



INSTITUTO TECNOLÓGICO DE
BUENOS AIRES

DEEP LEARNING EN LA DETECCIÓN DE
ALZHEIMER UTILIZANDO IMÁGENES DE
RESONANCIA MAGNÉTICA

Proyecto Final de Carrera

Autora:

Valentina PANIZA

Tutor:

Ing. Carlos SELMO

Buenos Aires

22 de septiembre de 2020

Agradecimientos

Al Prof. Ing. Carlos Selmo por su predisposición, tiempo y recursos en su rol de tutor.

A la Iniciativa de Neuroimagen de la Enfermedad de Alzheimer (ADNI) por haberme brindado acceso a las imágenes utilizadas en el proyecto.

Por último, a mi familia por haber tolerado tantas frustraciones y angustias, y apoyarme en todo momento a lo largo de la carrera.

Resumen

Deep Learning en la detección de Alzheimer utilizando Imágenes de Resonancia Magnética

A lo largo de las últimas décadas diferentes modelos de aprendizaje automático han ido explorando diversas áreas de la medicina con el fin de brindarle herramientas de soporte a los profesionales de la salud. Particularmente, técnicas de aprendizaje profundo han resurgido debido a, en gran medida, el aumento de la capacidad computacional y la disponibilidad de nuevos conjuntos de datos masivos. Los avances en esta disciplina contribuyen en la identificación, clasificación y cuantificación de patrones en imágenes médicas. El presente trabajo tiene por objeto extraer patrones de neuroimágenes mediante algoritmos de aprendizaje profundo que den soporte en el diagnóstico de la enfermedad de Alzheimer en adultos mayores en aplicaciones clínicas. El desafío yace en poder distinguir pacientes sanos de enfermos en imágenes donde los patrones cerebrales e intensidades son muy similares.

Se utilizaron imágenes de resonancia magnética (IRM) obtenidas de la base de datos ADNI (Alzheimer's Disease Neuroimaging Initiative)¹ en distintas arquitecturas de redes convolucionales (CNN de su sigla en inglés, Convolutional Neural Networks) con el fin de diferenciar a individuos en condiciones normales de aquellos que padecen Alzheimer. Estos modelos se ejecutaron en una unidad de procesamiento de gráficos (GPU) de alto rendimiento, habiendo previamente analizado y preprocesado los datos.

¹Los datos utilizados en la preparación de este artículo se obtuvieron de la base de datos de la Iniciativa de Neuroimagen de la Enfermedad de Alzheimer (ADNI) (adni.loni.usc.edu). ADNI se lanzó en 2003 como una asociación público-privada, dirigida por el investigador principal Michael W. Weiner, MD. El objetivo principal de ADNI ha sido probar si la resonancia magnética en serie (MRI), la tomografía por emisión de positrones (PET), otros marcadores biológicos y la evaluación clínica y neuropsicológica se pueden combinar para medir la progresión del deterioro cognitivo leve (MCI) y la detección temprana Enfermedad de Alzheimer (EA). Para obtener información actualizada, consulte www.adni-info.org.

Lista de Abreviaturas

- APOE** Apolipoproteína E [9]
- CE** corteza entorrinal [13]
- CNN** redes convolucionales [iv, 22, 23]
- DCL** Demencia con Cuerpos de Lewy [11, 36]
- DV** Demencia Vascular [10, 11]
- EA** enfermedad de Alzheimer [1, 2, 11-13, 17, 19, 28-30, 33-37]
- FLAIR** FLAIR (Fluid-attenuated inversion recovery) [17]
- GBD** Global Burden of Disease [5]
- GPU** Graphic Processing Unit [iv]
- IA** Inteligencia artificial [19, 20]
- IRM** Imagenes de resonancia magnetica [iv, 12, 13]
- LCR** líquido cefalorraquídeo [13, 16, 36]
- MAE** Mean Absolute Error [27]
- MCI** deterioro cognitivo leve [12, 34]
- MMSE** Mini-Mental State Exam [9, 12]
- MoCA** Montreal Cognitive Assessment [12]
- MP-RAGE** MP-RAGE [17]
- MSE** Mean Squared Error [27]
- OMS** Organización Mundial de la Salud [7]
- RF** radiofrecuencia [14-16]
- RM** resonancia magnética [14-17, 35, 36, 38]
- RRNN** redes neuronales [20, 22]
- TC** tomografía computarizada [14, 17, 36]

TE tiempo de eco [15, 16]

TI tiempo de inversión [15]

TR tiempo de repetición [15, 16]

Lista de Contenidos

Agradecimientos	iii
Resumen	iv
Lista de Abreviaturas	v
1 Introducción	2
2 Marco Teórico	4
2.1 Demencias	4
2.1.1 Prevalencia e Incidencia	6
2.1.2 Enfermedad de Alzheimer	11
2.2 Imagenología en Demencias: Resonancia Magnética	15
2.3 Redes Neuronales	20
2.3.1 Redes Convolucionales	24
2.4 Evaluación del modelo	29
2.4.1 Criterios de comparación	34
3 Metodología	40
3.1 Adquisición de Imágenes de RM	40
3.2 Preprocesamiento	42
3.3 Modelos de CNN Implementados	45
3.3.1 AlexNet	47
3.3.2 VGG-16	49
3.3.3 Modelos Adaptados	51
3.4 Resultados	52
4 Conclusiones y trabajos futuros	56
Bibliografía	59
A Apéndice	66
A.1 Modelos de tipo VGG16	66
A.2 Modelos de tipo AlexNet	68

1 Introducción

A lo largo de los años la comunidad científica ha orientado su búsqueda hacia nuevos métodos que puedan brindar soluciones a cuestiones de salud que afectan a cada vez más personas a nivel mundial. Los trastornos mentales no son una excepción: son más de 50 millones de personas las que padecen algún tipo de demencia, y dos terceras corresponden a la **Enfermedad de Alzheimer (EA)**. La ausencia de una cura, los problemas socio-económicos que acarrea, el carácter de irreversibilidad, y la falta de un método diagnóstico capaz de identificar a individuos con **EA** de manera temprana, convierten esta cuestión en un problema de gran magnitud e impacto social.

El dilema se torna más complejo si se considera que vivimos en un mundo con una población creciente en edad, en donde la esperanza de vida es cada vez mayor, y el número de personas que alcanzan edades avanzadas es cada vez más grande. Asimismo, hay que resaltar que el 60% de los implicados reside en países de ingresos bajos y medios, y sabiendo que es una enfermedad que afecta no sólo a quienes la padecen sino también a gran parte del entorno que los circunde, se convierte en una problemática mundial social, económica y emocional de gran escala. Encontrar una herramienta que sirva de soporte a los profesionales de la salud en el diagnóstico temprano de la enfermedad es una tarea pendiente primordial.

Las imágenes médicas han sido utilizadas a lo largo del último siglo como herramientas para estudios clínicos, para el diagnóstico de anomalías o enfermedades, como instrumento a la hora de planificar tratamientos, y recientemente también en cirugías asistidas por computadoras. La digitalización que surgió a partir de la década del 70 ha acrecentado sustancialmente el uso de dichas imágenes y facilitado la accesibilidad, brindando valiosos beneficios y oportunidades de desarrollo en diversas áreas de la medicina. El nivel de detalle que proporcionan es cada vez más preciso. Con la tomografía computarizada, la resonancia magnética y la ecografía, los órganos normales o patológicos ahora se pueden visualizar

con un grado de detalle tal que la imagen digital es casi un duplicado del órgano real. Los niveles de ruido han mejorado considerablemente, y los tiempos de adquisición son cada vez menores.

Pero una de las mayores ventajas que se ha manifestado con la digitalización es la creciente disponibilidad de datos. Cada vez son más las bases que almacenan grandes volúmenes de datos que son accesibles de manera remota para ser utilizados en investigación y desarrollo, y la importancia de contar con ellos es poder generar nuevos modelos que permitan automatizar procesos y dar soporte a los profesionales en la toma de decisiones; modelos que sean más certeros y precisos, que minimicen el error del ojo humano y sean capaces de brindar diagnósticos de manera más temprana.

El incremento en las capacidades computacionales no sólo ha expandido el rango de visualización remota hacia cualquier lugar con acceso a internet, sino que, junto con el avance de nuevos métodos de predicción y categorización de patologías ha virado la investigación hacia diferentes campos de aplicación de la inteligencia artificial. Numerosos estudios han revelado que la incorporación de distintas técnicas de aprendizaje profundo parecería ser una alternativa prometedora a la hora de diagnosticar enfermedades [1, 2, 3] y reconocer patrones o estructuras [4, 5].

En el presente trabajo se evaluarán distintos modelos de redes neuronales, particularmente redes convolucionales, aplicados en imágenes de resonancia magnética previamente procesadas con el objeto de construir un modelo capaz de discriminar entre individuos sanos e individuos con EA. La finalidad del trabajo es generar una herramienta capaz de diagnosticar a aquellas personas que tienen EA.

2 Marco Teórico

2.1 Demencias

La demencia es un síndrome que atormenta a más de 50 millones de personas a nivel mundial, de las cuales aproximadamente el 60% vive en países de ingresos bajos y medios. La ausencia de una cura o tratamiento efectivo y el hecho de que la prevalencia se duplique cada 20 años, hacen de este trastorno un desafío no sólo de salud, sino también social y económico, perjudicando tanto al afectado como al entorno que lo rodea. Si bien es cierto que prevalece principalmente en personas mayores, el envejecimiento de la sociedad hace que cada vez sean más las personas que padecen los síntomas que la demencia acarrea.

De acuerdo a la 10º edición de la Clasificación Internacional de Enfermedades (CIE-10), se entiende por demencia al “*síndrome debido a una enfermedad del cerebro, generalmente de naturaleza crónica o progresiva, en la que se producen alteraciones de múltiples funciones corticales superiores, entre ellas la memoria, el pensamiento, la orientación, la comprensión, el cálculo, la capacidad de aprendizaje, el lenguaje y el juicio.*” La conciencia no se ve afectada, mientras que el déficit cognitivo es generalmente acompañado, y en muchas ocasiones también precedido, de un deterioro en las capacidades de control social, emocional o motivacional [6].

La demencia produce un deterioro apreciable y usualmente paulatino en el funcionamiento intelectual. Para poder realizar un diagnóstico adecuado, se debe cumplir que haya evidencia de un déficit en la memoria y al menos una de las siguientes alteraciones cognitivas: afasia, apraxia, agnosia, o una falla en el funcionamiento ejecutivo. Asimismo, la discapacidad cognitiva debe ser lo suficientemente severa como para interferir con la independencia en las actividades cotidianas [7].

En lo que a la memoria refiere, los individuos suelen tener dificultades para aprender nuevo material, o bien para recordar lo previamente aprendido. Un método para determinar

el deterioro de la memoria es evaluando la capacidad de registro, retención y reconocimiento de la información. En el caso de información nueva, se suele dar una lista de palabras al individuo y se le pide que la repita (registro). Luego de un tiempo determinado, se le pide que las repita nuevamente en el orden dado (retención), y finalmente que las reconozca de una lista de múltiples palabras (reconocimiento). Para analizar la memoria remota, se le puede hacer preguntas personales o que recuerde material de interés aprendido en el pasado.

Se entiende por afasia a un deterioro en la función del lenguaje, en la que el individuo presenta dificultades para comunicarse mediante el habla, la escritura o la mimica, debido a lesiones en áreas cerebrales especializadas en estas funciones. La manera más sencilla de evaluar esta condición es pidiéndole al paciente que repita palabras o frases, que nombre elementos presentes o que siga instrucciones dadas.

La apraxia en cambio es la incapacidad de ejecutar las tareas motoras intencionadas y aprendidas con antelación como consecuencia de una lesión cerebral, pese a la capacidad física y a la voluntad para hacerlo [8]. Las alteraciones en las habilidades motrices se evalúan pidiéndole al individuo que ejecute determinadas tareas, desde actividades cotidianas como cepillarse los dientes hasta tareas más finas como ordenar bloques pequeños.

En cuanto a la agnosia, se la define como la incapacidad de procesar o reconocer información sensorial. Hay una dificultad en el reconocimiento de objetos, personas, sonidos, olores o tamaños, pero el sentido en sí mismo no presenta fallas, y tampoco hay un problema de pérdida de memoria. En estadíos avanzados de agnosia, los individuos pueden llegar a no reconocer a personas muy allegadas a ellos.

La disfunción en funcionamiento ejecutivo implica la pérdida, parcial o total, de la capacidad de pensar de manera abstracta, así como planificar, iniciar, secuenciar, controlar y detener el comportamiento complejo [7]. Comúnmente está relacionada con trastornos en el lóbulo frontal o las vías subcorticales asociadas (hipocampo, cerebelo, amígdala y ganglios basales). Se puede producir a su vez una deficiencia en el flujo de ideas y el individuo puede encontrar difícil atender más de un estímulo a la vez. La capacidad de abstracción se

puede medir haciendo que el individuo encuentre diferencias o similaridades entre palabras, así como pedirle que recite el abecedario o cuente secuencialmente.

En síntesis, para que un individuo pueda ser diagnosticado con demencia se debe cumplir que: (i) haya evidencia de una deficiencia en la memoria, y (ii) que se presente alguna de las siguientes alteraciones: afasia, apraxia, agnosia o falla en el funcionamiento ejecutivo. Estas deben ser lo suficientemente severas como para ocasionar una discapacidad ocupacional o social (e.g trabajar, realizar tareas domésticas, vestirse, manejar, o realizar cualquier otra actividad de la vida cotidiana), y deben evidenciar un deterioro del funcionamiento respecto a los niveles anteriores.

Una de las variables que se debe tener presente a la hora de realizar el diagnóstico es que, tanto la naturaleza como el grado de discapacidad, dependen en gran medida del entorno social y cultural. La interacción de factores genéticos, psicológicos, socioeconómicos, ambientales y demográficos pueden repercutir de diversas maneras en cada individuo, y es un desafío poder discernir qué es y qué no una demencia según el sujeto en gestión [9, 10]. Se cree que es un síndrome que está sub-diagnosticado, ya sea por falta de recursos (económicos y tecnológicos) o falta de concientización y educación al respecto.

2.1.1 Prevalencia e Incidencia

Se entiende por prevalencia a la proporción de individuos de una población y período determinado que presentan un evento determinado. De acuerdo a un estudio llevado a cabo en el 2016 por el Global Burden of Disease (GBD) [11], el número global de personas que padecían algún tipo de demencia alcanzaba los 43.8 millones (27 millones correspondientes a mujeres y 16.8 a hombres), sufriendo un incremento del 117% respecto a los datos estimados en 1990, 20.2 millones. Un reporte publicado por la Alzheimer's Disease International en 2015 [12] estimó que dicho número rondaba los 46.8 millones, una diferencia de 3 millones respecto a lo indicado por el GBD en el 2016. A pesar de la pequeña diferencia, ambos estudios coinciden en que la cantidad de individuos que padecen demencias prácticamente

se duplica cada 20 años, y estiman que para el 2050 los casos de demencia superarán los 100 millones.

Los factores que influyen en la diferencia de la prevalencia entre ambos reportes son, entre otras, el sesgo de medición entre países o regiones al realizar los censos (utilización de criterios diferentes para diagnosticar demencia dan diferentes resultados), la precisión de los estudios en los que se basan dichos reportes, que en parte depende del tamaño de la población analizada y de la exhaustividad del estudio en cuestión, y la falta de datos confiables en determinadas regiones o países, que conduce a realizar estimaciones menos certeras.

De todos modos, los estudios revelan que la prevalencia se ve afectada principalmente por la edad, aunque también hay diferencias significativas según el género, el grupo étnico y el nivel de ingreso de las poblaciones. A modo de ejemplo, la tabla 2.1 muestra datos que fueron recopilados del Reporte Mundial de Alzheimer 2015: El Impacto Global de las Demencias [12]. Se han incluido sólo algunas regiones, aquellas que contaban con mayor número de estudios de respaldo, y se trató de abarcar distintas áreas representativas.

Región	Género	60-64	65-69	70-74	75-79	80-84	85-90	>90	Prevalencia Estandarizada
Latinoamérica	M	1.4	2.4	4.3	7.4	12.6	21.6	43.7	8.41
	F	1.3	2.5	4.7	8.9	16.5	30.7	69.4	
EEUU	M	1.3	2.1	3.7	6.8	12.3	21.6	45.2	6.77
	F	1.0	1.8	3.3	6.4	12.5	23.2	52.7	
Europa Occidental	M	1.1	1.8	2.8	4.7	7.8	12.6	23.7	6.67
	F	2.0	3.2	5.2	8.7	14.6	23.7	45.1	
Sur de Asia	M	1.2	1.9	3.0	5.1	8.5	13.8	26.2	5.63
	F	1.6	2.5	4.0	6.7	11.2	18.1	34.3	
África sub-sahariana	M	1.0	1.5	2.3	3.8	5.7	9.2	17.5	5.47
	F	2.0	3.0	4.6	7.5	11.5	18.6	35.8	

Tabla 2.1: Prevalencia por región, género y edad según el Reporte Mundial de Alzheimer 2015.

Observando los datos se puede apreciar cómo la prevalencia aumenta conforme aumenta la edad, habiendo saltos significativos entre los grupos de edades. Esto se debe principalmente al deterioro de la función cognitiva que se da con el paso de los años. Hay que resaltar que la prevalencia en los grupos de mayor edad no sólo es mayor porque las demencias se presentan sobre todo en estadíos de edad avanzada, sino que además la proporción de

enfermos aumenta en el grupo dado que el tamaño del grupo disminuye. Es decir, se incrementa el número de afectados (matemáticamente, aumenta el numerador) y disminuye la población sobre la que se miden (se reduce el denominador) debido a que es menos la gente que llega a edades avanzadas.

Por otro lado, se puede distinguir una diferencia de prevalencia según el género. Salvando unos pocos grupos, la prevalencia de demencia en el sexo femenino supera a la masculina, y en algunos casos, la diferencia es muy acentuada. Sin embargo, se debe tener en cuenta que globalmente, según los datos obtenidos de la [Organización Mundial de la Salud \(OMS\)](#), la esperanza de vida de los hombres es menor que la de las mujeres. Mientras que el sexo masculino presenta una esperanza de vida de 69.1 años, la de las mujeres es de 73.8, una diferencia de 4.7 años [13]. Si indagamos en las regiones en cuestión, la esperanza de vida en América (incluye las regiones latinoamérica y EEUU) es de 74 y 79.9 en hombres y mujeres respectivamente, en Europa (incluye región Europa occidental) es de 73.2 y 80.2, en la región de Asia (incluye región sur de Asia) es de 67.3 y 70.7, y en África (incluye región África sub-sahariana) es de 58.2 y 61.7. Estas diferencias favorecen que la prevalencia sea mayor para el sexo femenino en grupos de edades avanzadas, puesto que la edad es el mayor factor de riesgo en las demencias [14].

No obstante, no hay evidencia suficiente para respaldar que la incidencia en mujeres sea mayor que en los hombres. Se define como incidencia al número de casos nuevos de una enfermedad en una población y período determinados. En cierta medida, la incidencia se encuentra estrechamente relacionada con el riesgo de desarrollar dicha enfermedad. Numerosos estudios han investigado la incidencia según el sexo y la edad, y no se ha encontrado una diferencia significativa entre hombres y mujeres en la proporción de quienes desarrollan algún tipo de demencia [14, 15, 16]. En otras palabras, aún cuando la prevalencia es mayor en mujeres que en hombres, en la incidencia parecería no haber una brecha marcada.

En cuanto a la etnicidad, a pesar de que hay datos que demuestran una mayor incidencia

y prevalencia de demencia en afroamericanos y blancos hispanos respecto a blancos no-hispanos [17, 18, 19], no parecerían ser los factores genéticos los que influyen en esta disparidad [20]. Son más bien las variantes en las condiciones de salud, estilos de vida y niveles socioeconómicos los que aportan a esta desigualdad.

