



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 Gullén Perales



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

—
Granada, 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 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 disease 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 for making improves in the pacient life. The problem is that the diagnosis of the disease in an early stage is really difficult, adding the fact that it could be difficult to distinguish in respect to other damages of age. With the new development of Deep Learning techniques in the area of Computer Vision and Image Classification, the use of this techniques for pacient classification could be a usefull help for when deciding if a pacient is developing the disease, helping to treat it in an early stage. With this TFG, we are trying to answer if with this techniques we can make a good classification of 2D brain images of pacientes and if so, which are the best slices of the brain for doing 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 14 de Septiembre de 2017 .

D. **Luis Javier Herrera Maldonado (tutor1)**, Profesor del Área de XXXX del Departamento Arquitectura y Tecnología de Computadores de la Universidad de Granada.

D. **Alberto Guillén Perales (tutor2)**, Profesor del Área de XXXX 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:

**Luis Javier Herrera Maldonado (tutor1) Alberto Guillén Pe-
rales (tutor2)**

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 hubiese 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	13
3.2. Convolutional Neural Networks	13
3.2.1. Convolutional Layer	14
3.2.2. Pooling Layer	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. K-Fold Cross-Validation	19
3.3.2. Leave-one-out Cross-Validation	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. Primeros experimentos: 5-Fold Cross Validation con la capa 55 en un conjunto reducido y el conjunto completo	33
5.5.3. Selección de las mejores capas 2D: 5-Fold Cross Validation para ver cuál es la mejor capa	34
5.5.4. Verificación de resultados en LOO: Leaving-one-out Cross Validation 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
Resultados	41
7.1. Primeros experimentos: 5-Fold Cross Validation con la capa 55 en un conjunto reducido y el conjunto completo	41
7.2. Selección de las mejores capas 2D: 5-Fold Cross Validation para ver cuál es la mejor capa	42
7.3. Verificación de resultados en LOO: Leaving-one-out Cross Validation por paciente a las mejores capas	44
7.4. Comparación de resultados en [1]	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	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 et al. .	15
3.4.	Ejemplo de capa de pool.	16
3.5.	Ejemplo de la operación maxpool.	16
3.6.	Se puede observar como con el batch normalization se consiguen unos mejores resultados con un <i>learning rate</i> alto [3] . .	17
3.7.	Función de activación Softmax	18
3.8.	Función de activación ReLU	19
3.9.	Ejemplo de Cross Validation con K=4 [4]	20
3.10.	Ejemplo de <i>Leave one out Cross Validation</i> [5]	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 [6] del artículo [1]	35
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 [7]	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
7.22.	Comparación de los resultados obtenidos en [1] y nuestros resultados	45

Índice de cuadros

1.1.	Incidencias de la enfermedad en España respecto a la edad [8]	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
7.5.	Comparación entre nuestros resultados y los obtenidos en [1]	45

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 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 [9]. 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 retrasan 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 [10]:

- **Evaluación de estado de ánimo y estado mental:** El estado men-

tal 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, anormalidades 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 [11]. El número de bits con el que trabajan estos sistemas suele ser de 16 bits [11] lo cuál da una escala de grises de hasta 65536 niveles, muy superior a los niveles que el ojo puede diferenciar.

ÍNDICE DE CUADROS

Para tener un estudio más exhaustivo del estado del paciente, se aplican una serie de algoritmos de ordenador 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. He aquí la motivación de la realización del proyecto, poder ayudar a un diagnóstico temprano de la enfermedad para poder tratarla con procedimientos que reduzcan su avance, y mejoren la vida del paciente.

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 [12], 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[13]. 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 [8], 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 [8]

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 Alzheimer.
- **LMCI:** Acrónimo para "*Late Mild cognitive impairment*". El **Mild 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.

ÍNDICE DE CUADROS

- **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 termino 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 [14].

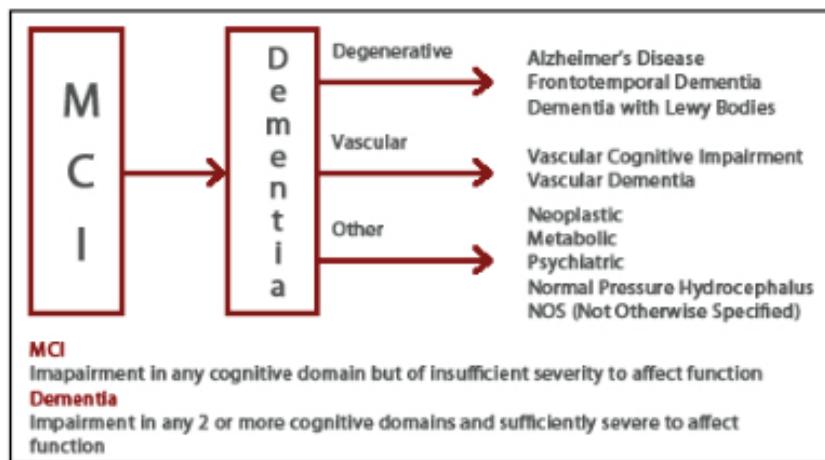


Figura 1.1: Caracterización de MCI que precede al diagnóstico de la enfermedad de Alzheimer

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 además de determinar cuál es la capa del cerebro que mejores resultados presenta a la hora de realizar esta clasificación.

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

Las redes neuronales artificiales son un intento de imitar el comportamiento de las redes neuronales biológicas en nuestro cerebro. 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. Los pesos son los hiperparámetros que se aprenden cuando entrenamos la red neuronal.

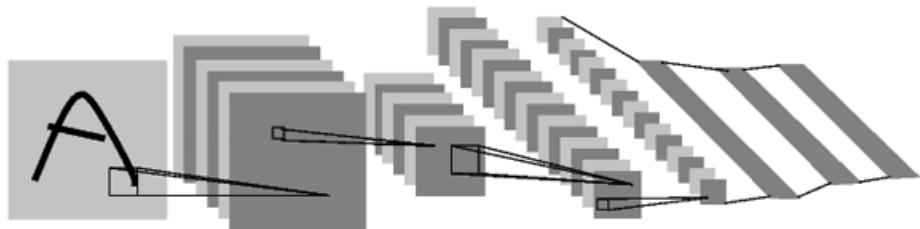
Las redes neuronales son la base de las técnicas utilizadas para la realización de este TFG, que son las *Convolutional Neural Networks*.

