Z. Así, se implementó un sistema de evaluación *5-Fold CV* con las capas de la 10 a la 100 de 5 en 5, obteniéndose los resultados dados en la tabla 7.2. Con estos en mano podemos observar que las mejores capas para la clasificación son la 40, la 50 y la 55, con unos rendimientos de **63'2912 %**, **64'1046 %** y **65'9467 %** respectivamente. Estas capas corresponden con capas centrales del eje Z, que de alguna manera es de esperar proporcionen más información para la clasificación que las más alejadas del punto medio.//

A continuación, realizamos un método de validación LOOCV a las capas del eje Z que mejor rendimiento nos dieron, siendo estas la 40, la 50 y la 55, obteniendo un **73'7139 %**, **69'6875 %** y **59'9851 %** respectivamente.

En cuanto a la comparación de los resultados obtenidos con los que se exponen en [8], se puede observar que el rendimiento obtenido en las imágenes 2D para la capa 40 del eje Z es levemente inferior al obtenido en 3D, entre ellos obteniendo nosotros un **73'7139 %** y en el trabajo se obtiene un **79 %**. Es de resaltar que ese rendimiento en la capa 40 del eje Z se obtuvo utilizando un LOOCV y no un 5CV, así esto apunta a que es necesaria la utilización de una base de datos mayor, y que esta aproximará paulatinamente los resultados a los obtenidos en un procesamiento 3D [8].

8.2. Trabajos futuros

En cuanto a los trabajos futuros a partir del actual, el principal sería continuar por la línea de comprobar si cuantas más imágenes mejores resultados. Actualmente solo se ha utilizado un subgrupo dentro de la base de datos donde se había aplicado el post procesamiento explicado en 6.1. Como continuación, nos gustaría aplicar el post procesamiento a un número mayor de imágenes y obtener nuevos resultados. Esto requiere un coste computacional y temporal muy grande, el post procesamiento de una sola imagen requiere de media 35 minutos, es por esto que no se han utilizado más datos para este trabajo, pero para una continuación nos gustaría post procesar muchas más imágenes para ver cuáles serían los resultados obtenidos con una base de datos extensa.

Para ello se está utilizando el software FreeSurfer [35], que es el que fue utilizado para normalizar las imágenes de la base de datos utilizada.

Una continuación sería la creación de una plataforma web donde un doctor pudiese subir las imágenes de un paciente y se produjese una predicción

que el doctor pudiese tomar como primera consulta a la hora de diagnosticar al paciente, teniendo los resultados guardados en un servidor web siempre accesible y con la historia de los pacientes.

Bibliografía

- [1] R. C. Petersen, “Mild cognitive impairment,” *New England Journal of Medicine*, vol. 364, no. 23, pp. 2227–2234, 2011. PMID: 21651394.
- [2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” in *Proceedings of the IEEE*, pp. 2278–2324, 1998.
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [4] “Batch normalization example.” <https://shuuki4.files.wordpress.com/2016/01/bn3.png?w=1000>. Accessed: 2017-09-02.
- [5] “Imagen k-fold by fabian flöck.” [https://en.wikipedia.org/wiki/Cross-validation_\(statistics\)#/media/File:K-fold_cross_validation_EN.jpg](https://en.wikipedia.org/wiki/Cross-validation_(statistics)#/media/File:K-fold_cross_validation_EN.jpg). Accessed: 2017-08-16.
- [6] K. Tastambekov, S. Puechmorel, D. Delahaye, and C. Rabut, “Aircraft trajectory forecasting using local functional regression in sobolev space,” vol. 39, pp. 1–22, 02 2014.
- [7] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *CoRR*, vol. abs/1409.1556, 2014.
- [8] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, “Residual and plain convolutional neural networks for 3d brain MRI classification,” *CoRR*, vol. abs/1701.06643, 2017.
- [9] “Mricon webpage.” <http://people.cas.sc.edu/rorden/mricon/index.html>. Accessed: 2017-08-16.
- [10] F. Bermejo Pareja, J. Benito-León, M. Vega, S. Mendrano, and G. Román, “Incidence and subtypes of dementia in three elderly populations of central spain,” *Journal of the Neurological Sciences*, 2008.
- [11] “Alzheimer risks factos.” https://www.alz.org/research/science/alzheimers_disease_causes.asp. Accessed: 2017-04-23.