Datos epidemiológicos sugieren que determinados factores de riesgo directamente asociados al desarrollo de demencia de tipo Alzheimer prevalecen en grupos afroamericanos e hispanos frente a los blancos no-hispanos. Enfermedades cardiovasculares (hipertensión, enfermedad de la arteria coronaria e infartos), que son más frecuentes en afroamericanos, así como también la diabetes, que prevalece tanto en afroamericanos como hispanos, tienen un peso considerable en la contribución de Alzheimer y otras demencias [19, 21]. Es esta relación la que vincula indirectamente a la etnicidad con la prevalencia e incidencia de demencias.

A nivel sociocultural y económico, un menor nivel de educación, mayor nivel de pobreza y menor ingreso, tienden a acrecentar la diferencia de prevalencia entre afroamericanos e hispanos contra los blancos no-hispanos. Censos y estadísticas confirman que el nivel de ingreso medio de los blancos no-hispanos supera en gran medida al de los afroamericanos e hispanos. Un estudio realizado por la Oficina de Censos de los Estados Unidos en 2018 ratificó que el ingreso medio de los blancos era de U\$S 67.937, mientras que el de los hispanos se situaba en U\$S 51.404, y U\$S 41.511 para los afroamericanos [22]. Mayor niveles de ingresos significa un mejor acceso a recursos médicos (ya sea para tratamiento o diagnóstico), y acceso a niveles de educación de mejor calidad, entre otras cosas.

Del mismo modo, en lo que a la educación respecta, hay una serie de hipótesis que explican por qué menores niveles educativos están relacionados con una mayor prevalencia e incidencia de demencia. En primer lugar, se cree que una educación deficiente viene acompañada de otros factores que influyen directamente en el riesgo de desarrollar algún tipo de demencia, tales como malnutrición, exposición a traumas, y atención médica inadecuada. Por otro lado, se cree que la educación a lo largo de la vida ayuda a promover la

memoria y mantener viva la actividad cerebral, construyendo y fortaleciendo estructuras neurobiológicas y patrones de comportamiento que subsanan los efectos de los agentes deteriorantes del cerebro [17]. Varios estudios han demostrado que personas con bajos niveles de alfabetización y/o una educación deficiente obtuvieron peores resultados en evaluaciones cognitivas, como en el **Mini-Mental State Exam (MMSE)**, uno de los tests más utilizados a nivel mundial para evaluar el deterioro cognitivo [19, 20, 23].

En contraposición, genéticamente hablando, a pesar de que no existen pruebas fehacientes que expliquen la desigualdad entre razas, numerosas investigaciones del genoma humano han desenmascarado una fuerte relación entre la **Apolipoproteína E (APOE)** y el riesgo de contraer demencias, particularmente la de tipo Alzheimer. El gen que codifica para la **APOE** puede tomar diferentes formas, siendo la $\epsilon 3$ la más común, seguida por la $\epsilon 4$, y de manera menos frecuente la $\epsilon 2$. Mientras que heredar la versión $\epsilon 2$ disminuye las posibilidades de contraer Alzheimer, la versión $\epsilon 4$ aumenta el riesgo. El genotipo **APOE** $\epsilon 2/\epsilon 2$ (dos copias del gen $\epsilon 2$) disminuye las probabilidades un 50%, en tanto $\epsilon 3/\epsilon 4$ (una copia de $\epsilon 4$) y $\epsilon 4/\epsilon 4$ (dos copias de $\epsilon 4$) aumentan 3 y 8 veces respectivamente el riesgo de padecer la enfermedad [24]. Uno de los motivos principales es que la isoforma $\epsilon 4$ desata la acumulación de la proteína beta-amiloide en el cerebro de manera más temprana que la $\epsilon 3$, y esta a su vez, antes que la $\epsilon 2$ [25].

En resumidas cuentas, son diversos los factores que juegan un rol importante en la prevalencia e incidencia de las demencias. La tendencia global muestra que la prevalencia va en aumento. El crecimiento del 117% que se produjo entre 1990 y 2016 respalda este hecho, y las estimaciones para el 2050 no parecen ser más optimistas. Esto se justifica con que la longevidad de las personas es cada vez mayor y con que cada vez hay más personas que alcanzan edades avanzadas. Este cambio es apreciable, y cada vez será más notorio, sobre todo en regiones en vías de desarrollo, que de a poco van aumentando la esperanza de vida, así como en países con alto crecimiento demográfico (e.g. China e India), donde cada vez son más quienes sobreviven hasta alcanzar edades reproductivas. En países desarrollados

también se puede percibir y se estima un aumento venidero pero de manera más moderada.

En un mundo donde el crecimiento demográfico no cesa y la esperanza de vida es cada vez más alta, el desafío que representan las demencias es un asunto a tratar de manera urgente. La carga económica y social que representan a nivel mundial es sustancial. Según estimaciones realizadas por Alzheimer's Disease International del costo de la enfermedad [12], en el 2010 rondaba los U\$S 604 mil millones, alcanzando U\$S 818 mil millones en el 2015, y proyectando 1 billón (10^{12}) para el 2018. El costo anual en dólares por persona para economías de ingresos bajos, medio-bajos, medio-altos, y altos en 2015 fue de 939, 3.865, 10.467 y 39.595 respectivamente.

Existe de este modo una dicotomía entre la distribución global de prevalencia, predominante en economías de ingresos medio-bajos, y los costos que están en su mayoría representados por economías de ingresos altos. Esto refleja menores costos por persona en economías de ingresos medio-bajos debido a menores ingresos salariales, acompañado de una mayor proporción de atención informal en lugar de cuidadores profesionales. En otras palabras, cada vez hay más casos de personas que padecen de demencia y no pueden costear un tratamiento adecuado.

Considerando la dimensión epidémica que está tomando la demencia, y dado que aún no existe una cura o tratamiento efectivo, es primordial abordar este conflicto activamente. Tratar de encontrar un diagnóstico temprano es fundamental para retrasar el desarrollo de la enfermedad así como aminorar los síntomas, siempre y cuando se reciba la estimulación y el cuidado necesarios.

2.1.2 Enfermedad de Alzheimer

Particularmente, dentro de los numerosos tipos de demencias que se han identificado, la demencia de tipo Alzheimer es la más común representando aproximadamente el 60% de los casos, seguida por la Demencia Vascular (DV) [26]. Se estima que hay otro 20% más que surge de las demencias mixtas, caracterizadas por tener más de una causa. Comúnmente

hay cuatro tipos de estas y son en primer lugar Alzheimer combinado con **DV**, seguido por Alzheimer con **Demencia con Cuerpos de Lewy (DCL)**, Alzheimer con tanto **DV** como **DCL**, y en menor medida, **DV** con **DCL** [21]. El Alzheimer es dentro de las enfermedades cerebrales degenerativas el mayor causante de demencia.

Para que la demencia sea de tipo Alzheimer, se debe cumplir junto con los requisitos que caracterizan a las demencias previamente mencionados en la sección 2.1 (i) y (ii), que: (iii) el progreso de la enfermedad sea de inicio gradual y con un deterioro cognitivo continuo; (iv) las deficiencias cognitivas no sean fruto de (a) otras fallas en el sistema nervioso que afecten progresivamente la memoria y cognición, (b) fallas a nivel sistémico que se conoce que pueden generar demencia (e.g hipotiroidismo, deficiencia de ácido fólico o vitamina B12, hipercalcemia, etc), (c) condiciones inducidas por sustancias; (v) las deficiencias no se produzcan exclusivamente durante el curso de un delirio; y (vi) la discapacidad no se describa mejor con otro trastorno [6, 7].

En cuanto a los síntomas que se pueden presentar en la **EA**, cabe resaltar que varían según cada individuo y que van evolucionando conforme la enfermedad progride. En estadíos tempranos, lo primero que se suele notar es que el sujeto tiene dificultades para aprender nueva información. Esto se debe a que las primeras neuronas en verse afectadas son aquellas que se encuentran en regiones del cerebro que cumplen la función de formar nueva memoria [21].

Otros síntomas que pueden advertir sobre la posible presencia de **EA** incluyen cambios en la memoria que dificultan las actividades de la vida cotidiana; problemas a la hora de planificar o resolver asuntos; dificultad para llevar a cabo tareas habituales en la casa, trabajo o tiempo libre; desorientación espacio-temporal; problemas en la comunicación por medio oral o escrito; colocación de objetos en lugares que no son los correspondientes; disminución del juicio; y cambios en el humor o personalidad [27].

En cuanto a los síntomas neuropsiquiátricos, se destacan apatía (falta de interés, motivación o entusiasmo), ansiedad, cambios en el apetito, alteración en el sueño, irritabilidad,

agresividad, y depresión. En menor medida, se han encontrado casos en los que el individuo presenta una actividad motora entorpecida o puede tener cuadros de euforia [28, 29].

El diagnóstico clínico se obtiene a partir de un examen físico, análisis de laboratorio, imágenes del cerebro y un examen mental. El estado mental se evalúa mediante pruebas neuropsicológicas, tales como el MMSE que está enfocado en la orientación y memoria principalmente, o el Montreal Cognitive Assessment (MoCA) que evalúa una amplia gama de habilidades cognitivas. Otra evaluación que está avanzando en la práctica es el "*test del reloj*", que consiste en dibujar un reloj a modo de medir la disfunción espacial y negligencia. No obstante, si bien hay algunos biomarcadores que se miden en líquido cefalorraquídeo (LCR), así como estudios histológicos ya sea en biopsias o autopsias, en la práctica el diagnóstico es exclusivamente clínico, acompañado de tests mentales e imágenes, y no son concluyentes a la hora de determinar la presencia de Alzheimer.

Del mismo modo, es un desafío poder diferenciar a los sujetos con deterioro cognitivo leve (Mild Cognitive Impairment (MCI)) que desarrollarán EA en el futuro respecto a quienes no lo harán. Los síntomas en estadios tempranos de EA pueden ser iguales o muy similares a los que presentan individuos con MCI, pero es sustancial encontrar un método que permita discernir unos de otros. Distintos estudios, tanto post mortem como in vivo, han logrado obtener información que describe características típicas de los cerebros afectados con Alzheimer.

En este sentido, si bien su etiología es desconocida, la EA presenta características neuroquímicas y neuropatológicas específicas. Está asociada a una pérdida progresiva de neuronas debido a una acumulación de proteínas beta-amiloïdes (placas amiloïdes) por fuera de las neuronas cerebrales, y a una acumulación de proteína tau (ovillos neurofibrilares) dentro de las neuronas, que conduce al entorpecimiento de la función neuronal [30].

Por otro lado, es sabido que en la EA el núcleo basal colinérgico de Meynert se ve afectado en gran medida. Esta estructura es la fuente principal de aportes colinérgicos que estimulan la acción de los nervios parasimpáticos postganglionares [31]. La degeneración se

puede apreciar mediante autopsias (post y ante mortem) que revelan la presencia de proteína Tau y una reducción en la densidad sináptica y en el conteo de neuronas presentes [32].

Gracias a los avances tecnológicos en imagenología médica, se han descubierto otros patrones que suelen estar presentes en cerebros con EA. IRM han demostrado una disminución significativa de la materia blanca, correlacionada con la pérdida de axones mielinizados y la denudación del epéndima ventricular [33]. Las células ependimarias cumplen un rol importante en la producción y regulación del LCR, y la denudación de estas hace que el mismo se disperse en la cavidad craneal.

Asimismo, numerosos estudios post mortem han indagado en la atrofia de la formación hipocampal y de la corteza entorrial (CE), y se ha evidenciado una reducción considerable en el volumen de ambas en pacientes con EA [34]. Estas estructuras están estrechamente ligadas con la memoria y la orientación espacio-temporal, y son las primeras en verse afectadas en el transcurso de la enfermedad. Investigaciones in vivo con IRM confirmaron que las estructuras del lóbulo temporal medial se ven reducidas simétricamente con mayores grados de atrofia en la CE, seguido por el hipocampo y la amígdala [35, 36, 37].

Sin embargo, los estudios parecerían indicar que esta degeneración comienza antes de que aparezcan los síntomas y se obtenga el diagnóstico. Individuos levemente afectados pueden presentar una atrofia con reducciones de la CE de hasta un 20-30% y del hipocampo de un 15-20%, lo que sugiere que la enfermedad lleva un tiempo progresando sin ser percibida [38].

Muchos científicos están orientando sus esfuerzos a la búsqueda de una técnica que permita realizar un diagnóstico temprano de Alzheimer. El carácter de irreversibilidad, la ausencia de un tratamiento efectivo y la magnitud epidemiológica que está adquiriendo, la convierten en objeto de estudio de múltiples centros de investigación que han ido explorando distintos métodos para extraer patrones de los datos de neuroimágenes.

En la actualidad se cuenta con varias técnicas de adquisición de imágenes, y la elección de la misma dependerá del problema que se quiera abordar. En reglas generales, hay que

contemplar el valor diagnóstico, y ponerlo en la balanza con la seguridad del paciente y el costo del estudio. Mientras que la tomografía computada es un estudio rápido y de gran disponibilidad, para una cuestión tan trascendente como es el caso del Alzheimer, la resonancia magnética resulta ser la técnica más elegida por los profesionales debido a su gran sensibilidad a la hora de examinar la estructura del parénquima cerebral y los tractos de materia blanca, los límites entre materia gris-blanca, la fosa posterior y el tronco encefálico. La tomografía computada suele ser utilizada en traumatismos, tumores o como una primera aproximación a un posible diagnóstico, más aún la resonancia es mayormente aplicada en estudios cerebrales, de articulaciones, y de corazón entre otros.

Aún así, es importante mencionar las diferencias entre las diferentes metodologías, detallando las ventajas y desventajas de una sobre a la otra.

2.2 Imagenología en Demencias: Resonancia Magnética

La **resonancia magnética (RM)** es una poderosa herramienta de adquisición de imágenes médicas que ha adquirido mucha popularidad debido a su alta capacidad de resolución y diferenciación entre tejidos de diferentes densidades. A diferencia de la **tomografía computarizada (TC)**, en lugar de emplear rayos X, aplica campos magnéticos y señales de radiofrecuencia para obtener información anatómica en cualquier dirección deseada [39].

Esta técnica parte de la base de que el átomo de hidrógeno, presente en todo el cuerpo humano, posee una masa atómica impar (sólo cuenta con un protón) que le proporciona un momento magnético y espín. En condiciones normales, dichos momentos se alinean aleatoriamente dando como resultado un momento magnético neto cero. En presencia de un campo magnético externo los espines pueden alinearse de manera paralela o antiparalela en la misma dirección que este generando un momento magnético M_0 distinto de cero. Si en ese momento se aplica un pulso de **radiofrecuencia (RF)** adecuado, los átomos que perciban dicho pulso se excitarán a niveles de energía más altos.

Acto seguido, los átomos volverán a su estado basal de energía emitiendo energía

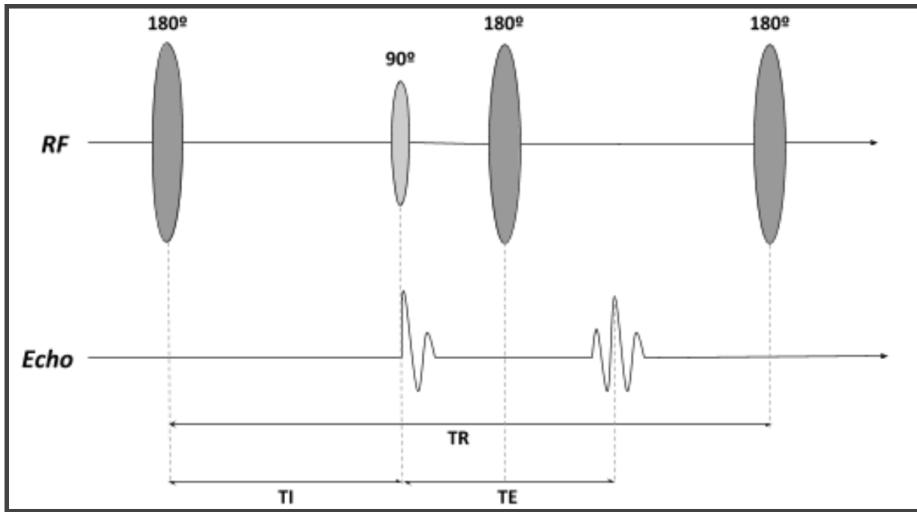


Figura 2.1: Diagrama de los tiempos involucrados en RM [41].

electromagnética de la misma frecuencia que el pulso aplicado. Unas antenas receptoras (bobinas) que se colocan sobre la zona de interés son las encargadas de captar dicha energía para luego generar un mapa de densidades de distribución de hidrógeno a través de algoritmos matemáticos. Utilizando tres campos magnéticos de gradiente lineal superpuestos se puede lograr cortes en todas las direcciones (axial, sagital y coronal). Con la transformada de Fourier luego se convierte la información de la frecuencia contenida en la señal para cada punto del plano considerado y se crea un mapa de intensidades en escala de grises [40].

La variación en la secuencia de pulsos de RF afecta el tipo de imagen que se obtiene a la salida. "Repetition Time" (TR) es el tiempo que transcurre entre secuencias de pulsos sucesivos sobre el mismo corte. "Time to Echo" (TE) es el tiempo entre la emisión del pulso y la recepción de la señal por la antena. Según los valores en los que se los configure, al tejido se lo puede caracterizar según dos tiempos de relajación: T1 y T2. Algunas secuencias también cuentan con un tiempo "Inversion Time" (TI), tiempo que se tarda desde que se lanza el pulso de 180° hasta que se lanza el pulso de excitación de 90° . En la imagen 2.1 se puede ver cómo serían los tiempos mencionados en una secuencia con TR, TE y TI.

El tiempo T1, también conocido como tiempo de relajación longitudinal, es una constante de tiempo que determina el tiempo que requiere el vector de magnetización del sistema en

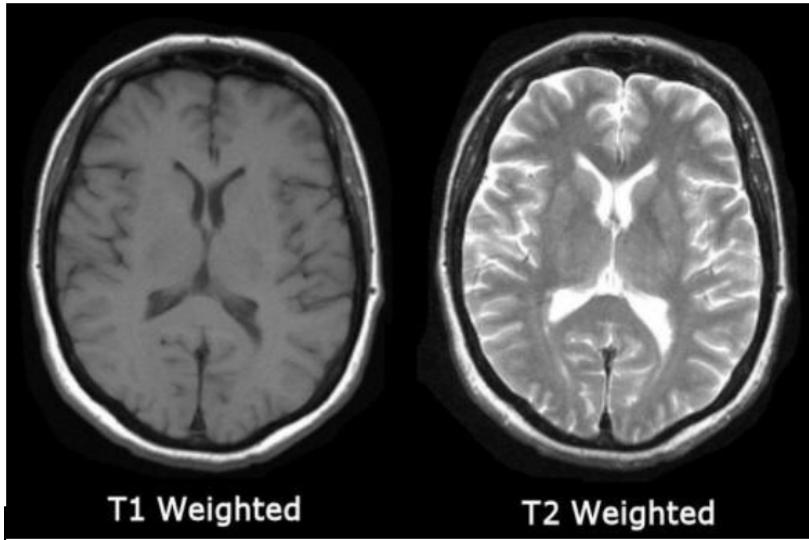


Figura 2.2: Imagen de RM de tipo T1 (izquierda) vs. T2 (derecha).

volver al 63% de su magnitud inicial. Es una medida que refleja el tiempo que tardan los protones en realinearse con el campo magnético externo, y depende en gran medida de la intensidad de este. Las imágenes potenciadas en T1 requieren tiempos TR y TE cortos (en general <500 ms y <30 ms respectivamente). Largos períodos de TR minimizan los efectos T1 puesto que los tejidos no tienen tiempo suficiente para recuperarse de la excitación.

T2, o tiempo de relajación transversal, es el tiempo que requiere la componente transversal de magnetización en llegar al 37% de su magnitud inicial luego de la excitación. Los tiempos de TR y TE en este tipo de imágenes son más largos (>1500 ms y >90 ms respectivamente). Períodos cortos de TE minimizan los efectos T2 dado que no se le dio el tiempo suficiente para que los protones excitados decaigan, y por ende, los tejidos no se logren diferenciar bien [42].