3.2. Convolutional Neural Networks

Las *Convolutional Neural Networks (CNN)*, o redes convolucionales en su traducción al español, son un tipo de redes neuronales profundas, es decir, con un número grande de capas con las cuales se han obtenido resultados más que notables en el campo de Visión por Computador.

Las *Convolutional Neural Networks* se componen de una capa de entrada y una de salida además de múltiples *hidden layers* o capas ocultas en su traducción al español. Dentro de estas hidden layers podemos encontrar diversos tipos, pero las principales son las capas de *Convolution*, 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

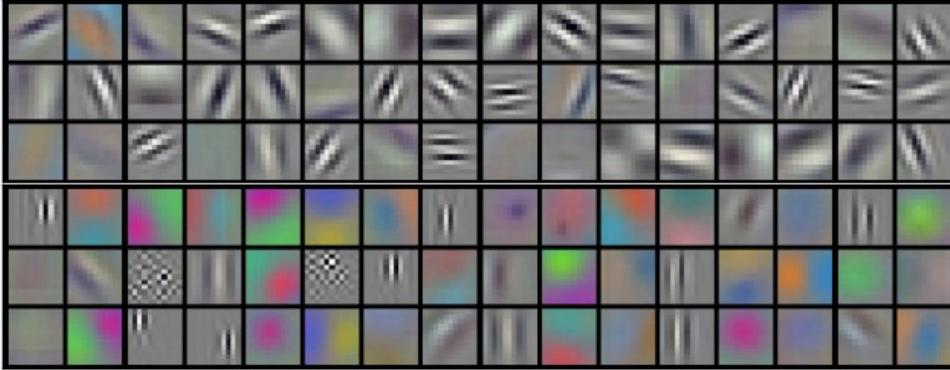
3.2.1. Convolutional Layer

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 red 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 activen cuando vean cierto tipo de característica, como por ejemplo una esquina. Tendremos un filtro por neurona, por lo que habrá tantos filtros como neuronas tenga la capa de convolución.

ÍNDICE DE CUADROS

Figura 3.3: Ejemplo de filtros aprendidos por parte de Krizhevsky et al.



3.2.2. Pooling Layer

Es bastante normal 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 *overfitting*, o sobre ajuste en su traducción en español. El sobre ajuste se produce cuando se aprenden unos hiperparámetros en el entrenamiento que son tan específicos del conjunto de entrenamiento que la red generalice a la hora de que lleguen nuevos ejemplos, por lo que tendremos resultados muy dispares entre los conjuntos de entrenamiento y el de *test*. La capa de *pooling* opera independientemente en cada porción de profundidad de la entrada y le cambia el tamaño espacialmente, usando la operación de MAX. Lo que esto implica es que cada vez que aplicamos una capa de *pooling*, vamos reduciendo el tamaño espacial de la imagen.

Un ejemplo. Si tenemos una imagen de 224x224 y utilizamos una capa de *pooling* de 2x2, a la salida de esa capa, nuestra imagen, o mejor dicho la representación de la misma dentro de la red, tendrá un tamaño de 112x112, ya que la hemos reducido a la mitad con la operación de *pool*. Con la operación de MAX, lo que hacemos es que nos quedamos con el valor del máximo dentro de un grupo de píxeles.

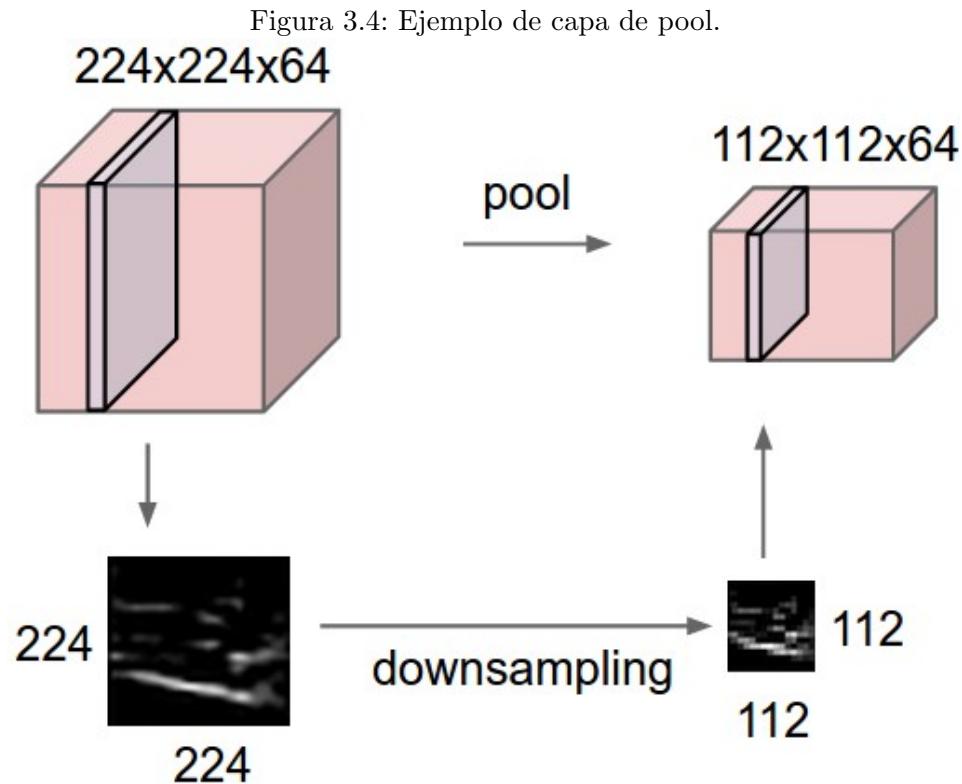
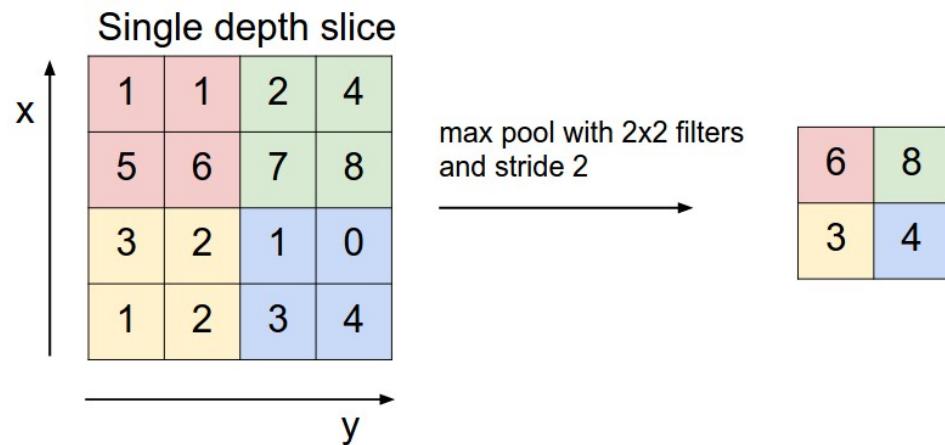


Figura 3.5: Ejemplo de la operación maxpool.