- [12] “Diagnóstico.” <https://www.alz.org/espanol/treatment/diagnostico.asp>. Accessed: 2017-04-23.
- [13] “Formatos de imágenes.” <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3948928/>. Accessed: 2017-04-23.
- [14] R. Brookmeyer, S. Gray, and C. Kawas, “Projections of alzheimer’s disease in the united states and the public health impact of delaying disease onset,” *American Journal of Public Health*, 1998.
- [15] K. A. Kellett and N. M. Hooper, “Prion protein and alzheimer disease,” *US National Library of Medicine National Institutes of Health*, 2009.
- [16] R. C. Petersen, “Mild cognitive impairment,” *New England Journal of Medicine*, vol. 364, no. 23, pp. 2227–2234, 2011. PMID: 21651394.
- [17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [18] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *International Conference on Machine Learning*, pp. 448–456, 2015.
- [19] P. Scheltens, N. Fox, F. Barkhof, and C. D. Carli, “Structural magnetic resonance imaging in the practical assessment of dementia: beyond exclusion,” *The Lancet Neurology*, vol. 1, no. 1, pp. 13 – 21, 2002.
- [20] M. Hattori, S. Koyama, Y. Kodera, Y. Kogure, Y. Ido, and H. Asano, “Automated measurement of medial temporal lobe atrophy by computed tomography,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 1, pp. 321–330, Apr 2007.
- [21] P. J. Visser, F. R. J. Verhey, P. A. M. Hofman, P. Scheltens, and J. Jolles, “Medial temporal lobe atrophy predicts alzheimer’s disease in patients with minor cognitive impairment,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 72, no. 4, pp. 491–497, 2002.
- [22] E. L. G. E. Koedam, M. Lehmann, W. M. van der Flier, P. Scheltens, Y. A. L. Pijnenburg, N. Fox, F. Barkhof, and M. P. Wattjes, “Visual assessment of posterior atrophy development of a mri rating scale,” *European Radiology*, vol. 21, pp. 2618–2625, Dec 2011.
- [23] J. K. Chung, E. Plitman, S. Nakajima, F. Caravaggio, S. Shinagawa, Y. Iwata, P. Gerretsen, J. Kim, H. Takeuchi, R. Patel, M. M. Chakravarty, A. Strafella, and A. Graff-Guerrero, “The effects of cortical hypometabolism and hippocampal atrophy on clinical trajectories in mild cognitive impairment with suspected non-alzheimer’s pathology: A brief report,” *Journal of Alzheimer’s Disease*, vol. Preprint, pp. 1–7, Aug 2017.

BIBLIOGRAFÍA

- [24] N. Fox, E. Warrington, P. Freeborough, P. Hartikainen, A. Kennedy, J. Stevens, and M. N. Rossor, “Presymptomatic hippocampal atrophy in alzheimer’s disease: A longitudinal mri study,” *Brain*, vol. 119, no. 6, pp. 2001–2007, 1996.
- [25] S. Chaplot, L. Patnaik, and N. Jagannathan, “Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network,” *Biomedical Signal Processing and Control*, vol. 1, no. 1, pp. 86 – 92, 2006.
- [26] J.-D. Lee, S.-C. Su, C.-H. Huang, J. J. Wang, W.-C. Xu, Y.-Y. Wei, and S. T. Lee, “Combination of multiple features in support vector machine with principal component analysis in application for alzheimer’s disease diagnosis,” pp. 512–519, 2009.
- [27] D. Jaramillo, I. Rojas, O. Valenzuela, I. García, and A. Prieto, “Advanced systems in medical decision-making using intelligent computing. application to magnetic resonance imaging,” June 2012.
- [28] E. Hosseini-Asl, R. Keynton, and A. El-Baz, “Alzheimer’s disease diagnostics by adaptation of 3d convolutional network,” *CoRR*, vol. abs/1607.00455, 2016.
- [29] “Caffe webpage.” <http://caffe.berkeleyvision.org>. Accessed: 2017-08-17.
- [30] “Theano webpage.” <http://wwwdeeplearningnetsoftwaretheano/>. Accessed: 2017-08-17.
- [31] “Tensorflow webpage.” <https://www.tensorflow.org>. Accessed: 2017-08-17.
- [32] “Keras webpage.” <https://keras.io>. Accessed: 2017-08-17.
- [33] “Nibabel webpage.” <http://nipy.org/nibabel/>. Accessed: 2017-08-17.
- [34] S. Sarraf and G. Tofighi, “Classification of alzheimer’s disease using fmri data and deep learning convolutional neural networks,” *CoRR*, vol. abs/1603.08631, 2016.
- [35] “Freesurfer webpage.” <http://freesurfer.net>. Accessed: 2017-09-6.