En imágenes cerebrales, T1 y T2 se pueden diferenciar fundamentalmente por el brillo que toma el LCR. Mientras que en T1 se ve más bien oscuro/negro, en las imágenes T2 se ve brillante/blanco. Además, la materia blanca se ve clara en T1 y en T2 más bien gris oscuro. Las infecciones o desmielinizaciones son oscuras en T1 y brillantes en T2 [40]. En la imagen 2.2 se puede comparar una resonancia magnética de tipo T1 contra una T2.

En resumidas palabras, la RM parte del principio que los protones tienen un momento

angular que se polariza en presencia de un campo magnético. Un pulso de RF puede alterar el estado de energía de los protones, y cuando se apaga, regresan a su estado de energía basal emitiendo una señal de RF que es capturada por unas antenas que decodifican la señal. Combinando de diversas maneras los gradientes y pulsos, se pueden diseñar numerosas secuencias para adaptarse a diferentes características del tejido. En la EA particularmente, la RM estructural se puede orientar a la evaluación de la atrofia o volúmenes (principalmente utilizando RM potenciada en T1), o a los cambios en las características del tejido que generan alteraciones en las señales reflejándose como hiperintensidades de la materia blanca (RM potenciada en T2) [38]. Algunas secuencias muy utilizadas hoy en día son RM por difusión, secuencia FLAIR (Fluid-attenuated inversion recovery) para imágenes potenciadas en T2, o MP-RAGE (Magnetization Prepared Rapid Acquisition Gradient Echo) en imágenes potenciadas en T1, entre otras.

Una de las ventajas que ofrece la técnica de adquisición de imágenes por RM frente a TC es que no hay exposición a ningún tipo de radiación ionizante, minimizando los efectos biológicos nocivos en el paciente. En cuanto a las imágenes obtenidas, la TC es muy útil para distinguir huesos y anomalías en tejidos, más aún la RM es fundamentalmente aplicada en tejidos blandos, como ligamentos, tendones, cerebro y tumores. Sin embargo, una desventaja de la RM es que los resultados suelen ser más ruidosos debido a que la adquisición de la imagen es un proceso que puede durar desde 10-15 minutos hasta 2 horas según qué tejidos se analicen, y dependen en gran medida del movimiento del individuo.

En las últimas décadas ha ido mutando el uso que se le da a las neuroimágenes en demencias. Tradicionalmente, han sido utilizadas principalmente como apoyo a la hora de descartar causas conocidas y tratables. En la mayoría de las patologías neurodegenerativas el diagnóstico es aún exclusivamente clínico y se basa en la amamnesis (información que proporciona el paciente durante la entrevista clínica), síntomas y signos. Consecuentemente, un posible tratamiento es recién posible cuando las alteraciones neuronales son lo suficientemente extensas como para que los síntomas sean perceptibles.

Actualmente, con el avance de la investigación científica y el desarrollo tecnológico, se ha extendido notablemente el rol de las neuroimágenes hacia aplicaciones en etapas diagnósticas y pronósticas. La necesidad de realizar un diagnóstico precoz para evitar el deterioro irreversible que se suele percibir en el momento de la detección, ha orientado los esfuerzos hacia el desarrollo de avanzadas técnicas que permitan identificar la enfermedad en períodos preclínicos donde el cuadro es todavía asintomático.

En particular, en patologías cerebrales el uso de esta metodología se ha propagado considerablemente gracias a su capacidad de detectar atrofia cortical. Las estimaciones en la pérdida de tejido en áreas tales como el hipocampo o corteza entorrinal han demostrado ser buenos indicadores del desarrollo de deterioro cognitivo leve en Alzheimer, diferenciándose de otras neuropatologías [34, 35, 36, 37, 38, 43, 44].

La función de las imágenes en medicina ha ido mutando a lo largo de los últimos años gracias a los nuevos usos y tecnologías que han surgido. Mientras que en primera instancia los profesionales las empleaban como herramientas para confirmar diagnósticos o como soporte para la comprensión de ciertas enfermedades, hoy en día el potencial que suponen es sumamente prometedor. La inteligencia artificial es uno de los propulsores más emblemáticos, ayudando en etapas preclínicas donde el ojo humano no es suficiente para detectar anomalías. Diferentes estudios han revelado que la aplicación de algoritmos puede ser más eficiente incluso que los especialistas en imágenes en la predicción de enfermedades neurodegenerativas tal como el Alzheimer [45, 46]. La importancia radica en encontrar métodos generalizables, rápidos y precisos capaces de discernir entre distintos tipos de trastornos o alteraciones.

Para poder lograr esto, es necesario generar una herramienta capaz de clasificar las imágenes de entrada de acuerdo a la condición clínica de cada individuo. En este caso específicamente, se debe crear un clasificador que distinga a individuos con condiciones normales de aquellos que presentan Alzheimer.

Un paso crítico al que se le debe dar suma importancia es el preprocesamiento de los

datos de entrada. Se entiende por preprocesamiento a la limpieza, integración y reducción de los datos, y es una transformación previa a alimentar el algoritmo de clasificación. El fin de esta etapa es poder extraer los parámetros principales o dominantes de las imágenes que la hacen característica (*feature extraction*). La reducción del ruido, segmentación, registro y redimensionamiento son algunos pasos fundamentales a la hora de extraer información sustancial. En imágenes neuronales también es aconsejable eliminar regiones no-cerebrales, como puede ser el cuello y cráneo.

En lo que a la clasificación respecta, diferentes dominios de la matemática, estadística e informática deben ser considerados. El clasificador es precisamente una red neuronal. La manera más simple de entender cómo funciona es considerarla como una gran función que se aplica a los datos de entrada para devolver una salida categórica. En este caso la salida sería más bien binaria: la EA está presente o no.

Un ejemplo de como podría funcionar una red es suponer que se quiere discernir imágenes donde hay tanto rosas rojas como jazmines blancos. El paso previo al clasificador sería hacer un preprocesamiento adecuado para extraer las características principales, mediante por ejemplo un análisis de las intensidades (no tienen el mismo valor los píxeles donde predomina el rojo que aquellos donde abunda el blanco). Usando umbrales, promediando, o aplicando otras operaciones se puede lograr obtener información sustancial representativa de cada imagen, que luego será utilizada por el clasificador a la hora de aprender a separarlas.

En el campo de la medicina, las redes neuronales son un poderoso recurso que está siendo explotado cada vez en mayor medida. En el diagnóstico de enfermedades neurodegenerativas especialmente, todavía los hallazgos no han alcanzado los estándares necesarios, y es necesario continuar con las investigaciones y evaluaciones.

2.3 Redes Neuronales

En una era donde la recopilación y almacenamiento de datos es cada vez más accesible a nivel mundial, el desarrollo de técnicas robustas y eficientes de IA

cumple un rol primordial en la optimización de procesos. La IA es un término que se utiliza para describir la capacidad de las computadoras y tecnología para simular el comportamiento inteligente humano. A lo largo de los años ha ido adquiriendo un papel muy importante en la investigación científica y en la resolución de problemas complejos en diversas aplicaciones, especialmente en áreas donde abundan los datos pero faltan herramientas o modelos que faciliten su uso.

Las redes neuronales (RRNN) son una subclase dentro de la IA. El Dr. Robert Hecht-Nielsen, creador de la primera neurocomputadora, las definió como “un sistema computacional hecho de un número de elementos simples e interconectados que procesan información a través de una respuesta dinámica a entradas externas” [47]. Las RRNN pretenden hacer uso de los conocimientos que hay acerca de la funcionalidad de los sistemas biológicos para modelar sistemas no-lineales en los que la relación entre las variables es desconocida o muy compleja, tales como reconocimiento de patrones, predicción, optimización, asociación, aproximación de funciones y control.

Inspiradas en el sistema nervioso, idealmente las características que deben imitar las RRNN son [48]:

1. alto grado de paralelismo: permite un procesamiento rápido de la información.
2. tolerancia a fallas: mayor adaptabilidad al ruido, brindando un mejor desempeño.
3. capacidad de aprender y manejar información imprecisa: permite que el sistema actualice su estructura interna en respuesta a un entorno cambiante.
4. no-linealidad: mejor ajuste a los datos.
5. capacidad de generalizar: posibilita que el modelo se pueda utilizar con datos no aprendidos.

La capacidad de aprender es una de las características más atractivas de las RRNN. En este proceso la red es capaz de modificar su arquitectura, ya sea ajustando los pesos, quitando

o estableciendo conexiones, o bien cambiando funciones de activación. El aprendizaje se hace de manera iterativa, y el rendimiento suele mejorar con el número de iteraciones [49].

Existen dos métodos de aprendizaje: supervisado y no-supervisado. Se considera aprendizaje supervisado a aquel en el que para cada patrón de entrada, hay una respuesta correcta determinada. Es decir, hay un agente externo que controla (supervisa) la respuesta que debería generar la red a una entrada determinada, a modo que los pesos se actualicen para acercarse al valor correcto. En cambio, en aprendizaje no-supervisado no hay una respuesta correcta asociada a cada entrada. La red explora los datos y su estructura o correlaciones, para luego organizar los patrones en categorías que surjan a partir de ellos.

En cuanto a la estructura, una red neuronal está formada por una serie de “neuronas” (o nodos) interconectados que se organizan en distintas capas. Cada unión entre nodos tiene un peso determinado que varía según la fuerza de la conexión. La estructura básica consiste en una capa de entrada, una o varias capas ocultas, y una capa de salida. El número de nodos en cada capa dependerá de la complejidad del problema a resolver.

Los nodos en la capa de entrada reciben información externa y la transfieren a la primera capa oculta, y ésta a su vez a las capas ocultas subsiguientes, hasta llegar a la capa de salida. Cada nodo es una unidad de procesamiento en la que se aplican funciones matemáticas a la información recibida mediante una función de entrada, una de activación, y una de salida. La función de entrada suele ser una sumatoria en la que se suman cada valor de entrada multiplicado por su peso correspondiente. La función de activación transforma la salida de la función de entrada en un valor (estado) de activación. Usualmente los rangos utilizados van de (0 a 1) o (-1 a 1). Finalmente, la función de salida determinará qué valor se transfiere a los nodos conectados, y generalmente depende de un umbral. Si el valor está por debajo del umbral, ninguna salida se transfiere al nodo siguiente.

La figura 2.3 ilustra un esquema básico de una red neuronal. Los nodos representados por X_i representan la capa de entrada, y cada nodo tiene sus pesos asociados representados por W_i (input) que caracterizan la fuerza de la unión con los nodos de la primera capa oculta.

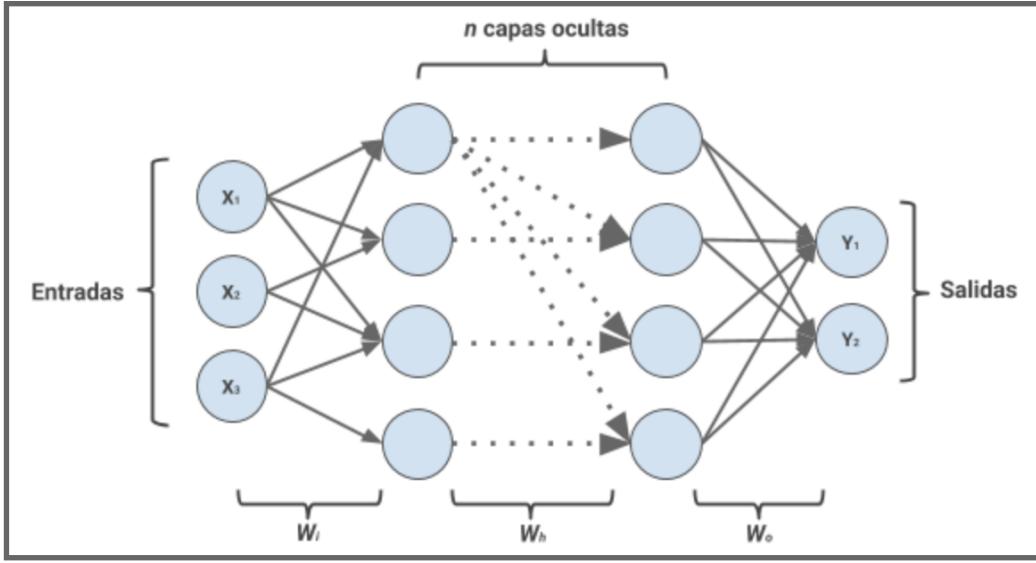


Figura 2.3: Diagrama de una red neuronal básica con n capas ocultas, 3 nodos de entrada, y 2 de salida.

En las n capas ocultas, las conexiones entre nodos también tienen pesos característicos, representados en el esquema como W_h (hidden). Finalmente, de la última capa oculta salen los pesos W_o (output) que determinan la unión con los nodos a la salida, Y_i .

Matemáticamente, si se toma un nodo (figura 2.4), la salida se puede expresar como:

$$y = f\left(\sum_{i=0}^n x_i \cdot w_i - \theta\right) \quad (2.1)$$

donde n representa el número de entradas del nodo, w_i es el peso asociado a la entrada x_i , y θ es un valor asociado al nodo conocido como “bias”. La función f es la función de activación, generalmente la función signo, la función sigmoidea, o la función umbral.

Por otro lado, según como sean las conexiones inter-nodales, las **RRNN** se pueden clasificar en redes feed-forward, en las que las señales van siempre hacia capas posteriores (hacia adelante), o recurrentes (RNN de su sigla en inglés Recurrent Neural Networks) que permiten conexiones tanto para adelante como hacia atrás. Teniendo esto presente, existen diversas topologías de redes que dependen fundamentalmente del número de capas, número de nodos por capa, el grado de conectividad y el tipo de conexiones entre nodos. Algunos ejemplos de topologías feed-forward son perceptrón simple o multicapa, redes de función

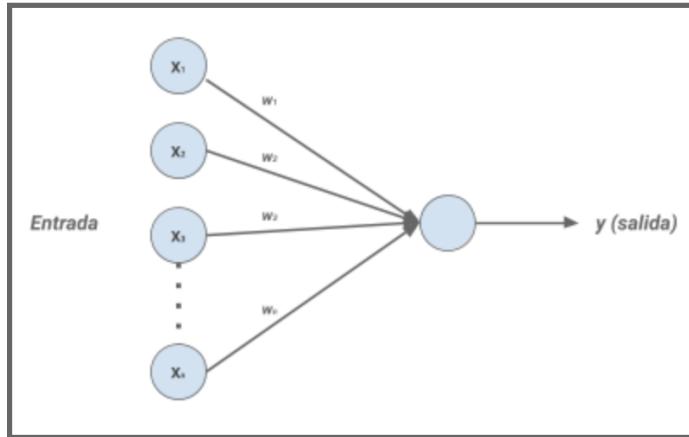


Figura 2.4: Diagrama de un nodo con n entradas y sus pesos asociados.

de base radial, o **CNN**. Las redes de Hopfield, mapas auto-organizados y redes competitivas son algunos ejemplos de redes recurrentes.

2.3.1 Redes Convolucionales

Las **CNN** son un tipo específico de redes neuronales feed-forward que se basan principalmente en algoritmos de aprendizaje profundo, caracterizándose por tener una arquitectura con numerosas capas ocultas [50]. Es precisamente esta arquitectura la que las convierte en una poderosa herramienta en reconocimiento y clasificación de imágenes. En medicina han ido adquiriendo un rol protagónico en imagenología especialmente en el diagnóstico de enfermedades, así como en el reconocimiento de objetos tales como tumores, quistes o nódulos [51, 52, 53].

La arquitectura típica de las **CNN** consiste en una capa de entrada, seguida de numerosas capas ocultas de diversos tipos, y una capa de salida. En las capas ocultas se encuentran numerosas capas convolucionales, ‘pooling’ y ‘fully connected’ (completamente conectadas).

La función de las capas convolucionales es extraer parámetros (tales como bordes, formas o colores) de las imágenes de entrada mediante una convolución entre estas y unos filtros que se desplazan a lo largo de cada imagen. La convolución es una operación lineal

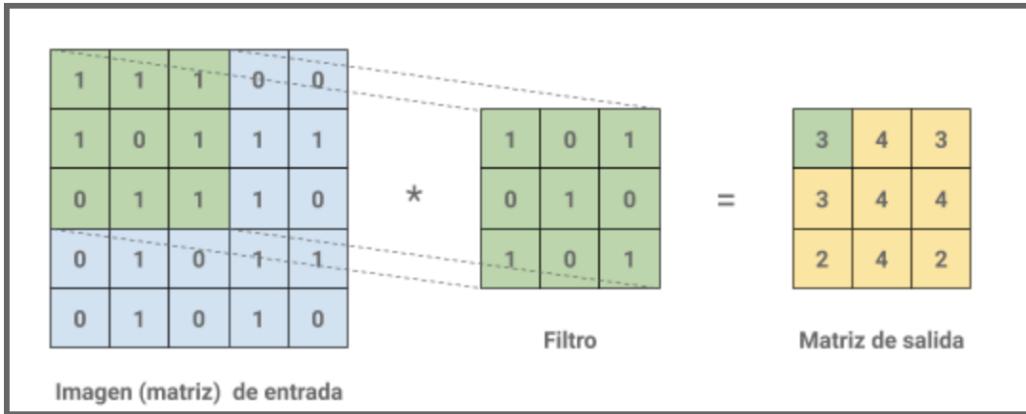


Figura 2.5: Ejemplo de una convolución aplicada a un tensor de entrada de 5×5 con un filtro de 3×3 .

en la que una matriz llamada *filtro* se aplica a la matriz de entrada correspondiente a la imagen, llamada *tensor*. Se realiza una “multiplicación inteligente” (element-wise product) entre cada elemento del filtro y del tensor para cada región que ocupe el filtro sobre el tensor, y se suman los valores para formar una matriz de salida conocida como *mapa de características* [54].

Si se tiene una imagen de dimensión $h \times w \times d$ (*height* - altura, *width* - ancho, *depth* - profundidad), y un filtro de $f_h \times f_w \times f_d$, a la salida se obtendrá un mapa de características de dimensión $(h - f_h + 1) \times (w - f_w + 1) \times 1$. Por ejemplo, si en la entrada se toma una imagen de $5 \times 5 \times 1$ y se aplican filtros de $3 \times 3 \times 1$, los mapas de patrones a la salida tendrán una dimensión de $3 \times 3 \times 1$. En la figura 2.5 se ilustra el ejemplo anterior para una imagen cuyos valores de píxeles por simplicidad valen 0 o 1, y donde la profundidad es 1. Generalmente la profundidad de las imágenes es 3, dado que corresponde a los 3 canales RGB.

La distancia entre dos filtros sucesivos se lo llama *paso* (*stride* en inglés). Este parámetro se lo puede fijar en distintos valores. Por lo general se lo fija en 1, aunque muchas veces se toman pasos mayores para lograr una reducción de dimensión. Otra forma para conseguir esto es mediante la capa *pooling*, como se describe más abajo.

Por otro lado, para permitir que los elementos más céntricos de los filtros se superpongan con los elementos más externos del tensor, existe lo que se llama *padding*. Esta técnica agrega filas y columnas a cada extremo del tensor, generalmente con ceros (*zero-padding*), a

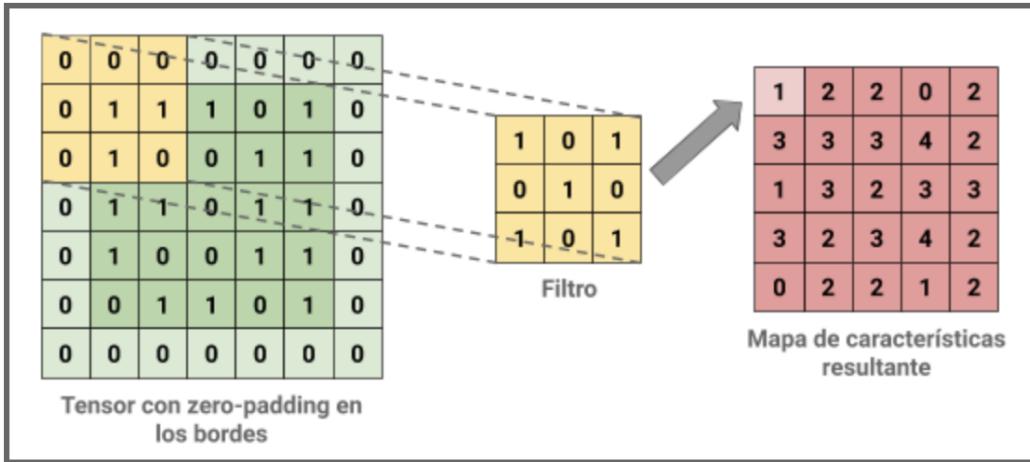


Figura 2.6: Ejemplo de zero-padding aplicado a un tensor de entrada de 5×5 . De esta forma no varía la dimensión del tensor a la salida.

modo que el centro del filtro logre superponerse con los elementos exteriores del tensor, tal como se puede ver en la figura 2.6. Nótese que sin aplicar zero-padding la dimensión del tensor se vería reducida cada vez que se realice una convolución.