3.2.3. Batch Normalization

Hay un tipo de capa que en los últimos años se ha utilizado ampliamente en la bibliografía y es una capa de *batch normalization* [15]. Con esta capa lo que se consigue es forzar las activaciones a través de la red a tomar una

ÍNDICE DE CUADROS

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: Se puede observar como con el batch normalization se consiguen unos mejores resultados con un *learning rate* alto [3]

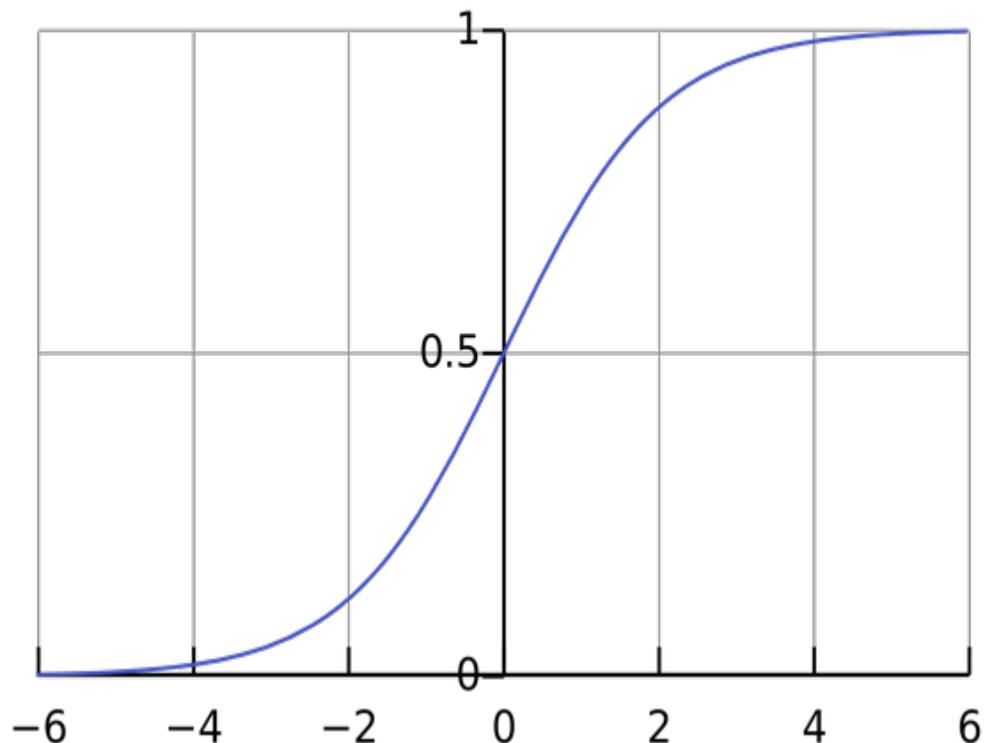


3.2.4. Funciones de activación

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.

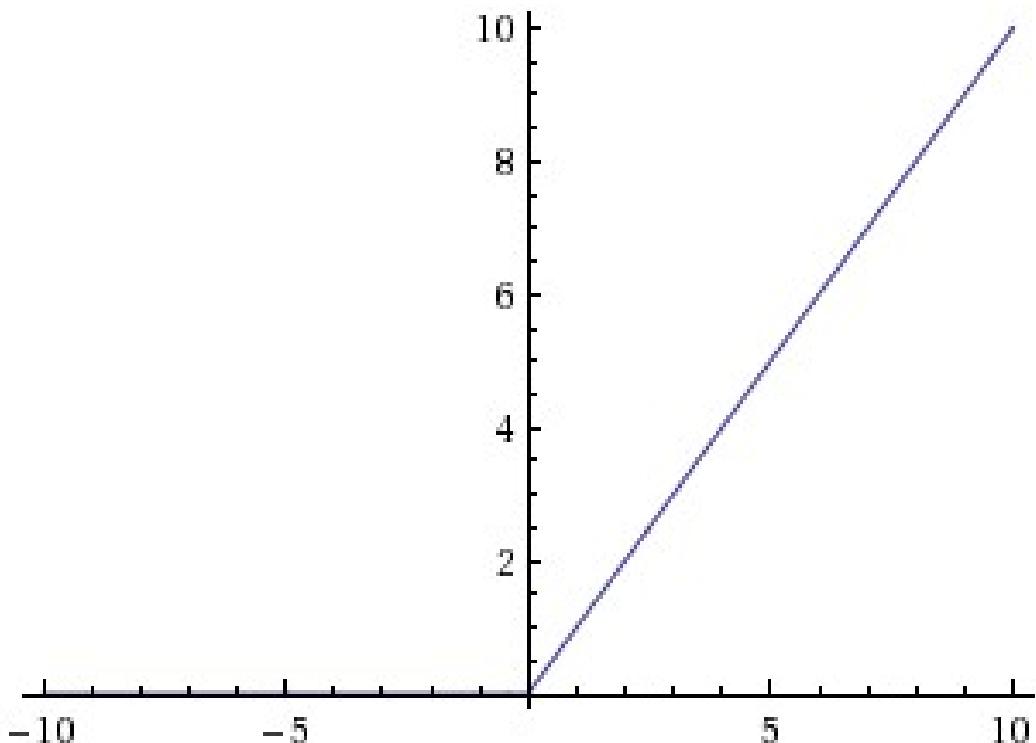
En nuestro caso usamos dos funciones de activación. Una es la función conocida como ReLU(*Rectified Linear Unit*) y la otra es la función conocida como Softmax.

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.