Seguido de cada capa convolucional se aplica una operación no lineal, generalmente la función *ReLU* (de su sigla en inglés *Rectified Linear Unit*) que consiste en una operación no lineal en la que todos los píxeles con valores negativos pasan a valer 0. El objeto de esta operación es justamente introducir la no-linealidad dado que la información del mundo real suele ser no lineal. Otras funciones que se suelen emplear son la sigmoidea o tangente hiperbólica.

La capa ‘pooling’ se utiliza para reducir las dimensiones de las imágenes y así acelerar el tiempo de procesamiento computacional. El objetivo es extraer los parámetros principales de las matrices resultantes de las capas anteriores, tomando un tamaño de ventana o filtro determinado. Si se toma un *stride* igual al tamaño del filtro pooling, entonces no habrá superposición de regiones, mientras que si se toma un *stride* menor al tamaño del filtro, se logrará una superposición y la reducción de dimensión será menor. Los valores obtenidos dependerán del tipo de pooling que se aplique. En general, se extraen los valores máximos de cada ventana, pero hay casos donde conviene sacar el promedio o la suma. En el ejemplo

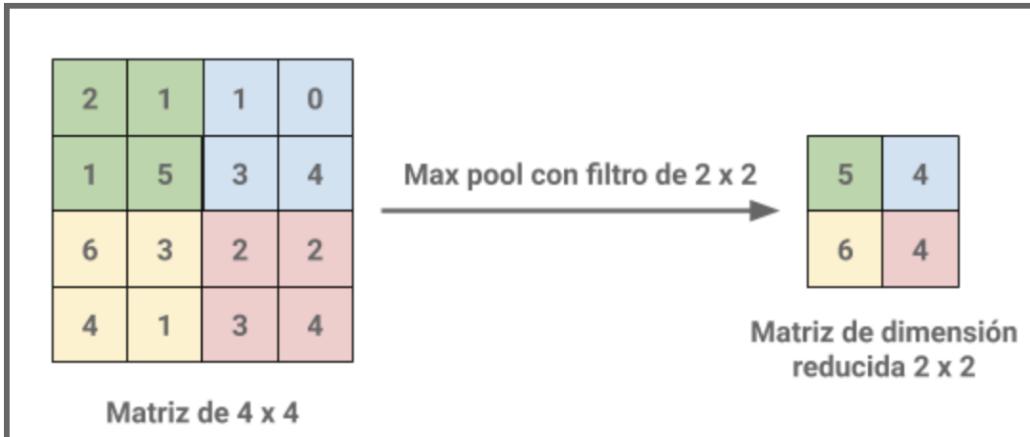


Figura 2.7: Aplicación de *max-pooling* de 2×2 y un stride de 2. No hay superposición de regiones. Se puede ver la reducción en la dimensión de la imagen de entrada.

de la figura 2.7 se puede ver como se reduce la dimensión tomando una capa pooling de valores máximos de 2×2 con *stride* 2.

Los mapas de características de la última convolución o de la última pooling son luego transformados en un arreglo de elementos unidimensional que se conecta con una capa densa completamente conectada (*fully connected*) en la que cada entrada se comunica con cada salida mediante un peso que se puede aprender [54]. El propósito de esta capa es usar las características más representativas obtenidas en los mapas de características para clasificar las imágenes en diferentes clases. La última capa densa tiene usualmente la misma cantidad de nodos que de clases a la salida.

La función de activación de la última capa densa debe ser propiamente elegida en función al problema de clasificación en cuestión. Para problemas de clasificación múltiple, donde se deben distinguir numerosas categorías, se suele aplicar la función ‘softmax’ que normaliza los valores de salida según las probabilidades de pertenecer a cada clase. A cada clase le corresponde un valor entre 0 y 1, y la suma de probabilidades da 1. En problemas binarios, en general se emplea la función sigmoidea.

Finalizado el proceso de modelado de la red, se debe evaluar su desempeño con los datos que se dispongan. Estos por lo general se dividen en tres conjuntos: datos para entrenamiento (*training set*), datos para testeо (*testing set*), y datos de validación (*validation*

set).

2.3.1.1 Entrenamiento de la Red

Una vez constituida la arquitectura de la red, se debe proceder a entrenarla con los datos previamente separados para ello. El objetivo de esta etapa es iterar con el training set sobre la red a modo de minimizar la diferencia entre las predicciones y las etiquetas reales. El entrenamiento consta de dos fases: una fase hacia adelante y una hacia atrás.

En la primera, los datos de entrada atraviesan todas las capas del modelo hacia adelante y se calcula el rendimiento para determinados filtros y pesos mediante una función de costo (*loss function*). Las funciones de costo comúnmente usan técnicas como entropía cruzada, error cuadrático medio (**MSE** de su sigla en inglés Mean Squared Error) o error absoluto medio (**MAE** de su sigla en inglés Mean Absolute Error). La primera es mayormente aplicada en problemas de clasificación, en tanto las últimas dos en problemas de regresión. De esta fase se obtienen los costos (*loss values*).

En la segunda fase se actualizan hacia atrás los parámetros aprendibles, i.e. filtros y pesos, de acuerdo a lo obtenido en la fase anterior. Los métodos de propagación hacia atrás (*backpropagation*) y gradiente descendente son frecuentemente utilizados en conjunto para optimizar tales parámetros.

2.3.1.2 Validación y Testeo del Modelo

La validación es el proceso mediante el cual se evalúa el modelo durante el entrenamiento, con un conjunto de datos diferente al utilizado para entrenarlo. Se utiliza el set de validación para ajustar los hiperparámetros tales como número de nodos, cantidad de capas ocultas, o tipos de funciones de activación. El objeto es decidir si los resultados que cuantifican las relaciones entre las variables describen correctamente los datos. De aquí se puede estimar que tan bien generaliza el modelo durante el entrenamiento. Cabe remarcar que ni los pesos ni los filtros son actualizados en esta etapa.

Una de los motivos por el cual es conveniente aplicar el set de validación es para evitar que el modelo se sobreajuste a los datos de entrenamiento (*overfitting*). Un modelo sobreajustado obtiene muy buenos resultados en la clasificación de los datos de entrenamiento, pero cuando se agregan datos desconocidos para la red, el desempeño se reduce notablemente dado que no es capaz de generalizar.

Hay diferentes métodos para validar el modelo. El más simple es, previo a entrenar la red, tomar un porcentaje de los datos y separarlos para validar. A este se lo conoce como “*Holdout*”. Una desventaja de esto es que se pierden datos de entrenamiento que podrían ser muy útiles e incluso representativos, y aumenta el riesgo de generar un modelo sub-ajustado.

La finalidad del conjunto de testeo o de prueba es evaluar la performance de la red una vez que ya fue entrenada. Los datos son diferentes a los de los conjuntos de validación y entrenamiento, y son desconocidos para la red. La idea es proporcionarle al modelo datos imparciales de forma tal que los resultados en la predicción y exactitud sean más certeros, brindando información acerca de la capacidad de predicción y capacidad de generalización.

2.4 Evaluación del modelo

Para poder comprender mejor la calidad del modelo, es importante explicar primero ciertos conceptos. En primer lugar, hay que entender qué es una matriz de confusión y qué valores podemos obtener a partir de ella. En un problema de clasificación binario como el planteado, hay que distinguir entre dos clases: clase **positiva**, en este caso si se tiene la enfermedad de Alzheimer, y clase **negativa**, en caso que el individuo esté sano. Por otro lado, tenemos predicciones: predicción **verdadera**, es decir se acierta el valor real, o predicción **falsa**, en caso que no se acierte. De esta forma, en nuestra problemática se tiene cuatro posibles escenarios para nuestra hipótesis nula "el individuo tiene EA":

1. el individuo tiene EA y se predice correctamente que tiene (*verdadero positivo (VP)*).
2. el individuo no tiene EA y se predice incorrectamente que tiene (*falso positivo (FP)*) o

error tipo I);

3. el individuo no tiene EA y se predice correctamente que no tiene (*verdadero negativo (VN)*);
4. el individuo tiene EA y se predice incorrectamente que no tiene (*falso negativo (FN)* o *error tipo II*);

En general, los errores de tipo I suelen ser de mayor envergadura dado que implican rechazar erróneamente la hipótesis nula. En el campo de la medicina, esto sería comunicarle a un individuo que tiene una enfermedad/condición cuando realmente no es así. Un FN, error tipo II, podría ser incluso más grave dado que erróneamente se estaría comunicando al paciente que no tiene una enfermedad cuando lo cierto es que sí la tiene. El individuo volvería tranquilamente a su vida normal cargando con una enfermedad sin saberlo y dejando en muchos casos que la misma progrese.

En este caso en particular, comunicarle a una persona que tiene o va a desarrollar Alzheimer cuando en realidad no es así (error tipo I - FP), podría generar angustia, dolor y frustración. La persona podría tratarse para frenar el avance del deterioro, mediante ejercicios mentales o de coordinación por ejemplo, pero el costo más alto sería probablemente emocional. Incluso, el paciente seguramente se sometería a futuros estudios para chequear el progreso de la enfermedad, y con suerte se evidenciaría que el primer diagnóstico era erróneo.

Un error de tipo II, FN, sin duda tendría efectos negativos más severos. Se le comunicaría a la persona que está en condiciones normales y que no tiene EA, cuando efectivamente tiene o tendrá. En este caso, el paciente volvería a su vida normal, no trataría la enfermedad en progreso, y dejaría inconscientemente que la misma avance y se vaya apoderando de su cerebro. Podrían pasar muchos años hasta que la persona note nuevamente que algo anda mal, siendo quizás demasiado tarde para frenar o revertir el deterioro. La persona habría sido privada de la oportunidad de hacerse un tratamiento adecuado.

En cuanto a la veracidad de las predicciones, se tiene que hacer distinción entre dos términos: **sensibilidad** y **especificidad**. La **sensibilidad**, también conocida como tasa de verdaderos positivos (TPR de su sigla en inglés True Positive Rate) o *recall*, mide la proporción de individuos con la condición nula que son correctamente identificados. En este caso, individuos con EA a quienes se los identifica como tales. La **especificidad**, o tasa de verdaderos negativos (TNR de su sigla en inglés True Negative Rate) en cambio es la proporción de individuos que no poseen la condición y son correctamente catalogados (individuos sanos con condiciones normales a quienes se los identifica como tales).

En otras palabras, la sensibilidad en pruebas médicas de diagnóstico es una medida del grado en que los VP no son pasados por alto. A mayor grado de sensibilidad, menor sería el número de FN. Mide la probabilidad de un resultado positivo dado que el paciente tiene la condición. La especificidad mide el grado en que los VN son bien clasificados, y a mayor grado de especificidad menor es el número de FP. Se mide la probabilidad de un resultado negativo dado que el paciente no tiene la condición.

Una prueba médica altamente sensible rara vez pasa por alto un caso positivo (quienes tienen la condición serán identificados como tal), mientras que una prueba altamente específica rara vez pasa por alto un caso negativo (quienes no tienen la condición son correctamente identificados). Estas tasas se pueden calcular:

$$Sensibilidad = \frac{\# VP}{\# VP + \# FN}$$

$$Sensibilidad = \frac{\# VP}{\# Total\ de\ individuos\ enfermos}$$

$$Especificidad = \frac{\# VN}{\# VN + \# FP}$$

$$Especificidad = \frac{\# VN}{\# Total\ de\ individuos\ sanos}$$

Estos dos indicadores suelen representarse gráficamente en lo que se conoce como la curva ROC (*Receiver Operating Characteristic*). Esta curva representa la tasa de verdaderos positivos (TPR) en función a la tasa de falsos positivos (FPR - False Positive Rate). El FPR se define como:

$$FPR = 1 - Especificidad$$

quedando de este modo una curva de sensibilidad en función a la especificidad. El objetivo de dicha gráfica es determinar el punto de corte en una escala continua en el que se alcanza la especificidad y sensibilidad más alta y así poder evaluar la capacidad discriminativa del test, es decir, diferenciar entre sanos y enfermos. Más importante que la curva en sí misma es el área bajo la misma, conocida como AUC-ROC (de su sigla en inglés *area under curve ROC*). Es una medida que indica el grado de separabilidad de las clases.

El AUC es un número que oscila entre 0 y 1. Un modelo que tiene asociado un AUC de 0 significa que hace todas las predicciones de manera incorrecta. Un AUC de 1 por el contrario está asociado a un modelo que clasifica todas las muestras correctamente. Un AUC que vale 0.5 describe a un modelo que no tiene capacidad discriminativa, es decir que tiene un 50% de probabilidades de predecir bien así como de predecir mal. En la figura 2.8 se tienen representadas a modo de ejemplo dos curvas ROC de modelos diferentes. En ese caso, el test A supera al test B en tanto tiene mayor capacidad discriminativa (mayor AUC). La línea naranja representa la curva de no-discriminación. Una guía aproximada para clasificar la capacidad discriminativa de una prueba de diagnóstico es el sistema académico tradicional de puntos [55]:

- 0.9 - 1: Excelente;
- 0.8-0.9: Bueno;
- 0.7-0.8: Aceptable;

- 0.6-0.7: Pobre;
- 0.5-0.6: Malo / Falla;

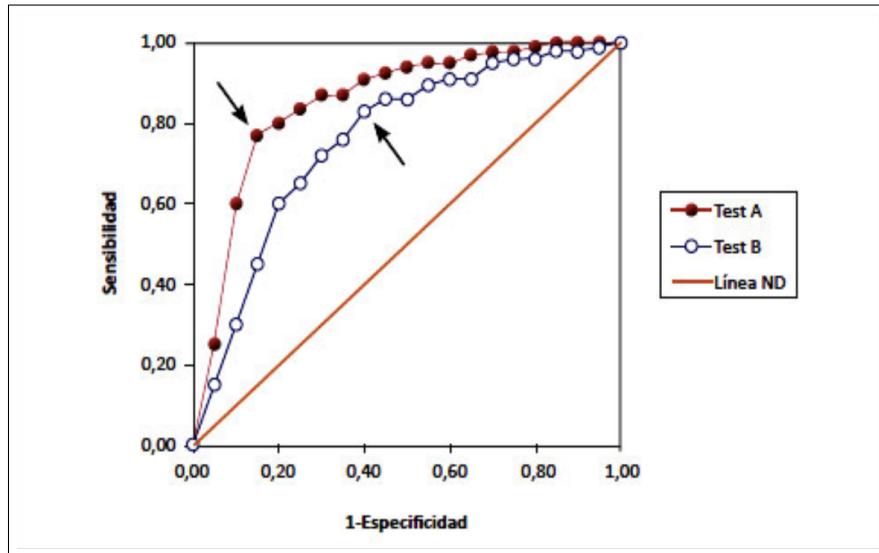


Figura 2.8: Ejemplo de curvas ROC. El test A tiene un mejor desempeño a la hora de clasificar que el test B. La línea naranja representaría a un modelo que tiene un 50% de probabilidades de clasificar tanto bien como mal.

Otros indicadores que son muy útiles a la hora de evaluar modelos son exactitud, precisión, y valor F1, F0.5 y F2. **Exactitud (accuracy)** y **precisión (precision)** son dos términos que suelen confundirse fácilmente. El primero hace referencia al grado de dispersión de los valores medidos respecto a un valor específico; que tan cerca está el valor obtenido respecto al valor real. Cuanto menor sea el sesgo, más exacta será la medición. En general se expresa mediante el error absoluto, esto es, la diferencia entre el valor medido y el valor real. En análisis predictivo, la exactitud se puede calcular como las mediciones verdaderas sobre el total de mediciones:

$$EXACTITUD = \frac{\#VP + \#VN}{\#VP + \#VN + \#FP + \#FN}$$

Precisión es en cambio una medida de la dispersión de los valores obtenidos de mediciones repetidas. También se la conoce como *Valor Predictivo Positivo (VPP)*, y mide la

probabilidad de tener la enfermedad si el resultado de la prueba diagnóstica es positivo. Es un indicador de la variabilidad de las mediciones, a menor dispersión o variabilidad, mayor es la precisión. Se puede calcular como:

$$PRECISIÓN \text{ o } VPP = \frac{\#VP}{\#VP + \#FP}$$

Por otro lado, existe también lo que se conoce como *Valor Predictivo Negativo (VPN)*, que contrariamente al VPP, mide la probabilidad de no tener la enfermedad si el resultado de la prueba diagnóstica es negativo.

$$VPN = \frac{\#VN}{\#VN + \#FN}$$

Los valores F son una medida que combina la precisión con el recall/sensibilidad. En términos generales se calcula como:

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{precisión} \cdot \text{recall}}{(\beta^2 \cdot \text{precisión}) + \text{recall}}$$

Si $\beta = 1$, entonces se trata del valor F1, en donde se pondera por igual tanto al recall como a la precisión. Los otros valores que se le suele dar a β son 0.5 para una ponderar en mayor medida a la precisión, y 2 para ponderar más a la sensibilidad.

2.4.1 Criterios de comparación

Como se ha mencionado en la sección 2.1.2, diversas estrategias se han abordado a lo largo de los años para diagnosticar EA. Los criterios de diagnóstico originalmente fueron prácticamente clínicos, siguiendo con los criterios propuestos en 1984 por Instituto Nacional de Trastornos Neurológicos, Comunicativos y Accidentes Cerebrovasculares - Asociación de la Enfermedad de Alzheimer y Trastornos Relacionados (NINCDS - ADRDA) [56] y por el Manual Diagnóstico y Estadístico de Trastornos Mentales (hoy en día está disponible la quinta edición, "Guía de consulta de los criterios diagnósticos del DSM-5") [57], pero con el

avance en las últimas décadas en biomarcadores, neuroimágenes y genética, se han propuesto nuevas recomendaciones a la hora de hacer un diagnóstico. En el 2011 el Instituto Nacional del Envejecimiento y la Asociación de Alzheimer (NIA - AA) [58, 59] definieron nuevas pautas que abarcan varios niveles de diagnóstico en los que se incluyen MCI, demencia (todas las causas), demencia con EA probable y demencia con EA posible.

Múltiples estudios llevados a cabo han demostrado falta de precisión en los estudios clínicos [60, 61, 62, 63]. Los indicadores de sensibilidad y especificidad alcanzaron rangos que van desde 34% a 96% y 33% a 100% respectivamente. El hecho de que los valores varíen tanto entre los diferentes estudios sugiere falta de objetividad y precisión en el método, probablemente a causa de las diferencias en las habilidades, experiencia y conocimiento de quienes diagnostican, así como también de la falta de procedimientos estandarizados.

El estudio publicado en 1998 por Jobst et al. [61] basado en los criterios de diagnóstico de NINCDS-ADRDA y del DSM-IV (1994) en 198 sujetos obtuvo un índice de sensibilidad de 96% para EA posible o probable (NINCDS-ADRDA) y 51% para demencia senil o presenil de tipo EA (DSM-III -1980), y un índice de especificidad de 61% y 97% respectivamente. En contraposición a esto, Mayeux et al. analizaron 1707 sujetos con las guías del NINCDS-ADRDA y 66 con las del DSM-III (1980), y obtuvieron 93% de sensibilidad y 55% de especificidad, con valores de especificidad variando considerablemente según la edad de los individuos (rango de 23% a 66%).