3.3. Métodos de validación de modelos

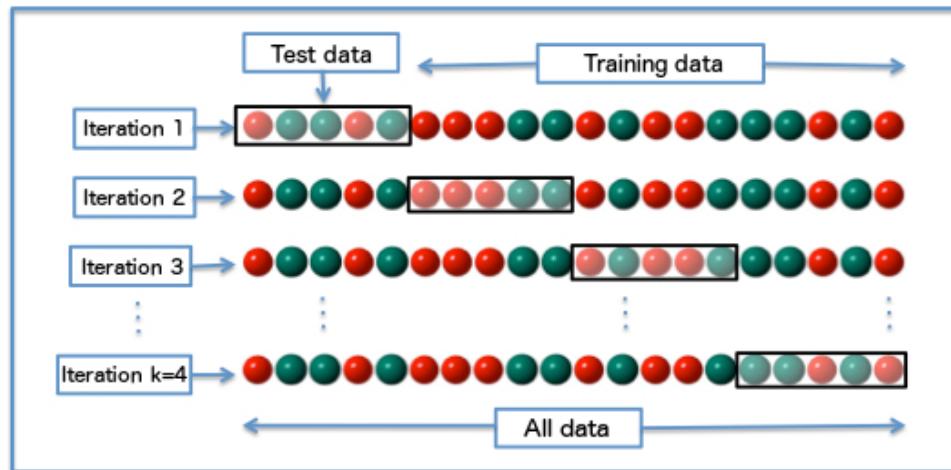
Para la validación del modelo empleado se han utilizado dos métodos, los cuáles son *K-Fold cross-validation* y el *Leave-one-out cross-validation*, que a continuación serán explicados brevemente.

3.3.1. K-Fold Cross-Validation

En el *K-Fold Cross-Validation* 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 *training*. Esto se repite K veces, sin repetir un subconjunto en *test*. Entonces, se obtiene

una media de los resultados en *test* y con esto podemos hacernos una idea de como funciona nuestro modelo y de si generalizará bien. Esto incrementa el tiempo de entrenamiento y comprobación, por lo que no es un método muy usado en los los conjuntos de datos cuyo tamaño es muy grande, pero es de gran ayuda en conjuntos de datos más pequeños lo que nos permite validar mejor como será la generalización del modelo.

Figura 3.9: Ejemplo de Cross Validation con K=4 [4]

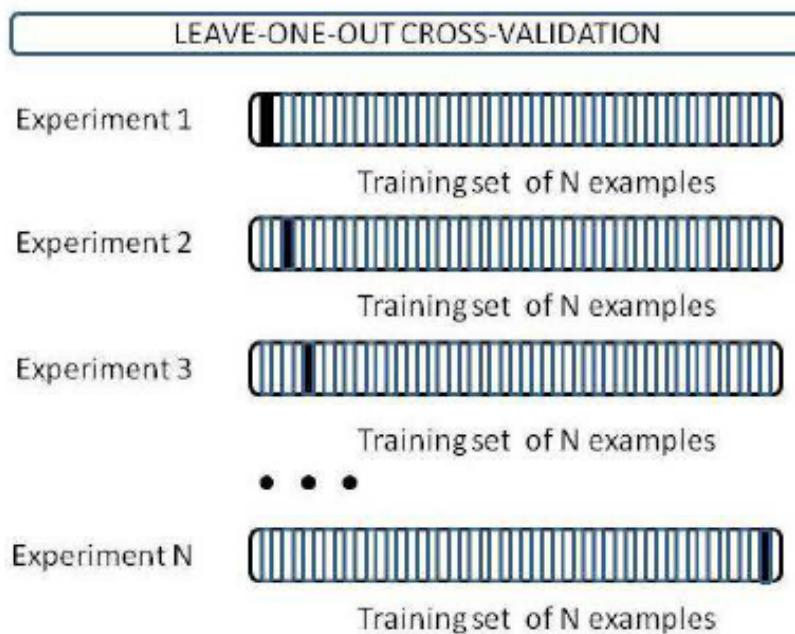


3.3.2. Leave-one-out Cross-Validation

Este método podría ser un tipo de *K-Fold cross-validation* en el que K es igual al número de elementos dentro del conjunto de datos. Esto implica que el tiempo de validación sea muy alto, ya que vamos a realizar K entrenamientos y K evaluaciones, pero es un método muy robusto para la validación del modelo.

ÍNDICE DE CUADROS

Figura 3.10: Ejemplo de *Leave one out Cross Validation* [5]



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 opcionales. 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. [16].

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 tofas 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

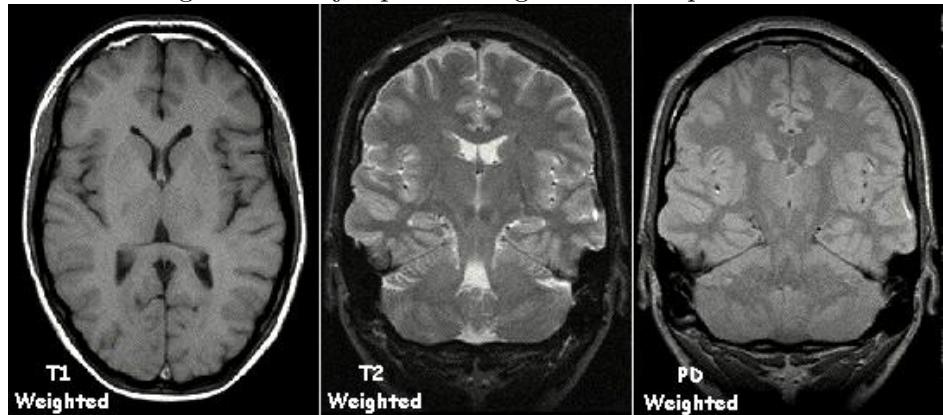
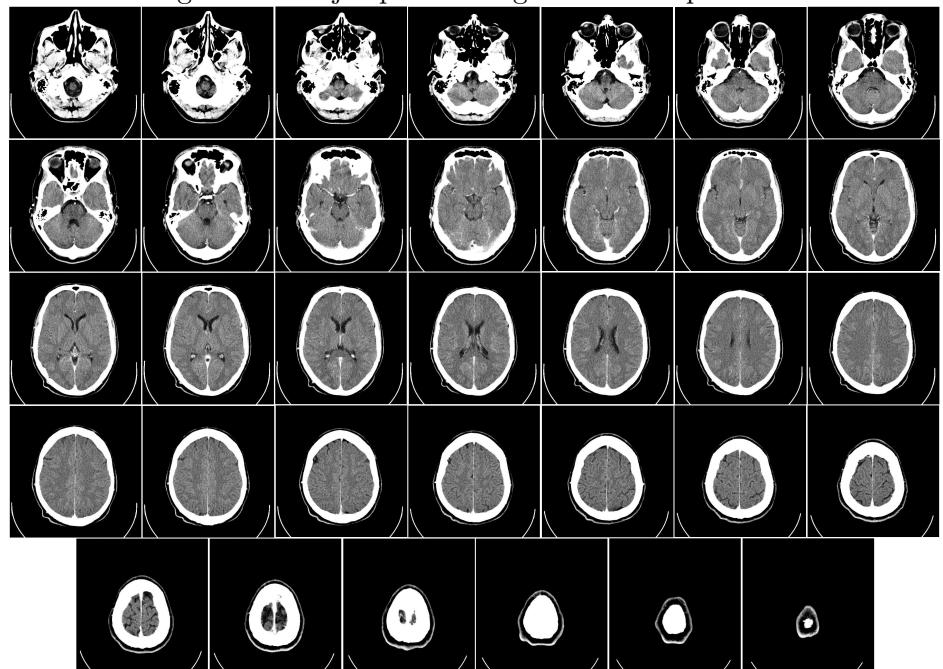


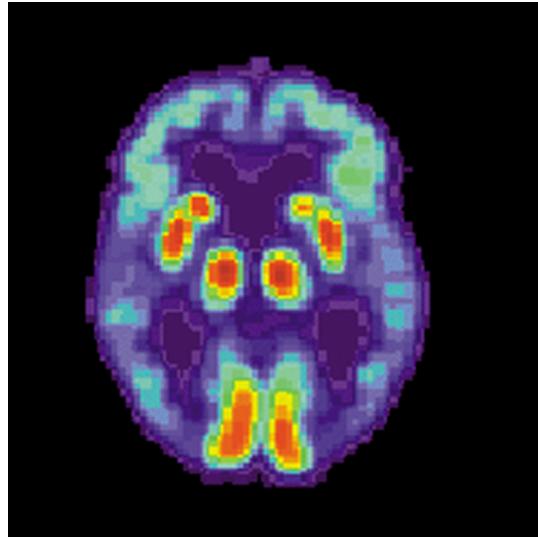
Figura 4.12: Ejemplo de imagen obtenida por CAT



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 [17], y esto se ha utilizado para la predicción de pacientes que padecen la enfermedad [18].

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 [19].

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. [20].

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 [21].

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 *Machine Learning* es un campo que en la última década ha cobrado mucha fuerza a la hora del diagnóstico de la enfermedad. En [22] podemos observar el uso ondas 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 *accuracy*.

Estudios siguientes continuaban la misma línea de trabajo, usando como características las ondas cerebrales y luego utilizando algún algoritmo de *Machine Learning* como clasificador. Además, también se incluyeron métodos como el PCA(*Principal Component Analysis*) para reducir el tamaño dimensional de los datos. [23]

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 d el uso del PCA para la reducción de la dimensión, permite obtener unos buenos resultados a la hora de la clasificación, como se puede comprobar en [24].

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 [1], 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 [25].

Como se puede observar, la mayoría de estudios realizados hacen usos de redes convolucionales en tres dimensiones, lo que produce que se necesite mucha capacidad de computación y tiempo para obtener los resultados del

ÍNDICE DE CUADROS

entrenamiento. Este es un campo en el que aún queda mucho desarrollo, sobre todo para llegar a unas técnicas que se pudiesen implantar como ayuda a la toma de decisión de un médico en caso de diagnóstico.

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 exlicación 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**[26] o **Theano**[27] 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**[28]. 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*[29]. 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

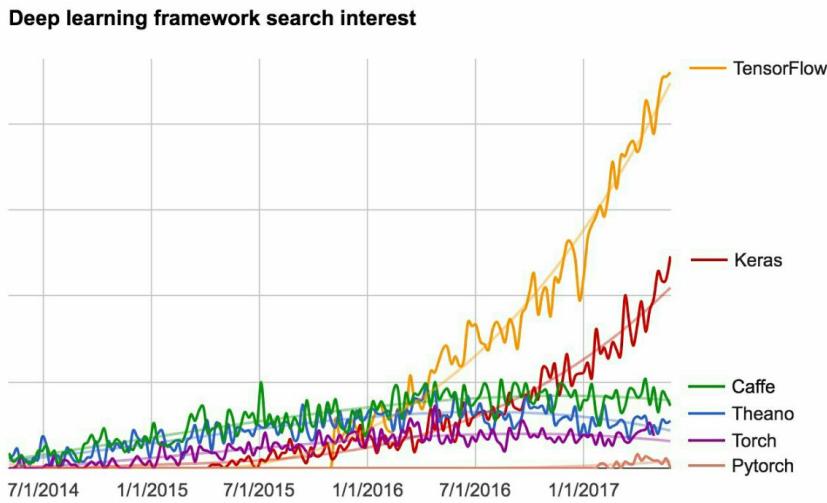


Figura 5.14: Interés de búsqueda de Frameworks de Deep Learning. Fuente: François Chollet

new stars from 2017-04-20 to 2017-07-06		
#1:	7929	tensorflow/tensorflow
#2:	2465	fchollet/keras
#3:	1894	caffe2/caffe2
#4:	1526	BVLC/caffe
#5:	1250	pytorch/pytorch
#6:	1233	Microsoft/CNTK
#7:	979	dmlc/mxnet
#8:	709	deepmind/sonnet
#9:	690	tflearn/tflearn
#10:	485	deeplearning4j/deeplearning4j
#11:	458	Theano/Theano
#12:	452	davisking/dlib
#13:	341	torch/torch7
#14:	303	baidu/paddle
#15:	243	pfnet/chainer

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, debido a que la capacidad de computación requerida para trabajar con las imágenes en 3D no era posible, 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**[30]. 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 slice 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 [1] y [31].

Por parecernos más concluyentes los resultados de [1] 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** [6].

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 [1] también se prueba otro tipo de arquitectura, un tipo de arquitectura residual. Pero en los resultados no se encuentra una mejora sustancial como para aplicar una arquitectura más compleja, y que requiera

ÍNDICE DE CUADROS

más tiempo de computación.

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, eligiendo la capa del cerebro más significativa para la clasificación.

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 *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. Primeros experimentos: 5-Fold Cross Validation 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 nuestro caso, al ser las imágenes de **110x110**, la capa media del cerebro es la capa 55.. Como tenemos un conjunto de datos reducido, ya que solo disponemos de 389 imágenes 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 [1] previamente citado, se redujo el conjunto de datos al utilizar solo una imagen por paciente, ya que hay pacientes de los cuales disponemos más de una imagen de distinto período temporal, dejando el conjunto de datos a una cantidad de 131 imágenes. Para ello se implementó un *5-Fold Cross Validation*, ya que es lo que se utilizó en [1] 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 Cross Validation* pero repitiendo imágenes del mismo paciente en distintos conjuntos, para comprobar que se producía filtración de información entre imágenes de un mismo paciente.