Los resultados obtenidos mediante estudios que involucran imágenes, en cambio, han demostrado ser más concisos con rangos de variabilidad sustancialmente más acotados [62, 63, 64]. Si bien es cierto que sigue habiendo subjetividad a la hora de interpretar imágenes, el proceso de obtención de las mismas es algo que está regulado y estandarizado, y del que se pueden obtener medidas cuantitativas, eliminando gran parte de la varianza a causa de interpretaciones.

La figura 2.9 muestra las varianzas en la sensibilidad y especificidad de los distintos métodos de diagnóstico, clínicos y por imágenes, en diferentes estudios recopilados [62]. Se

puede ver como hay un alto grado de variabilidad en los métodos clínicos, así como también se evidencian tasas de especificidad y sensibilidad menores. Asimismo, las imágenes de RM han demostrado tener los mejores resultados, con las tasas más elevadas y varianzas acotadas.

Por otro lado, en la figura 2.10 se comparan los criterios de diagnóstico de NINCDS-ADRDA de 1984 y NIA-AA de 2011 para EA probable, EA posible y EA posible o probable según Harris et al. [65]. Tal como era de suponer, el criterio propuesto por NIA-AA basado no sólo en análisis clínicos sino también en biomarcadores e imágenes, prevalece sobre todo cuando se trata de EA posible o probable donde la sensibilidad alcanza 80.9% y la especificidad 94.0%.

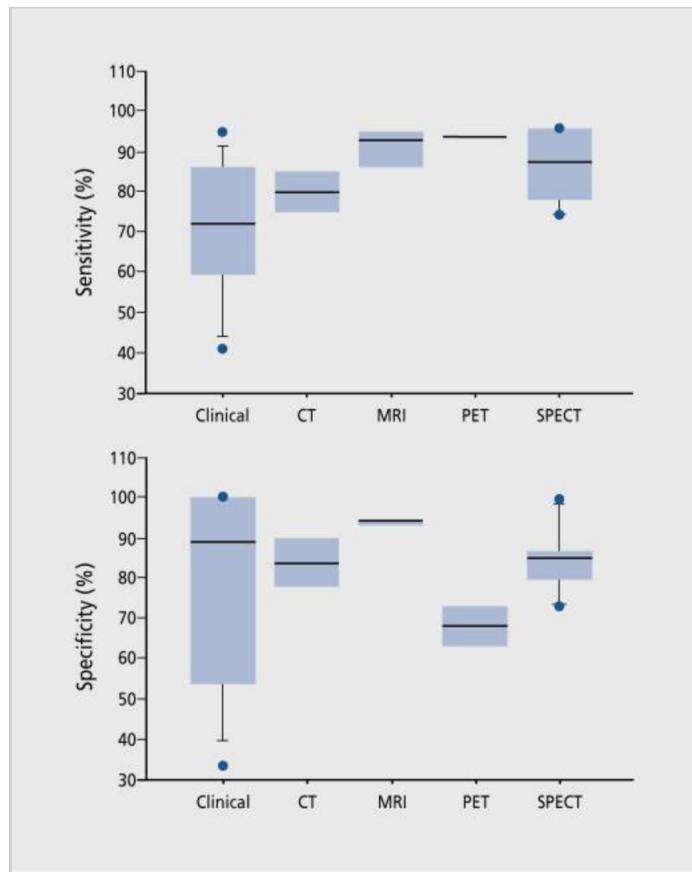


Figura 2.9: Imagen obtenida de '*Sensitivity and specificity of neuroimaging for the diagnosis of Alzheimer's disease*' [62] donde se comparan las varianzas de diferentes métodos diagnósticos utilizados para la detección de EA.

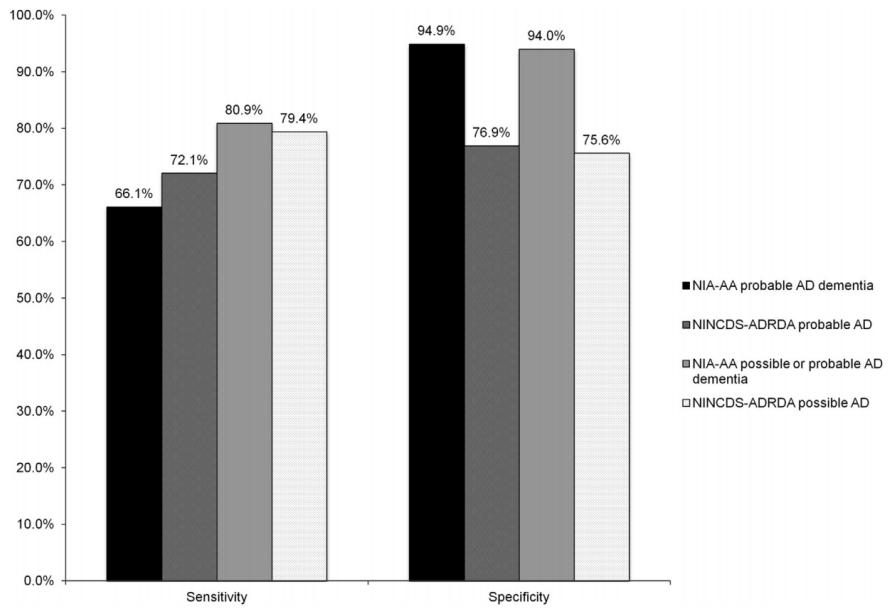


Figura 2.10: Imagen obtenida de '*Do NIA-AA criteria distinguish Alzheimer's disease from frontotemporal dementia?*' [65]. Se comparan las tasas de sensibilidad y especificidad de diferentes estudios utilizando los criterios de NINCDS-ADRDA y NIA-AA para EA posible, probable, y posible o probable.

Bloudek et al. [66] llevaron a cabo un meta análisis exhaustivo en el que se revisaron 119 publicaciones de estudios realizados entre 1990 y 2010 que analizan diferentes criterios de diagnóstico para EA con el objeto de estimar los valores de sensibilidad y especificidad. Las publicaciones revisadas consideraron criterios de diagnósticos clínicos, por imágenes de RM, TC, SPECT, FDG-PET o biomarcadores en LCR, en los que se comparaban alguna (o varias) de los siguientes clasificaciones: individuos con EA vs. (i) individuos con condiciones normales (sin ningún tipo de demencia); (ii) individuos con demencias de otro tipo (incluido DCL); y (iii) individuos con otro tipo de demencias (DCL excluido).

En las figuras 2.11 y 2.12 se incluyen los meta-análisis realizados para la sensibilidad y especificidad mediante criterios clínicos (8 estudios considerados) e imágenes de RM (26 estudios). Se puede ver por los resultados obtenidos para los indicadores combinados que los diagnósticos basados en imágenes resultaron ser más sensibles y específicos que los clínicos (más fiables), reforzando la importancia de incorporar a neuroimágenes en el diagnóstico de EA. Asimismo, el área bajo la curva ROC de los diagnósticos clínicos (donde

se compararon individuos con EA vs. individuos del caso (iii)) fue de 0.8, mientras que las áreas bajo la curva para diagnósticos mediante RM fueron de (i) 0.93, (ii) 0.86 y (iii) 0.90.

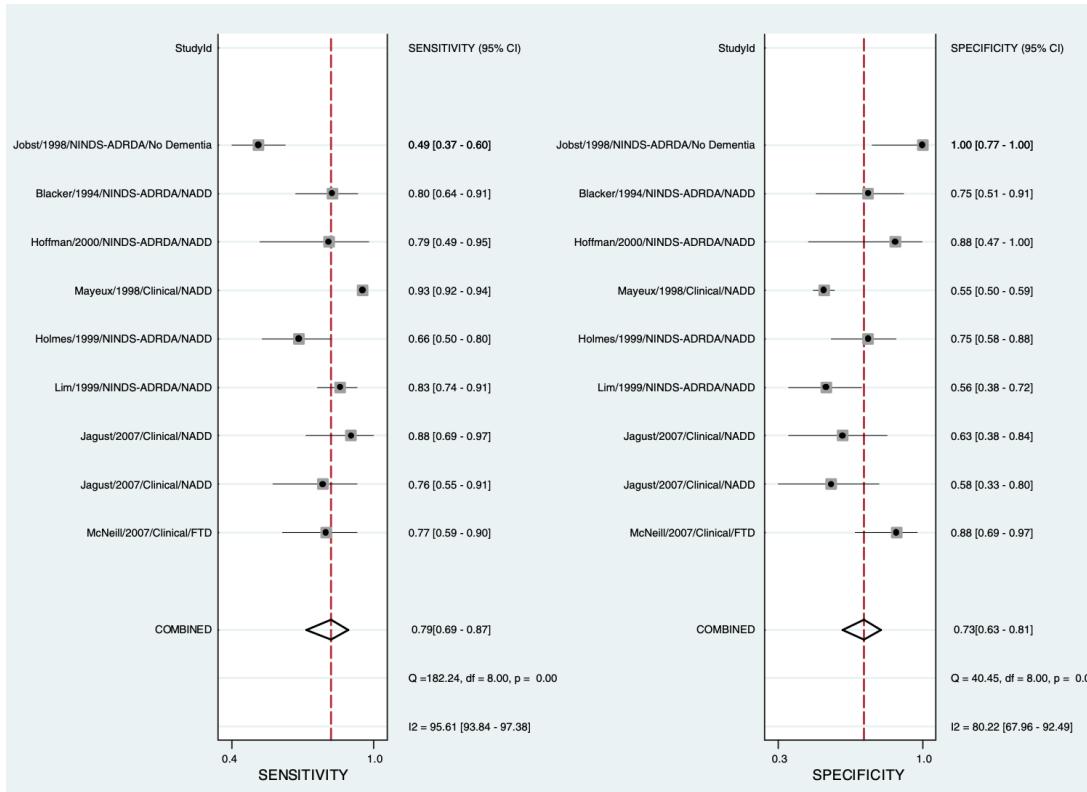


Figura 2.11: 8 Estudios considerados en el meta-análisis de sensibilidad y especificidad mediante diagnóstico clínico llevado a cabo por Bloudek et al. Al combinarlos obtuvieron una sensibilidad de 0.79 y especificidad 0.73.

Si bien en esta sección se incluyen estudios donde se revelan algunos indicadores, es importante destacar que el presente trabajo se encuentra restringido a un grupo demográfico determinado (individuos de Estados Unidos y Canadá, de edades comprendidas entre 55 y 93 años), y si bien esto no debe ser considerado como una validación externa dado que los datos demográficos no son particularmente comparables, sirve como punto de partida en los órdenes de magnitud de los resultados que se obtuvieron.

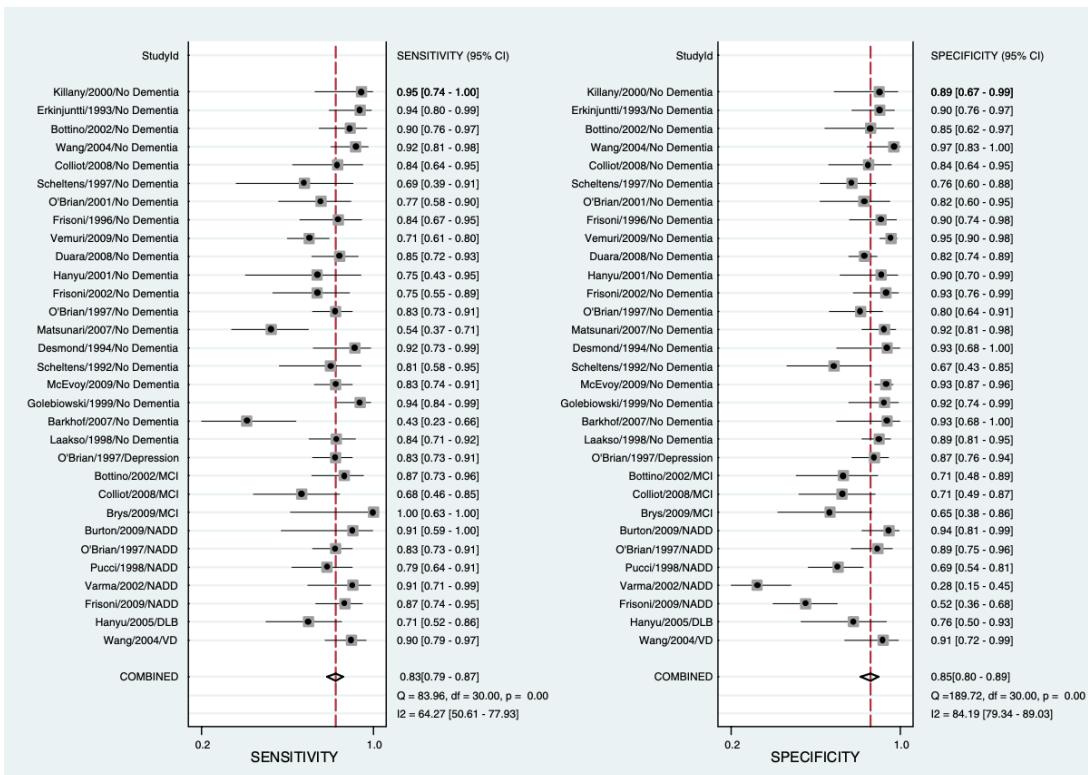


Figura 2.12: 26 Estudios considerados en el meta-análisis de sensibilidad y especificidad mediante diagnóstico por RM llevado a cabo por Bloudek et al. La sensibilidad combinada obtenida fue de 0.83 y la especificidad 0.85.

3 Metodología

3.1 Adquisición de Imágenes de RM

Se tomaron 1474 imágenes estructurales de RM ponderadas en T1 de la base de datos de ADNI, de las cuales 871 corresponden a individuos con diagnóstico de EA¹ y 603 a individuos con condiciones normales (CN). Todas las imágenes fueron adquiridas con secuencias 3D MP-RAGE, con parámetros de tiempo medio de: TE 3.44 ms, TR 1157.69 ms y TI 963.41 ms, y grosor de cortes medio de 1.211 mm. Se descargaron en formato NIfTI (Neuroimaging Informatics Technology Initiative).

		EA		CN	
		Femenino (127)	Masculino (140)	Femenino (58)	Masculino (55)
Edad	Media	74.76	76.68	78.55	78
	Desv. Estándar	8.07	7.55	6.56	6.13
MMSE	Media	21.82	21.25	28.84	28.76
	Desv. Estándar	3.66	4.01	1.75	1.63

Tabla 3.1: Distribución demográfica de edad y MMSE según condición neurológica y sexo.

La tabla 3.1 resume la distribución demográfica de los grupos estudiados, donde se incluye la edad media y la puntuación MMSE media por grupo, junto con las desviaciones estándar. Hay que resaltar que todos los individuos en cuestión pertenecen al territorio de Estados Unidos (EEUU) y Canadá, pero se desconoce el grupo étnico de cada uno. Por otro lado, los histogramas de la figura 3.1 muestran las frecuencias de edades para cada grupo en cuestión. Como se puede ver, la gran mayoría de los sujetos en cuestión son de edades avanzadas, donde en general los signos de la enfermedad suelen ser evidentes.

Cabe aclarar que el número de imágenes utilizadas supera al número de individuos de cada grupo, dado que en varios casos se consideraron más de un estudio por individuo, es

¹No es correcto afirmar que dichas personas tienen efectivamente Alzheimer dado que para poder determinarlo con certeza, se debe realizar una biopsia o necropsia, procedimiento que no fue realizado en la categorización de los individuos.

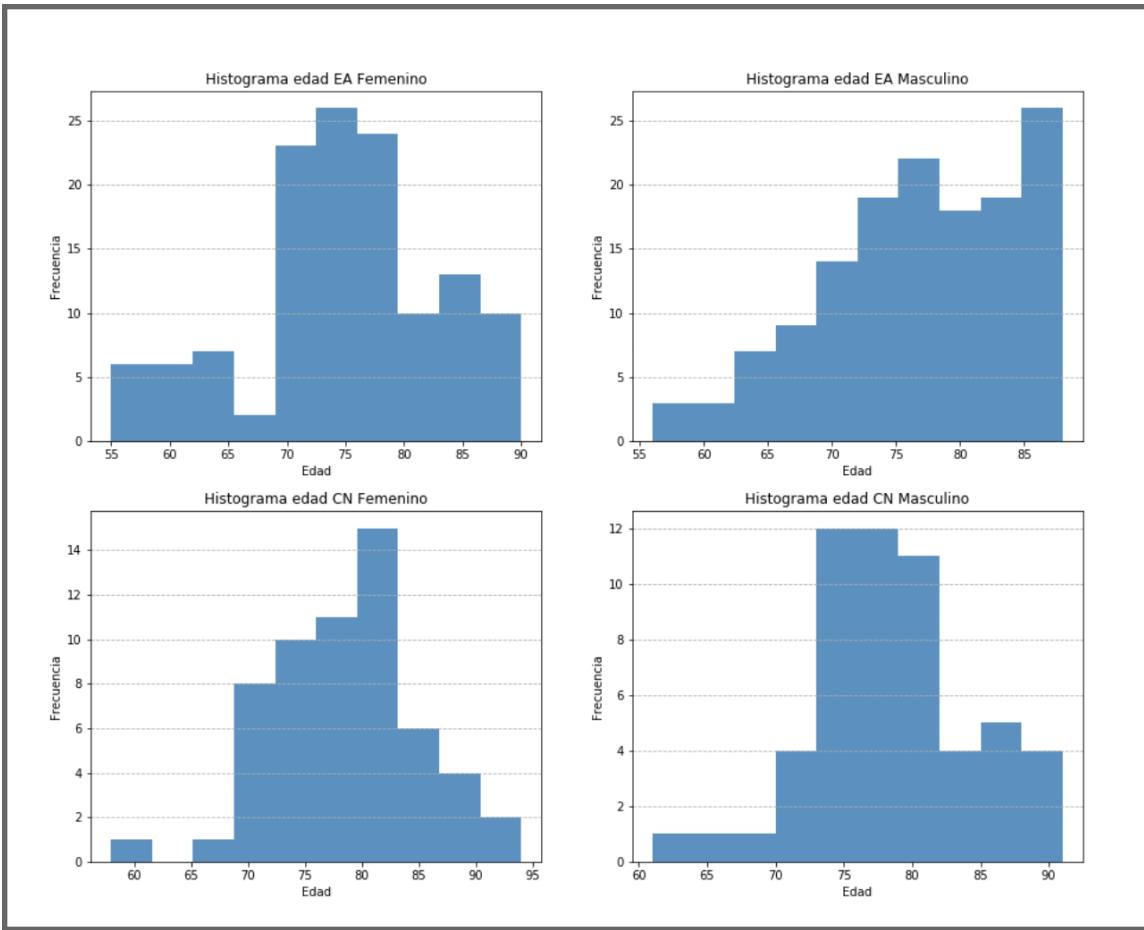


Figura 3.1: Histogramas de distribución de edad según rango etario y condición neurológica.

decir algunos pacientes cuentan con más de un estudio de **RM** asociado, llegando al total de 1474 imágenes mencionadas. Asimismo hay que destacar que a cada imagen de **RM** le corresponden numerosos cortes que varían según cada estudio, y por ende, el total de imágenes que considerará la red es ampliamente mayor.

En resumidas cuentas, se tienen 380 individuos de EEUU y Canadá, de los cuales 267 corresponden a individuos con diagnóstico de demencia tipo Alzheimer (127 femeninos vs. 140 masculinos), y 113 a individuos sanos (58 femeninos vs. 55 masculinos). Particularmente, de las 871 imágenes de **EA**, 721 pertenecen a individuos mayores a 70 años, y de los 603 individuos sanos, 553 son mayores de 70. El número de imágenes disponibles de gente menor a 70 es considerablemente bajo (150 para **EA** y 50 para **CN**).

3.2 Preprocesamiento

Luego de descargar las imágenes en formato **NIfTI**, se utilizó el programa SPM12 (Statistical Parametric Mapping 2014 [67]) en MATLAB para centrar las imágenes, estableciendo el origen en la comisura anterior del cerebro (imagen 3.2 y 3.3).

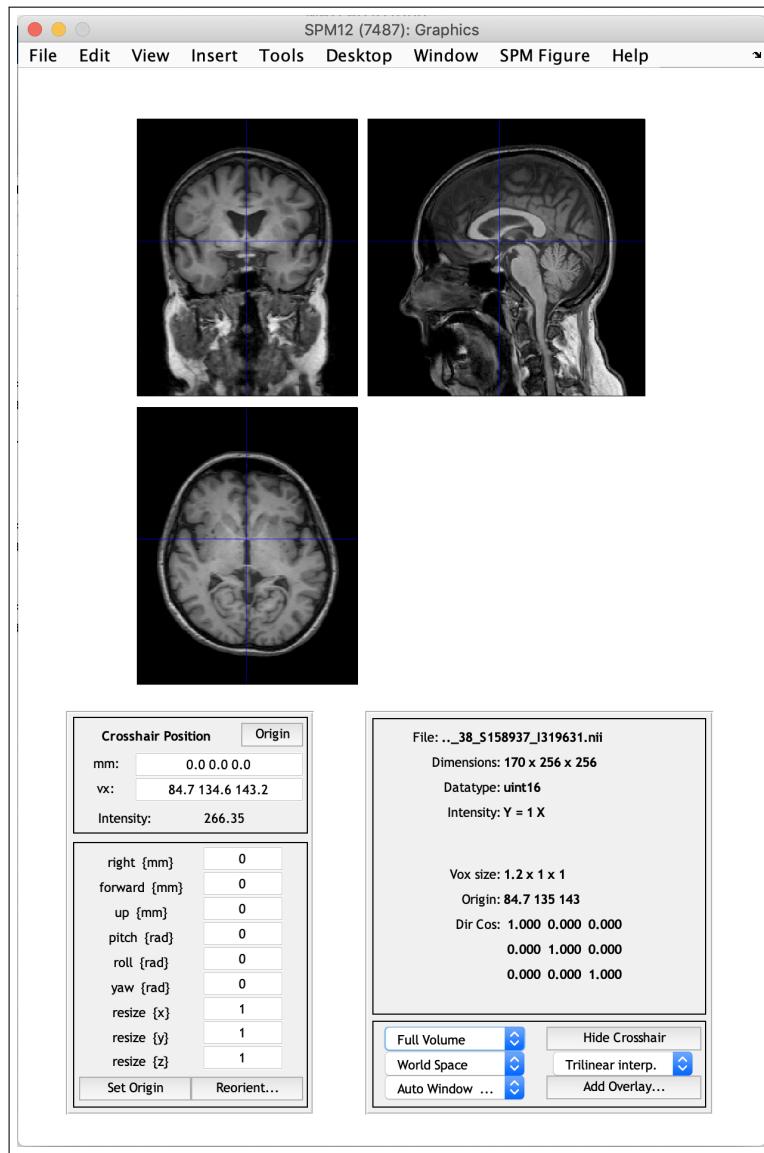


Figura 3.2: Visualización del programa utilizado para fijar el origen en la comisura anterior del cerebro, SPM12.

Una vez fijado el centro de cada una, se procedió por reducir el ruido mediante un paquete para imágenes de **RM** (MRI Denoising Software [68]). Se aplicó el filtro PRINLM

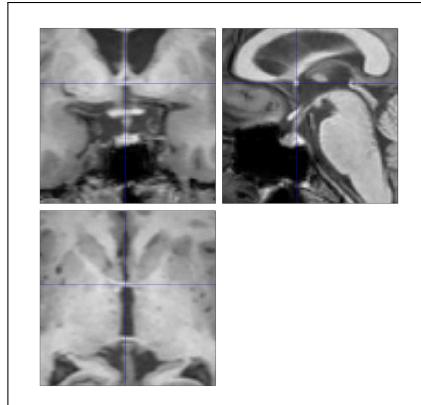


Figura 3.3: Imágenes ampliadas del cerebro que evidencian donde se fijó el origen.

(Prefiltered Rotationally Invariant NonLocal Means filter) y se consideró al ruido con distribución de Rician [69]. En la imagen 3.4 se puede ver la imagen previa y posterior a la aplicación del filtro, así como el ruido eliminado.

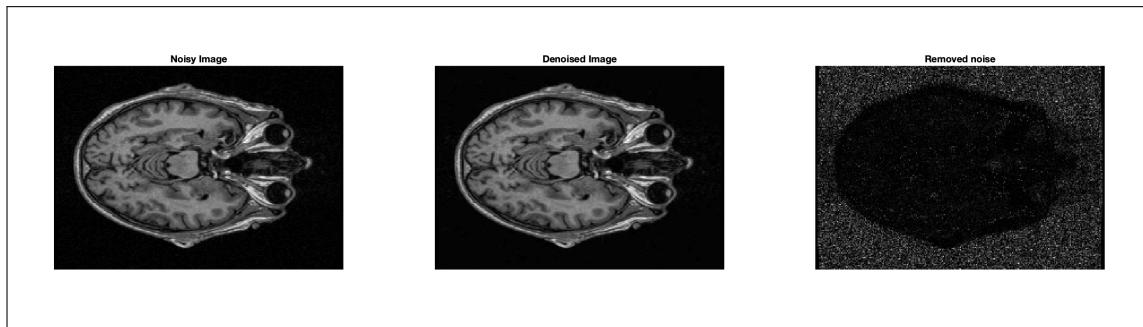


Figura 3.4: Se puede ver a la izquierda la imagen cruda con ruido. En el centro, la imagen luego de aplicarle el filtro PRINLM. A la derecha se ve el ruido que fue eliminado.

Finalizado este proceso, el siguiente paso fue extraer el cerebro eliminando todas las regiones que no fueran de interés, e.g cráneo y cuello. Para ello se utilizó el programa FSL-BET, una librería para el análisis de fMRI (funcional), MRI y DTI (Diffusion Tensor Imaging) que cuenta con una herramienta para extraer el cerebro, Brain Extraction Tool (BET) [70]. Se probaron distintos valores para el umbral de intensidad fraccional $-f-$ y para el gradiente vertical de dicho umbral $-g-$, con el objeto de recortar la mayor cantidad de tejido no-cerebral sin perder porciones de cerebro. En las figuras 3.5 y 3.6 se puede observar diferentes resultados de extracciones de cerebro utilizando los mismo parámetros. En la 3.6

se puede ver como han quedado porciones de cráneo sin extraer, mientras que en [3.5] los resultados obtenidos fueron más precisos.

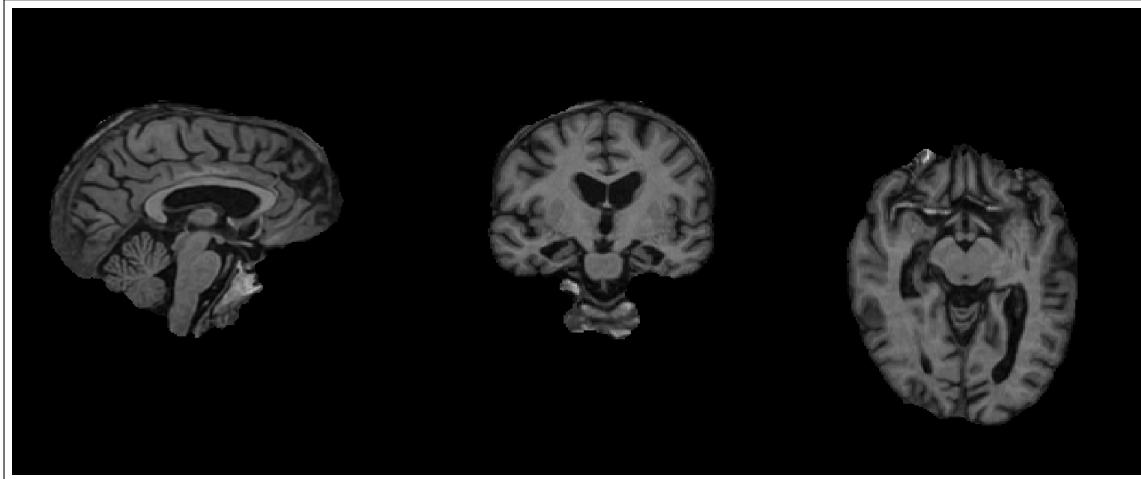


Figura 3.5: Imagen de un cerebro extraído correctamente utilizando la herramienta FSL-BET. Cortes sagital (izquierda), coronal (centro) y axial (derecha).

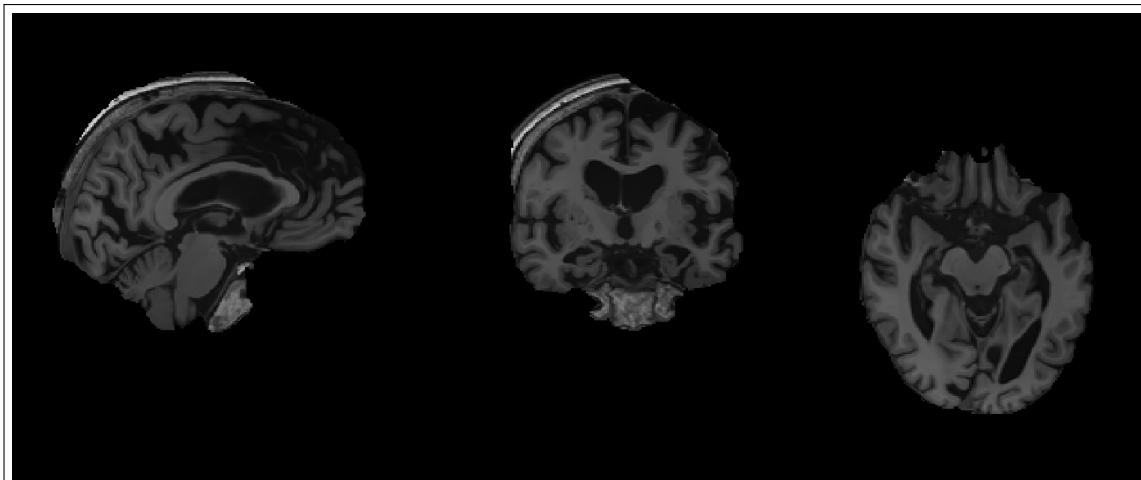


Figura 3.6: En este caso, el cerebro no logró extraerse correctamente. Se puede ver en los cortes sagital y coronal que han quedado porciones de cráneo.

A fines de segmentar el cerebro en materia blanca, materia gris y LCR, se implementó la herramienta FAST [71], también perteneciente a FSL. Esta herramienta es capaz de segmentar imágenes 3D así como también corrige las variaciones de intensidad espacial o inhomogeneidades de radiofrecuencia. En la imagen [3.7] se muestra en primer lugar una imagen luego de haber extraído el cerebro. Por debajo, se pueden ver las tres segmentaciones

obtenidas al aplicarle FAST. En primer lugar se ve el LCR, seguido de la materia gris, y finalmente la materia blanca. Con todas las imágenes de materia gris formadas, se tomó un modelo estándar del ICBM (International Consortium for Brain Mapping), particularmente la máscara de materia gris del modelo no-lineal simétrico del 2009 [72, 73, 74], y se las registró sobre el mismo utilizando la herramienta FLIRT [75, 76, 77] perteneciente a su vez a FSL. De este modo, todas las imágenes adquirieron las dimensiones del template, quedando de $197mm \times 233mm \times 189mm$.

Finalmente, antes de proceder con el armado de los modelos de red, se convirtieron las imágenes obtenidas en el proceso de registro anterior a formato PNG empleando una librería de Python que convierte imágenes en formato DICOM o NIfTI a PNG o JPG, *med2image* [78]. Se eliminaron luego todas las imágenes negras de cada dimensión (x , y y z) estableciendo un umbral con la media de las imágenes. El número total de imágenes resultantes fue de 736.697, de las cuales 653.284 se utilizaron para entrenar, 72.586 para la validación, y 10.827 para la evaluación del modelo.

3.3 Modelos de CNN Implementados

Se realizaron cuantiosas pruebas con diferentes versiones adaptadas de arquitecturas de redes convolucionales existentes. Los modelos que fueron seleccionados para entrenar como primera aproximación fueron LeNet-5 [79], AlexNet [80], VGG16 [81] y ResNet50 [82], adaptando únicamente los valores dimensionales de entrada con las imágenes obtenidas en el proceso de preprocesamiento, y considerando a su vez que la salida del problema en cuestión es binaria (clasificación en dos clases).

A grandes rasgos, los modelos empleados se pueden diagramar como un bloque de entrada, una serie de capas convolucionales seguidas (o no) de capas *pooling*, luego un bloque de capas *fully-connected* con función de activación *ReLU*, y en último lugar una *fully-connected* con función de activación sigmoidea para resolver el problema binario: condición normal (CN) vs. enfermedad de Alzheimer (AD). En la figura 3.8 se puede ver una estructura

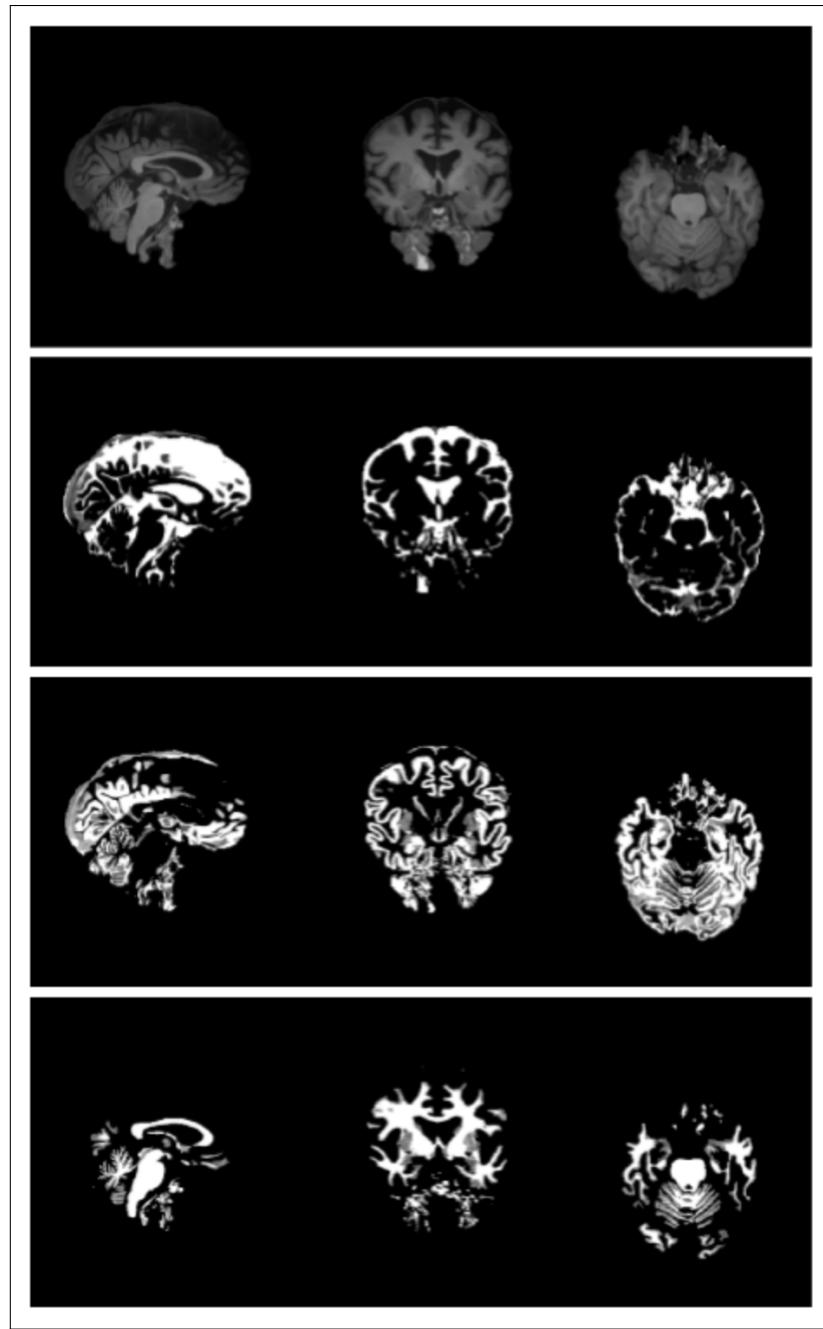


Figura 3.7: Por encima se muestra una imagen de un cerebro extraído, seguido de las tres segmentaciones utilizando el modelo del ICBM. En primer lugar el **LCR**, seguido de la materia gris y materia blanca respectivamente.

genérica que representa a los modelos implementados. Estos se distinguen unos de otros principalmente por la estructura comprendida en el bloque de capas convolucionales y pooling.

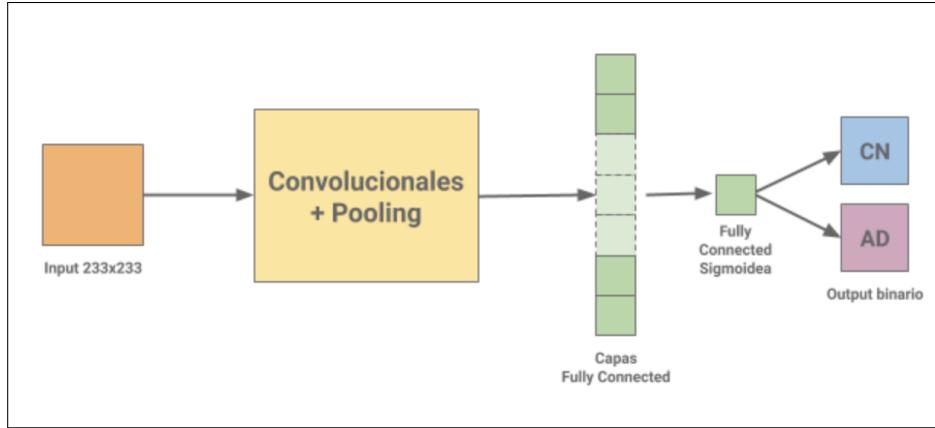


Figura 3.8: Diagrama en bloque de los modelos de redes implementados. El primer bloque corresponde a las imágenes de entrada de dimensión 233×233 . El segundo a las capas convolucionales junto con las pooling. Luego una serie de capas *fully-connected*, seguidas por una *fully-connected* de un nodo y función de activación sigmoidea, dando como resultado un output de dos clases.

Las primeras pruebas se llevaron a cabo a modo de identificar dentro de las arquitecturas mencionadas, las que mejores resultados obtuvieran con el propósito de reducir el número de modelos para entrenar. Los que prevalecieron fueron AlexNet y VGG16. Hay que destacar que estas no fueron las redes finales utilizadas, sino más bien el punto de partida para construir los modelos que acabaron por destacarse. Se construyeron diferentes versiones de modelos inspirados en estos resultados, y se entrenaron utilizando una GPU de Amazon Web Services (AWS).

3.3.1 AlexNet

La red AlexNet fue diseñada en el 2012 por Krizhevsky, Sutskever y Hinton con el propósito de clasificar las 1.2 millones de imágenes de alta resolución del dataset del *ImageNet Large Scale Visual Recognition Challenge 2010* (ILSVRC-2010) en 1000 clases diferentes [80].

La estructura se compone principalmente por 5 capas convolucionales, algunas seguidas de capas max-pooling, y finalmente 2 capas fully-connected seguidas por una softmax de 1000 nodos, dando un total de 60 millones de parámetros y 650.000 neuronas. En la figura 3.9 se puede ver una representación del modelo.

El conjunto de datos del ImageNet LSVRC-2010 se compone de alrededor de 1.2

millones de imágenes de entrenamiento, 50.000 imágenes de validación y 150.000 imágenes de testeo. Se redujeron todas a la misma dimensión, 256×256 píxeles, y con el objeto de aumentar la cantidad de datos disponibles para prevenir over-fitting, se realizó *data augmentation* mediante traslaciones y reflexiones horizontales con fragmentos aleatorios de 227×227 píxeles. Se logró de esta forma aumentar el tamaño del conjunto de entrenamiento por un factor de 2048.