5.5.3. Selección de las mejores capas 2D: 5-Fold Cross Validation 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.

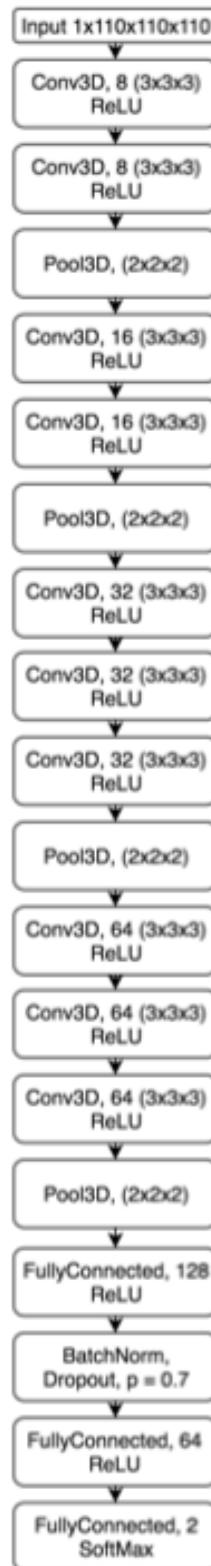
5.5.4. Verificación de resultados en LOO: Leaving-one-out Cross Validation por paciente a las mejores capas

En este caso se realizó fue una implementación especial del algoritmo de *Leave-one-out Cross-Validation* en el que 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 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 encontramos con otra imagen del mismo paciente no se vuelva a realizar lo mismo que se ha realizado previamente.

Este método de validación se lo aplicamos a las capas que mejores resultados habían obtenido en el experimento anterior 5.5.3 para obtener una referencia muy robusta sobre cuál sería el desempeño de la red a la hora de que llegasen nuevos ejemplos.

ÍNDICE DE CUADROS

Figura 5.16: Arquitectura utilizada basada en la arquitectura VGG [6] del artículo [1]



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.

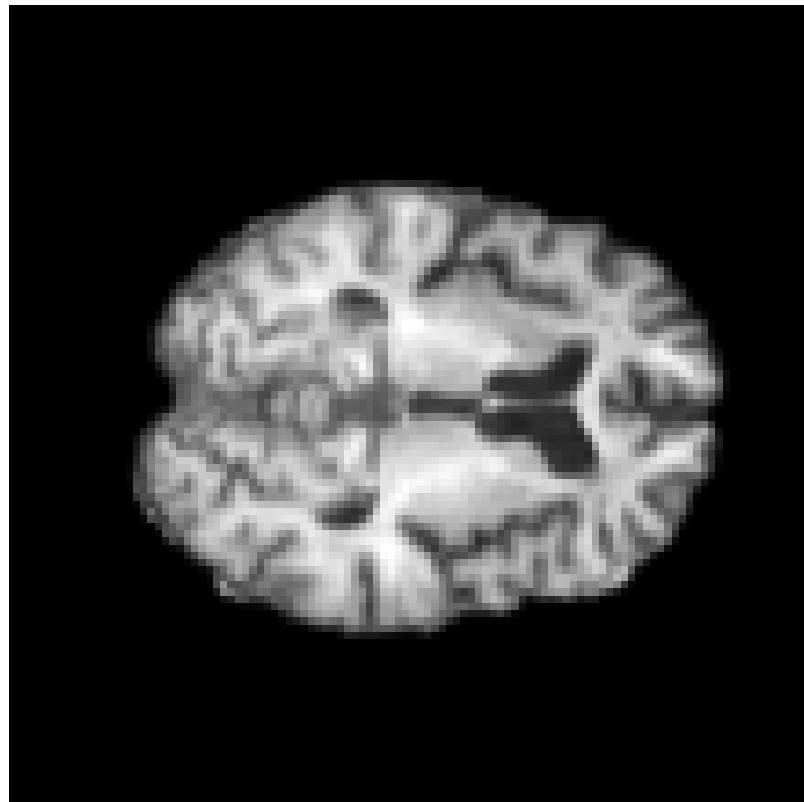
6.1. Post procesamiento

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

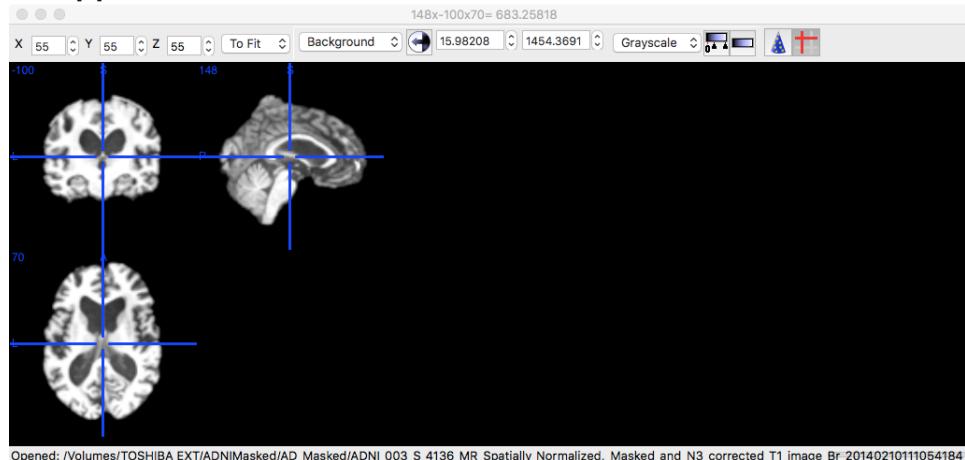


6.2. Software de visualización de las imágenes

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**[7]. 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 [7]



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.