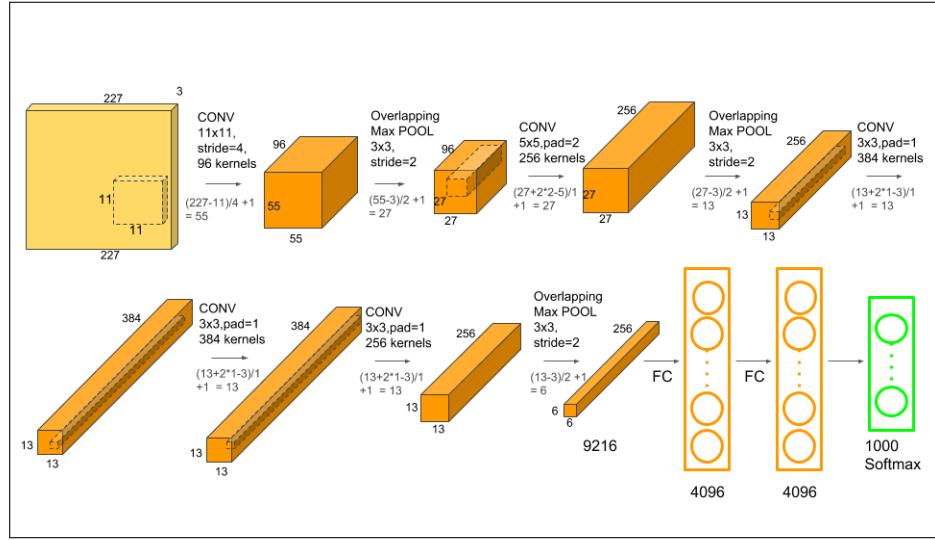


Figura 3.9: Arquitectura de una red convolucional AlexNet. La imagen fue obtenida de: <https://neurohive.io/en/popular-networks/alexnet-imagenet-classification-with-deep-convolutional-neural-networks/>.

Las particularidades más novedosas que adoptó este modelo son:

1. **Función de activación ReLU:** al reemplazar la función *tanh* por la ReLU en las capas convolucionales lograron obtener una tasa de error de entrenamiento del 25% seis veces más rápido.
2. **Múltiples GPUs:** debido al gran volumen de datos y a las limitaciones de memoria de las GPUs en aquel entonces, distribuyeron la red en dos GPUs para lograr una paralelización y reducir los tiempos de entrenamiento.
3. **Superposición de capas pooling:** al tomar un *stride* menor al tamaño de los filtros

pooling implementados, lograron superponer regiones de las imágenes de entrada obteniendo una reducción en las tasas de error.

En resumidas cuenta, la arquitectura se compone de una capa que recibe las imágenes de dimensión $227 \times 227 \times 3$ y las convoluciona con 96 filtros de 11×11 y un *stride* de 4, luego se realiza una normalización y un max-pooling con superposición (*stride* = 2), y la salida ingresa a la segunda capa convolucional. Esta consiste en 256 filtros de 5×5 , seguidos de una normalización y max-pooling con *stride* = 2. La tercera, cuarta y quinta capa se conectan directamente, sin aplicar normalización ni max-pooling entre ellas. La tercera está formada por 384 filtros de 3×3 , al igual que la cuarta. La quinta en cambio son 256 filtros de 3×3 seguidos de una capa de max-pooling con superposición (*stride* = 2). La salida se conecta a una serie de dos capas *fully-connected* de 4096 nodos cada una, y con dropout de 0.5. Finalmente, las imágenes atraviesan una *softmax* de 1000 nodos para la clasificación en clases.

3.3.2 VGG-16

VGG16 es una red neuronal convolucional que fue propuesta por Simonyan y Zisserman en la competencia ILSVRC-2014 [81]. El número 16 hace referencias a las 16 capas con pesos que posee la red. La singularidad de este modelo es que en lugar de tener una gran número de hiperparámetros, se caracteriza por tener filtros pequeños de 3×3 con un *stride* = 1 para preservar la resolución, y capas max-pooling de 2×2 y *stride* = 2. Los datos utilizados son los mismos que los que se usaron para AlexNet.

En la figura 3.10 se puede ver en detalle cómo está conformada esta estructura. Los tensores de entrada son imágenes en formato RGB de dimensión 224×224 . Como preprocesamiento, se restó únicamente el valor RGB medio para cada píxel. Las primeras dos capas convolucionales tienen 64 filtros de 3×3 y un *padding* = 1 tal que no se altera la dimensión de las imágenes de entrada, y son seguidas por una capa max-pooling de 2×2 con *stride* = 2. Las dos capas convolucionales siguientes tienen 128 filtros de 3×3 cada

una, y le sigue otra max-pooling igual a la anterior. Las capas subsiguientes consisten en tres convolucionales de 256 filtros de dimensión 3×3 , y a la salida una max-pooling con las mismas características que las anteriores. Luego hay dos series de tres convolucionales de 512 filtros de 3×3 y una max-pooling tal como las anteriores. Finalmente, las imágenes atraviesan 3 capas *fully-connected*, las primeras dos de 4096 nodos cada una, y la última una *softmax* de 1000 nodos para clasificar en el número de clases deseadas. Todas las funciones de activación son de tipo ReLU.

La red tuvo un desempeño que se destacó por sobre el resto de los modelos en la competencia y su facilidad a la hora de implementar, la convierten en una herramienta muy útil en reconocimiento de imágenes. Los resultados fueron mejores incluso que los obtenidos con AlexNet para el mismo dataset. De todos modos, uno de los desafíos que presenta este tipo de arquitectura es el tiempo de entrenamiento que conlleva debido al alto número de parámetros involucrados (alrededor de 138 millones), así como el tamaño de los pesos que son considerablemente grandes en relación con el disco y ancho de banda

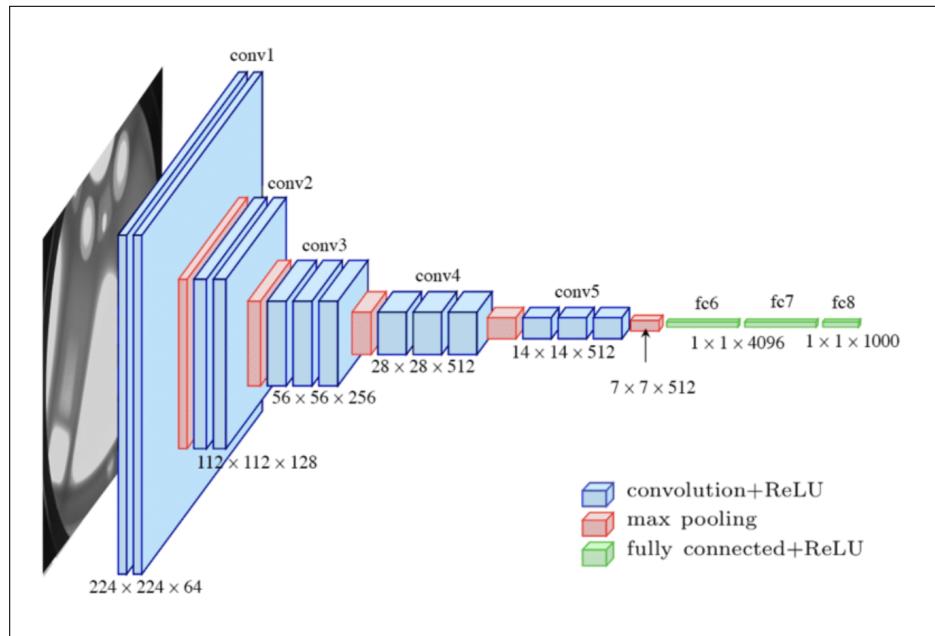


Figura 3.10: Arquitectura de una red convolucional VGG16. La imagen fue obtenida y adaptada de "Automatic localization of casting defects with convolutional neural networks" [82].

3.3.3 Modelos Adaptados

Tal como se mencionó previamente, partiendo de los dos modelos descriptos en las secciones anteriores 3.3.1 y 3.3.2, se diseñaron nuevas arquitecturas que adoptaron ciertas características de ellos. Las pruebas realizadas se podrían agrupar en dos tipos de modelos. Por un lado, modelos orientados/inspirados en AlexNet, y por otro, modelos que utilizan una VGG-16.

En las figuras 3.11 y 3.12 se pueden ver simplificadas ambas arquitecturas. El conjunto de modelos basados en AlexNet que obtuvieron el mejor desempeño consisten en una capa convolucional de 6 filtros de 3×3 y $stride = 1$ que recibe como entrada las imágenes dimensionadas a 233×233 , seguida de una capa max-pooling de 2×2 y $stride = 1$. Luego atraviesan una segunda convolucional de 16 filtros con la misma dimensión y stride, y una capa max-pooling igual a la anterior. La última capa convolucional consiste en 120 filtros de 3×3 y $stride = 1$. Esta capa no tiene una capa pooling asociada, y desemboca directamente en la primera de las capas *fully-connected*. Entre las arquitecturas ganadoras lo que varía son la cantidad de capas y nodos implementados en esta última fase del modelo. La salida es finalmente sometida a una capa uninodal de activación sigmoidea para la clasificación en dos clases.

Como se puede ver, la estructura es notoriamente diferente de la AlexNet original. Se ha ido modificando para la obtención de mejores resultados. Aún así, hay ciertas cosas que se mantuvieron:

1. Funciones de activación ReLU.
2. Capas convolucionales seguidas de capas pooling de tipo max-pooling.
3. Capas pooling con superposición.

En cuanto a los modelos semejantes a la VGG-16, se mantuvo la arquitectura original del modelo, adaptando las dimensiones de entrada a las imágenes preprocesadas en cuestión,

y se agregaron al final una serie de *fully-connected* que se fueron modificando. Fueron tanto los nodos como el número de capas y el porcentaje de dropout los que se fueron variando entre las distintas experiencias para alcanzar los resultados más altos.

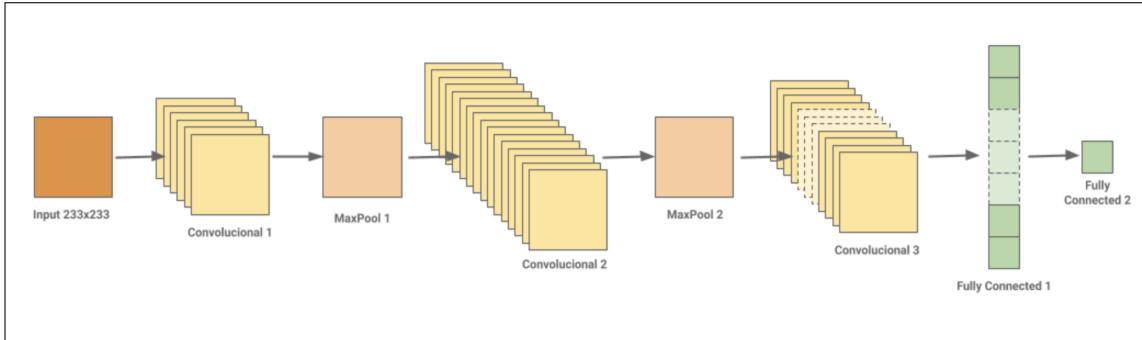


Figura 3.11

En la sección 3.4 se pueden ver los resultados de seis de los modelos que se destacaron, y en el apéndice A se incluyen secciones de código donde se detallan las estructuras de cada una de estas arquitecturas.

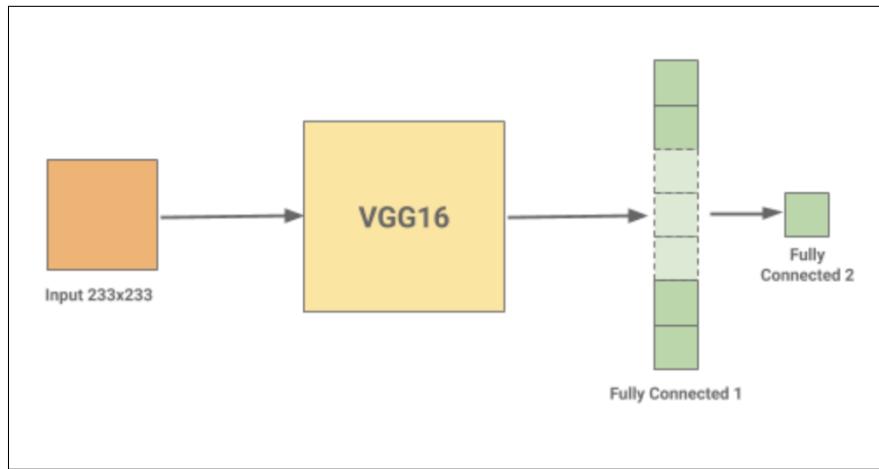


Figura 3.12

3.4 Resultados

Teniendo en consideración todos los indicadores mencionados en la sección 2.4, se han escogido los mejores tres modelos de tipo VGG16 y los mejores tres de tipo AlexNet

en función a la *exactitud* en la *validación* del modelo. Luego se tomó de esos seis el modelo ganador, y se calculó la matriz de confusión junto con el resto de los indicadores - sensibilidad, especificidad, exactitud, precisión (VPP), VPN, curva ROC, puntaje AUC-ROC, y los valores F1, F2 y F0.5 - para el set de datos previamente separados para *evaluación* final del modelo.

La tabla 3.2 muestra los seis modelos que mejor se desempeñaron. La primer columna hace referencia a la arquitectura utilizada. En la segunda y tercera columna están los resultados de entrenamiento, exactitud (*accuracy*) y función de costo (*loss-function*) respectivamente. Las dos últimas columnas muestran los resultados de validación. Tal como se puede ver, los modelos que tomaron a la red VGG16 como arquitectura principal prevalecieron frente a los que se basaron en AlexNet. En la tabla 3.3 se comparan los indicadores obtenidos para el mejor modelo de cada arquitectura, y se evidencia la superioridad del VGG por sobre AlexNet dado que todos los valores son mayores.

Arquitectura	Entrenamiento		Validación	
	Exactitud	Función de costo	Exactitud	Función de costo
Model-VGG16-(a)	0.9982	0.0036	0.9348	1.1009
Model-VGG16-(b)	0.9961	0.0075	0.9313	0.7211
Model-VGG16-(c)	0.9962	0.0071	0.9290	0.8158
Model-AlexNet-(a)	0.9938	0.0173	0.8944	0.8232
Model-AlexNet-(b)	0.9907	0.0252	0.8938	0.7280
Model-AlexNet-(c)	0.9949	0.0129	0.8929	0.8351

Tabla 3.2: Tabla de resultados de los modelos que mejor se desempeñaron. Se incluyen los mejores tres de cada tipo, VGG16 y AlexNet.

Modelo	AUC	ACC	VPP	VPN	SN	SP	F1	F0.5	F2
VGG16-(a)	0.99557	0.98273	0.98275	0.98270	0.99008	0.97012	0.98639	0.98420	0.98859
AlexNet-(a)	0.99517	0.97322	0.97452	0.97091	0.98335	0.95579	0.97892	0.97627	0.98157

Tabla 3.3: Tabla de indicadores obtenidos para el modelo ganador VGG16 y AlexNet utilizando el set de datos separados para evaluación final. ACC: exactitud, SN: sensibilidad, SP: especificidad

En la figura 3.13 se puede ver la matriz de confusión resultante del modelo ganador VGG16. La primer fila muestra los valores reales negativos y la segunda los positivos. El gráfico de la curva ROC se puede ver en la imagen 3.14, y en la tabla 3.3 se pueden ver los

valores obtenidas para los distintos indicadores considerados. El puntaje obtenido para el AUC, 0.99557, refleja según lo discutido en la sección 2.4 una alta capacidad discriminativa del modelo, también sostenida por los pocos casos de predicciones falsas - 187 (119 FP y 68 FN) de un total de 10.827 imágenes, dando una tasa de 0.01727 % (nótese que la exactitud (0.9827 %) es el complemento de este número).

Por otra parte, se puede ver que el valor más bajo es el de la especificidad (incluso en ambas arquitecturas), y el más alto corresponde a la sensibilidad (salvando el puntaje AUC). Esto está asociado a que el error tipo I (tasa de FP) es mayor que el error tipo II (tasa de FN). Como se mencionó previamente en la sección 2.4 en problemas de clasificación como este, si bien es cierto que los FP pueden tener un impacto negativo debido a las consecuencias de tratar una enfermedad inexistente (psicológicas, sociales y económicas), los FN pueden resultar sumamente perjudiciales puesto que imposibilitan un tratamiento adecuado; justamente es interesante en este modelo contar con una alta tasa de sensibilidad. Aún así, ambas tasas son considerablemente altas, más aún si se las compara con los estudios citados (figuras 2.10, 2.12).

Sin embargo, aún siendo la tasa de especificidad el valor más bajo obtenido, el VPN que indica la probabilidad de tener un diagnóstico correcto habiendo recibido una predicción negativa, es muy similar al VPP, es decir a la probabilidad de tener un diagnóstico correcto habiendo recibido una predicción positiva. Esto sugiere que una menor susceptibilidad a clasificar FN.

En cuanto a los valores F, el mejor desempeño lo tiene el F2. El valor F2 busca minimizar la cantidad de FN al darle más peso a la sensibilidad (recall) que a la precisión. Para aumentar la sensibilidad, es necesario minimizar el número de FN, mientras que para aumentar la precisión es necesario minimizar el número de FP. Este es un caso donde, tal como se ha mencionado anteriormente, el impacto de los FN es mayor que los FP. Por lo tanto, es en cierta medida conveniente otorgarle mayor importancia a la sensibilidad que a la precisión.

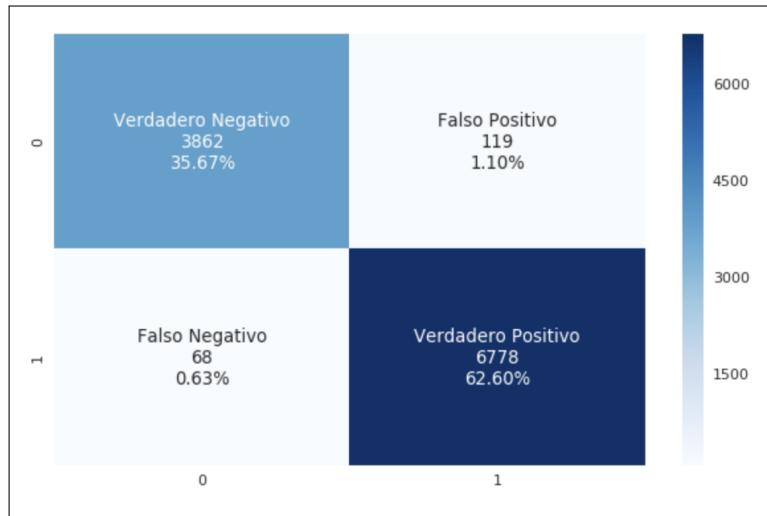


Figura 3.13: Matriz de confusión obtenida a partir del modelo ganador: Model-VGG16-(a)

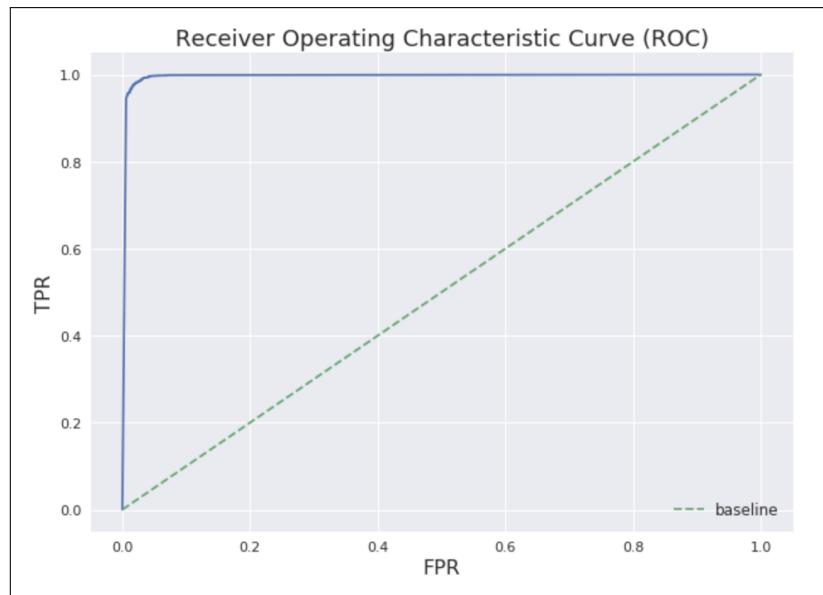


Figura 3.14: Curva ROC para el modelo ganador: Model-VGG16-(a)

4 Conclusiones y trabajos futuros

El diagnóstico de la enfermedad de Alzheimer es una tarea compleja, incierta, y la mayoría de las veces está condicionada por la subjetividad y experiencia del profesional médico involucrado. Asimismo, la aparición tardía de los síntomas dificulta la posibilidad de realizar un diagnóstico temprano, sustancial para poder realizar un tratamiento que retrase el progreso de la enfermedad. El avance computacional en el procesamiento de imágenes así como la incorporación de aprendizaje profundo en el análisis de las mismas, parece ser muy prometedor y es un área que aún le queda mucho por desarrollar.