Resultados

7.1. Primeros experimentos: 5-Fold Cross Validation 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.

Primero se intentaron repetir los resultados obtenidos en [1], 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, 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, obteniendo 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:

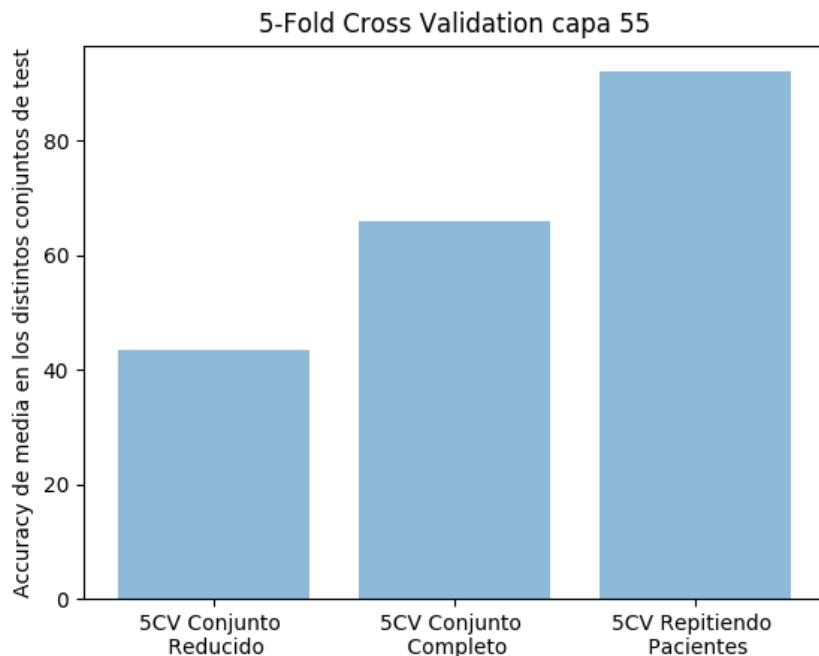
7.2. Selección de las mejores capas 2D: 5-Fold Cross Validation

42 para ver cuál es la mejor capa

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

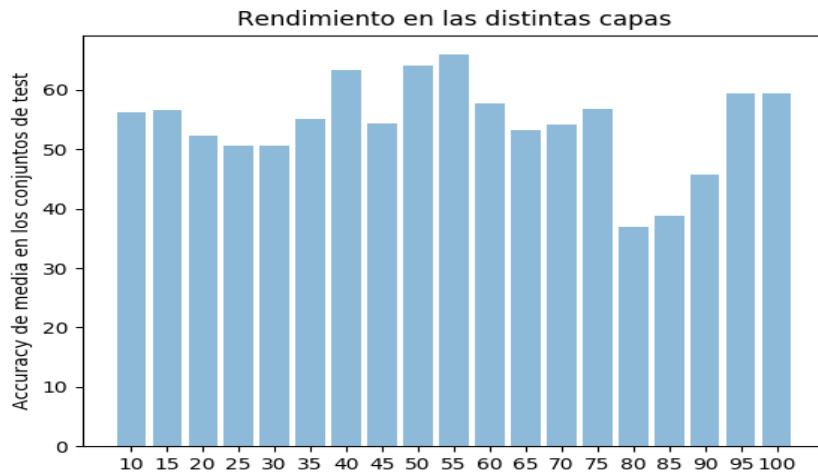


7.2. Selección de las mejores capas 2D: 5-Fold Cross Validation para ver cuál es la mejor capa

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 conjuntos de *test* en las distintas capas se pueden observar en la siguiente tabla ??:

ÍNDICE DE CUADROS

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 %

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

44 un método aún más robusto como es el *Leave-One-Out Cross Validation*.

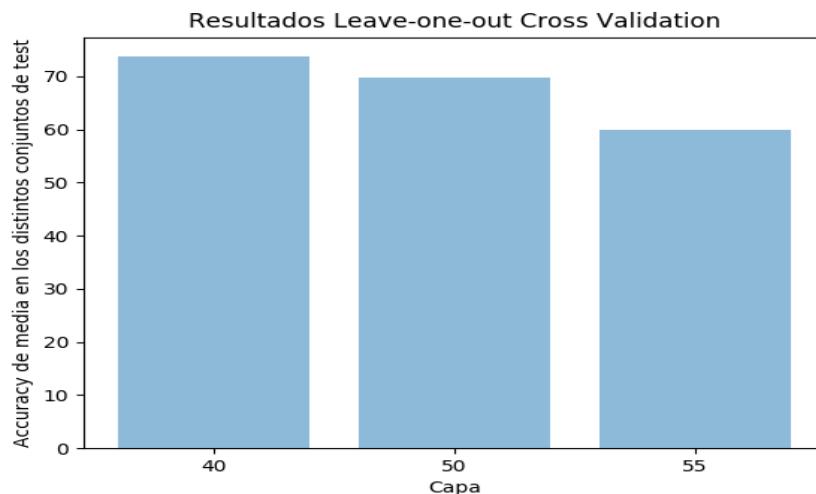
7.3. Verificación de resultados en LOO: Leaving-one-out Cross Validation por paciente a las mejores capas

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



7.4. Comparación de resultados en [1]

Por último, se va a realizar una comparación de los resultados que se ofrecen en [1], ya que es el principal estudio en el que nos hemos basado,

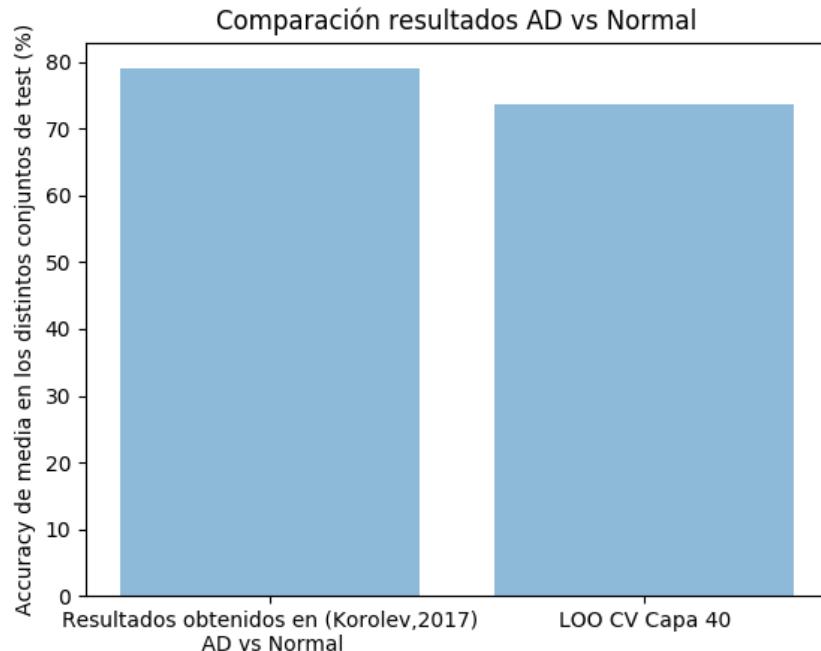
ÍNDICE DE CUADROS

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 encuentren imágenes de un mismo paciente en *training* y en *test*. En la siguiente tabla y gráfica se hace una comparación entre los resultados obtenidos en [1] con la arquitectura que nosotros también hemos propuesto para la clasificación entre alzheimer y normal y nuestros mejores resultados obtenidos en el experimento anterior 7.3:

Cuadro 7.5: Comparación entre nuestros resultados y los obtenidos en [1]

Resultado en [1] para AD vs Normal	LOO CV Capa 40
79 % \pm 0.8	73'7139 %

Figura 7.22: Comparación de los resultados obtenidos en [1] y nuestros resultados



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 intentaban dar respuesta a varias cuestiones. Una sería **si se pueden obtener buenos resultados usando técnicas de Deep Learning para la clasificación de imágenes de pacientes con Alzheimer y pacientes sin esta enfermedad usando imágenes en 2D**, al contrario que en estudios previos donde se utilizaban las imágenes en 3D [1] [31].

Como se puede observar en los experimentos previos 7.1, los resultados mejoran cuanto mayor es la base de datos que tenemos. Esto deja dilucidar que si aumentamos la base de datos cada vez con más imágenes se podrían obtener muy buenos resultados a la hora de la clasificación de pacientes que nos llegasen, 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 notablemente, dejando un camino abierto para que se pudiese realizar en un ordenador menos potente, lo cuál abriría más caminos para que se pudiese implantar un sistema así en un hospital.

La segunda cuestión a la que se intentaba dar respuesta era que, una vez visto que la clasificación de los pacientes con imágenes 2D es viable y se pueden observar mejoras, **cuál es la capa del cerebro con la que se debería hacer esta clasificación**. Para ello se realizó un *5-Fold CV* con las capas de la 10 a la 100 de 5 en 5, y se pueden ver sus resultados

en 7.2. Con estos en mano podemos observar que las mejores capas para la clasificación son la **40,55,50**. Con estos resultados se puede afirmar que en cuanto a la clasificación de imágenes cerebrales en cuanto a la enfermedad de Alzheimer, existen unas capas en el cerebro que funcionan mejor que otras.//

Además, como también hemos comprobado que hay una filtración de información entre distintas imágenes del mismo paciente con los resultados en 7.1 , siempre tenemos que realizar una separación de los pacientes y esto se debería realizar también a la hora de que se implantase este procedimiento en un centro médico.

Gracias a estos resultados podríamos afirmar que hay un futuro a la hora de realizar predicciones sin utilizar imágenes 3D, reduciendo los costes de computación y sin realizar un procesamiento de las imágenes muy costoso como se ha explicado en 5.3, pero que se debe disponer de una mayor base de datos para que incrementemos la generalización de la red. Además, se ha confirmado que la estructura de la red utilizada en [1] es útil también para la clasificación de imágenes 2D, obteniendo unos resultados aceptables en 7.2 y 7.3.

En cuanto a la comparación de los resultados obtenidos con los que se exponen en [1], se puede observar que no existe una diferencia muy significativa entre ellos obteniendo nosotros un **73'7139 %** y en el trabajo se obtiene un **79 %**. También se debe añadir, que ellos lo obtienen utilizando un método de validación como es el *5-Fold CV* mientras que nosotros utilizamos un *LOO CV*. Además ellos utilizan una base de datos más pequeña, pero también ellos utilizan las imágenes en 3D. Además, también nuestro método de validación es más robusto que un *5-Fold CV*.

Lo que también se puede afirmar es que un mayor número de imágenes de pacientes da lugar a unos mejores resultados, como se pudo observar en los primeros experimentos que se realizaron 7.1. Esto indicaría que si se pudiese normalizar la base de datos actual y aplicarle el post procesamiento 6.1 al que fue sometida la base de datos utilizada para este trabajo, se podrían obtener unos muy buenos resultados a la hora de la clasificación de los pacientes.

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.

Para ello se está utilizando el software FreeSurfer [32], que es el que fue utilizado para normalizar las imágenes de la base de datos utilizada.

Además, después de haber respondido a la pregunta de que capas funcionan mejor que otras, lo siguiente sería responder porque unas capas funcionan mejores que otras. Este es otro estudio muy interesante el cuál sería recomendable realizar para darle una respuesta a estas preguntas.

Como último, una continuación muy interesante una vez que se haya respondido a estas cuestiones y se hayan conseguido buenos resultados, 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. Esto sería realizable una vez que se hubiesen obtenido buenos resultados y estos tuviesen una robustez muy considerable.

Bibliografía

- [1] 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.
- [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] “Batch normalization example.” <https://shuuki4.files.wordpress.com/2016/01/bn3.png?w=1000>. Accessed: 2017-09-02.
- [4] “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.
- [5] 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.
- [6] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *CoRR*, vol. abs/1409.1556, 2014.
- [7] “Mricon webpage.” <http://people.cas.sc.edu/rorden/mricon/index.html>. Accessed: 2017-08-16.
- [8] 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.
- [9] “Alzheimer risks factos.” https://www.alz.org/research/science/alzheimers_disease_causes.asp. Accessed: 2017-04-23.
- [10] “Diagnóstico.” <https://www.alz.org/espanol/treatment/diagnostico.asp>. Accessed: 2017-04-23.
- [11] “Formatos de imágenes.” <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3948928/>. Accessed: 2017-04-23.

- [12] 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.
- [13] K. A. Kellett and N. M. Hooper, “Prion protein and alzheimer disease,” *US National Library of Medicine National Institutes of Health*, 2009.
- [14] R. C. Petersen, “Mild cognitive impairment,” *New England Journal of Medicine*, vol. 364, no. 23, pp. 2227–2234, 2011. PMID: 21651394.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.

BIBLIOGRAFÍA

- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] “Caffe webpage.” <http://caffe.berkeleyvision.org>. Accessed: 2017-08-17.
- [27] “Theano webpage.” <http://www.deeplearning.net/software/theano/>. Accessed: 2017-08-17.
- [28] “Tensorflow webpage.” <https://www.tensorflow.org>. Accessed: 2017-08-17.
- [29] “Keras webpage.” <https://keras.io>. Accessed: 2017-08-17.
- [30] “Nibabel webpage.” <http://nipy.org/nibabel/>. Accessed: 2017-08-17.
- [31] 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.
- [32] “Freesurfer webpage.” <http://freesurfer.net>. Accessed: 2017-09-6.