En este trabajo se han propuesto dos modelos de redes convolucionales basados en arquitecturas conocidas para afrontar el problema de clasificación temprana de EA. Los resultados obtenidos parecen ser alentadores. Renombrados estudios que utilizan criterios de diagnóstico mundialmente aceptados, tales como DSM-IV¹, NINCDS-ADRDA y NIAA-AA, se han comparado con los indicadores obtenidos en el presente estudio. En cuanto a sensibilidad y especificidad, si bien el rango de los valores obtenidos en la bibliografía citada es muy amplio, si se toman los de la NIA-AA (80.9% y 94.0% respectivamente) o el meta-análisis de Bloudek et al. (0.83% y 0.85% respectivamente), los modelos propuestos han evidenciado una mayor capacidad discriminativa con sensibilidades superiores a 98% y especificidades mayores a 95%. Más aún, todos los indicadores dieron resultados por encima de 95%, resaltando la calidad del modelo en cuestión.

De todos modos, nuevamente hay que hacer hincapié en que la superioridad de los indicadores obtenidos frente a los estudios en cuestión puede estar en gran medida sesgada por el grupo demográfico considerado, y sería erróneo considerarlo como validación externa. Habría que evaluar el modelo neuronal incorporando imágenes que no sean exclusivamente de individuos de Estados Unidos y Canadá, sino de todas partes del mundo, a modo de

¹Esta versión fue publicada en 1994. Hoy en día existe una nueva versión que fue publicada en el 2013 (DSM-V), donde los criterios diagnósticos se actualizaron sustancialmente.

lograr una herramienta que sea aplicable a nivel mundial.

En cuanto a los métodos diagnósticos, al momento es sugestiva la preminencia del diagnóstico por imágenes médicas por sobre el diagnóstico clínico para EA con los estudios comparados. Acorde a lo que evidencian los estudios externos analizados, la incorporación de neuroimágenes a la hora de detectar la EA ha brindado resultados sustancialmente más acertados y precisos que los exámenes clínicos. Particularmente, las IRM han prevalecido frente a las otras técnicas mencionadas. Sin embargo, parecería que la implementación de diversas formas de diagnóstico aumenta la posibilidad de realizar un diagnóstico correcto. No se trata de eliminar los análisis clínicos, sino complementarlos y respaldarlos con técnicas más robustas y objetivas. La biopsia y necropsia sigue siendo el método determinante a la hora de diagnosticar la enfermedad.

Una de las cuestiones que queda aún por resolver es la precocidad del diagnóstico. Uno de los mayores desafíos de esta enfermedad es poder detectarla de manera temprana. Si bien en este estudio se han incorporado imágenes de personas tanto sanas como enfermas menores a 65 años, la mayoría corresponden a individuos de edades avanzadas, donde el daño cerebral se puede advertir fácilmente. Para poder generalizar esta propuesta y aplicarla en pacientes más jóvenes, particularmente asintomáticos, es necesario reevaluar el modelo con un volumen de neuroimágenes de individuos menores a 65 años considerablemente más grande, y si bien la disponibilidad de datos es cada vez mayor, la generación de los mismos es una tarea engorrosa que requiere un largo período de seguimiento de los pacientes. El reto reside justamente en poder categorizar imágenes de individuos que desarrollarán Alzheimer, pero que al ojo humano parecerían ser normales; encontrar un modelo neuronal que perciba diferencias que a los ojos de los especialistas en la materia son inapreciables.

De este modo, dejando en evidencia las limitaciones de este trabajo, el próximo paso a seguir sería en primera instancia, considerar una base de datos más abarcativa, con individuos de diferentes ciudades y países, y pertenecientes a grupos étnicos variados y bien categorizados. Con esa información, se podrían obtener datos sumamente interesantes que

respalden el desarrollo de EA con el nivel de ingreso, educación y acceso a la salud, así como también información genética, tal como se discutió en el marco teórico.

Más importante aún sería poder contar con una amplia variedad de imágenes de individuos menores a 60-65 años que aún no evidencian grandes daños cerebrales para poder abordar el problema de diagnosticar tempranamente. Sería muy enriquecedor entrenar otro modelo con la misma red pero utilizando exclusivamente imágenes donde el daño cerebral es prácticamente imperceptible. De esta forma se podría lograr justamente anticipar quiénes desarrollarán EA con el paso del tiempo, aún cuando el ojo humano no puede discernir entre cerebros sanos y enfermos. Una herramienta preclínica podría ser de gran utilidad para empezar a tratar la enfermedad aún cuando los síntomas no están presentes.

Bibliografía

- [1] Jeffrey De Fauw, Joseph R Ledsam, Bernardino Romera-Paredes, Stanislav Nikolov, Nenad Tomasev, Sam Blackwell, Harry Askham, Xavier Glorot, Brendan O'Donoghue, Daniel Visentin, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature medicine*, 24(9):1342–1350, 2018.
- [2] Siqi Liu, Sidong Liu, Weidong Cai, Sonia Pujol, Ron Kikinis, and Dagan Feng. Early diagnosis of alzheimer's disease with deep learning. In *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*, pages 1015–1018. IEEE, 2014.
- [3] Wenqing Sun, Bin Zheng, and Wei Qian. Computer aided lung cancer diagnosis with deep learning algorithms. In *Medical imaging 2016: computer-aided diagnosis*, volume 9785, page 97850Z. International Society for Optics and Photonics, 2016.
- [4] Jie-Zhi Cheng, Dong Ni, Yi-Hong Chou, Jing Qin, Chui-Mei Tiu, Yeun-Chung Chang, Chiun-Sheng Huang, Dinggang Shen, and Chung-Ming Chen. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans. *Scientific reports*, 6(1):1–13, 2016.
- [5] MR Avendi, Arash Kheradvar, and Hamid Jafarkhani. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri. *Medical image analysis*, 30:108–119, 2016.
- [6] John Edward Cooper. *ICD-10 Classification of Mental and Behavioural Disorders*. Churchill Livingston, 1984.
- [7] Asociación Psiquiátrica Americana APA. Diagnostic and statistical manual of mental disorders. *DSM-IV*, 1994.
- [8] Juebin Huang. Apraxia - trastornos neurológicos. <https://www.msdmanuals.com/es/professional/trastornos-neurológicos/función-y-disfunción-de-los-lóbulos-cerebrales/apraxia>, 2017. Consultado: 17 Oct. 2019.
- [9] Julie Hugo and Mary Ganguli. Dementia and cognitive impairment: epidemiology, diagnosis, and treatment. *Clinics in geriatric medicine*, 30(3):421–442, 2014.
- [10] Swedish Council on Technology Assessment in Health Care et al. Dementia–etiology and epidemiology. a systematic review. *Stockholm: The Swedish Council on Technology Assessment in Health Care*, 2008.
- [11] Emma Nichols, Cassandra El Szoek, Stein Emil Vollset, Nooshin Abbasi, Foad Abd-Allah, Jemal Abdela, Miloud Taki Eddine Aichour, Rufus O Akinyemi, Fares Alahdab, Solomon W Asgedom, et al. Global, regional, and national burden of alzheimer's

- disease and other dementias, 1990–2016: a systematic analysis for the global burden of disease study 2016. *The Lancet Neurology*, 18(1):88–106, 2019.
- [12] Martin James Prince. *World Alzheimer Report 2015: the global impact of dementia: an analysis of prevalence, incidence, cost and trends*. Alzheimer's Disease International, 2015.
 - [13] World Health Organization, B Annex, et al. tables of health statistics by country, who region and globally. *En: World health statistics*, pages 103–20, 2016.
 - [14] Geneviève Chêne, Alexa Beiser, Rhoda Au, Sarah R Preis, Philip A Wolf, Carole Dufouil, and Sudha Seshadri. Gender and incidence of dementia in the framingham heart study from mid-adult life. *Alzheimer's & Dementia*, 11(3):310–320, 2015.
 - [15] Liesi E Hebert, Paul A Scherr, Judith J McCann, Laurel A Beckett, and Denis A Evans. Is the risk of developing alzheimer's disease greater for women than for men? *American journal of epidemiology*, 153(2):132–136, 2001.
 - [16] LL Barnes, RS Wilson, JA Schneider, JL Bienias, DA Evans, and DA Bennett. Gender, cognitive decline, and risk of ad in older persons. *Neurology*, 60(11):1777–1781, 2003.
 - [17] Alexander L Chin, Selamawit Negash, and Roy Hamilton. Diversity and disparity in dementia: The impact of ethnoracial differences in alzheimer's disease. *Alzheimer disease and associated disorders*, 25(3):187, 2011.
 - [18] M-X Tang, Peter Cross, H Andrews, DM Jacobs, S Small, K Bell, Carol Merchant, R Lantigua, Rosanne Costa, Yaakov Stern, et al. Incidence of ad in african-americans, caribbean hispanics, and caucasians in northern manhattan. *Neurology*, 56(1):49–56, 2001.
 - [19] RTI Press. Racial and ethnic disparities among individuals with alzheimer's disease in the united states: A literature review. 2014.
 - [20] Norman B Anderson, Rodolfo A Bulatao, Barney Cohen, Panel on Race, National Research Council, et al. Ethnic differences in dementia and alzheimer's disease. In *Critical perspectives on racial and ethnic differences in health in late life*. National Academies Press (US), 2004.
 - [21] Association Alzheimer's. 2015 alzheimer's disease facts and figures. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, 11(3):332, 2015.
 - [22] U.S. Census Bureau. U.S. median household income up in 2018 from 2017. <https://www.census.gov/library/stories/2019/09/us-median-household-income-up-in-2018-from-2017.html>, 2019. Consultado: 18 Oct. 2019.
 - [23] Barry J Gurland, David E Wilder, Rafael Lantigua, Yaakov Stern, Jiming Chen, Eloise HP Killeffer, and Richard Mayeux. Rates of dementia in three ethnoracial groups. *International journal of geriatric psychiatry*, 14(6):481–493, 1999.

- [24] Clement T Loy, Peter R Schofield, Anne M Turner, and John BJ Kwok. Genetics of dementia. *The Lancet*, 383(9919):828–840, 2014.
- [25] David M Holtzman, Joachim Herz, and Guojun Bu. Apolipoprotein e and apolipoprotein e receptors: normal biology and roles in alzheimer disease. *Cold Spring Harbor perspectives in medicine*, 2(3):a006312, 2012.
- [26] Liara Rizzi, Idiane Rosset, and Matheus Roriz-Cruz. Global epidemiology of dementia: Alzheimer's and vascular types. *BioMed research international*, 2014, 2014.
- [27] Alzheimer's Association. Alzheimer's disease and dementia - las 10 señales. <https://alz.org/alzheimer-demencia/las-10-senales>, 2018. Consultado: 17 Oct. 2019.
- [28] Carmen Echávarri, Saartje Burgmans, Harry Uylings, Manuel J Cuesta, Victor Peralta, Wouter Kamphorst, Annemieke JM Rozemuller, and Frans RJ Verhey. Neuropsychiatric symptoms in alzheimer's disease and vascular dementia. *Journal of Alzheimer's Disease*, 33(3):715–721, 2013.
- [29] Manuel Fernández-Martínez, Jessica Castro, Ana Molano, Juan J Zaranz, Rosa Monica Rodrigo, and Rafael Ortega. Prevalence of neuropsychiatric symptoms in alzheimer's disease and vascular dementia. *Current Alzheimer Research*, 5(1):61–69, 2008.
- [30] KM Cullen and GM Halliday. Neurofibrillary degeneration and cell loss in the nucleus basalis in comparison to cortical alzheimer pathology. *Neurobiology of aging*, 19(4):297–306, 1998.
- [31] Alan King Lun Liu, Raymond Chuen-Chung Chang, Ronald KB Pearce, and Steve M Gentleman. Nucleus basalis of meynert revisited: anatomy, history and differential involvement in alzheimer's and parkinson's disease. *Acta neuropathologica*, 129(4):527–540, 2015.
- [32] William Samuel, Robert D Terry, Richard DeTeresa, Nelson Butters, and Eliezer Masliah. Clinical correlates of cortical and nucleus basalis pathology in alzheimer dementia. *Archives of Neurology*, 51(8):772–778, 1994.
- [33] Ph Scheltens, F Barkhof, D Leys, E Ch Wolters, R Ravid, and W Kamphorst. Histopathologic correlates of white matter changes on mri in alzheimer's disease and normal aging. *Neurology*, 45(5):883–888, 1995.
- [34] Matthew Bobinski, MJ De Leon, J Wegiel, S Desanti, A Convit, LA Saint Louis, H Rusinek, and HM Wisniewski. The histological validation of post mortem magnetic resonance imaging-determined hippocampal volume in alzheimer's disease. *Neuroscience*, 95(3):721–725, 1999.
- [35] Paul M Thompson, Kiralee M Hayashi, Greig I de Zubicaray, Andrew L Janke, Stephen E Rose, James Semple, Michael S Hong, David H Herman, David Gravano,

- David M Doddrell, et al. Mapping hippocampal and ventricular change in alzheimer disease. *Neuroimage*, 22(4):1754–1766, 2004.
- [36] Dennis Chan, Nick C Fox, Rachael I Scahill, William R Crum, Jennifer L Whitwell, Guy Leschziner, Alex M Rossor, John M Stevens, Lisa Cipolotti, and Martin N Rossor. Patterns of temporal lobe atrophy in semantic dementia and alzheimer’s disease. *Annals of neurology*, 49(4):433–442, 2001.
- [37] Bradford C Dickerson, I Goncharova, MP Sullivan, C Forchetti, RS Wilson, DA Bennett, Laurel A Beckett, and L deToledo Morrell. Mri-derived entorhinal and hippocampal atrophy in incipient and very mild alzheimer’s disease. *Neurobiology of aging*, 22(5):747–754, 2001.
- [38] Keith A Johnson, Nick C Fox, Reisa A Sperling, and William E Klunk. Brain imaging in alzheimer disease. *Cold Spring Harbor perspectives in medicine*, 2(4):a006213, 2012.
- [39] Raghbir Singh Khandpur. *Biomedical instrumentation: Technology and applications*, volume 1. McGraw-Hill New York, 2005.
- [40] David C. Preston. Magnetic resonance imaging (mri) of the brain and spine: Basics. <https://casemed.case.edu/clerkships/neurology/Web%20Neurorad/MRI%20Basics.htm>, 2006. Consultado: 2 Nov. 2019.
- [41] Case School of Medicine Case Western Reserve University. Magnetic resonance imaging. <http://casemed.case.edu/clerkships/neurology/NeurLrngObjectives/MRI.htm>, 2004. Consultado: 14 Nov. 2019.
- [42] Vijay PB Grover, Joshua M Tognarelli, Mary ME Crossey, I Jane Cox, Simon D Taylor-Robinson, and Mark JW McPhail. Magnetic resonance imaging: principles and techniques: lessons for clinicians. *Journal of clinical and experimental hepatology*, 5(3):246–255, 2015.
- [43] Giovanni B Frisoni, Nick C Fox, Clifford R Jack Jr, Philip Scheltens, and Paul M Thompson. The clinical use of structural mri in alzheimer disease. *Nature Reviews Neurology*, 6(2):67, 2010.
- [44] Ronald J Killiany, Teresa Gomez-Isla, Mark Moss, Ron Kikinis, Tamas Sandor, Ferenc Jolesz, Rudolph Tanzi, Kenneth Jones, Bradley T Hyman, and Marilyn S Albert. Use of structural magnetic resonance imaging to predict who will get alzheimer’s disease. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 47(4):430–439, 2000.
- [45] Stefan Klöppel, Cynthia M Stonnington, Josephine Barnes, Frederick Chen, Carlton Chu, Catriona D Good, Irina Mader, L Anne Mitchell, Ameet C Patel, Catherine C Roberts, et al. Accuracy of dementia diagnosis—a direct comparison between radiologists and a computerized method. *Brain*, 131(11):2969–2974, 2008.

- [46] Adrien Payan and Giovanni Montana. Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks. *arXiv preprint arXiv:1502.02506*, 2015.
- [47] Robert Hecht-Nielsen. Neurocomputer applications. In *Neural computers*, pages 445–453. Springer, 1989.
- [48] Imad A Basheer and Maha Hajmeer. Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods*, 43(1):3–31, 2000.
- [49] Anil K Jain, Jianchang Mao, and K Moidin Mohiuddin. Artificial neural networks: A tutorial. *Computer*, 29(3):31–44, 1996.
- [50] Yoav Mintz and Ronit Brodie. Introduction to artificial intelligence in medicine. *Minimally Invasive Therapy & Allied Technologies*, 28(2):73–81, 2019.
- [51] Paras Lakhani and Baskaran Sundaram. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 284(2):574–582, 2017.
- [52] Sérgio Pereira, Adriano Pinto, Victor Alves, and Carlos A Silva. Brain tumor segmentation using convolutional neural networks in mri images. *IEEE transactions on medical imaging*, 35(5):1240–1251, 2016.
- [53] Wei Shen, Mu Zhou, Feng Yang, Caiyun Yang, and Jie Tian. Multi-scale convolutional neural networks for lung nodule classification. In *International Conference on Information Processing in Medical Imaging*, pages 588–599. Springer, 2015.
- [54] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4):611–629, 2018.
- [55] The Area Under an ROC Curve. The area under an roc curve. <http://gim.unmc.edu/dxtests/roc3.htm>. Consultado: 3 Jul. 2020.
- [56] Guy McKhann, David Drachman, Marshall Folstein, Robert Katzman, Donald Price, and Emanuel M Stadlan. Clinical diagnosis of alzheimer's disease: Report of the nincds-adrda work group* under the auspices of department of health and human services task force on alzheimer's disease. *Neurology*, 34(7):939–939, 1984.
- [57] American Psychiatric Association et al. *Guía de consulta de los criterios diagnósticos del DSM-5®: Spanish Edition of the Desk Reference to the Diagnostic Criteria From DSM-5®*. American Psychiatric Pub, 2014.
- [58] Guy M McKhann, David S Knopman, Howard Chertkow, Bradley T Hyman, Clifford R Jack Jr, Claudia H Kawas, William E Klunk, Walter J Koroshetz, Jennifer J Manly, Richard Mayeux, et al. The diagnosis of dementia due to alzheimer's disease: recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):263–269, 2011.

- [59] Marilyn S Albert, Steven T DeKosky, Dennis Dickson, Bruno Dubois, Howard H Feldman, Nick C Fox, Anthony Gamst, David M Holtzman, William J Jagust, Ronald C Petersen, et al. The diagnosis of mild cognitive impairment due to alzheimer's disease: recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):270–279, 2011.
- [60] Fran ois Boiler, Oscar L Lopez, and John Moossy. Diagnosis of dementia: clinico-pathologic correlations. *Neurology*, 39(1):76–76, 1989.
- [61] Kim A Jobst, Lin PD Barnetson, Basil J Shepstone, et al. Accurate prediction of histologically confirmed alzheimer's disease and the differential diagnosis of dementia: the use of nincds-adrda and dsm-iii-r criteria, spect, x-ray ct, and apo e4 in medial temporal lobe dementias. *International Psychogeriatrics*, 10(3):271–302, 1998.
- [62] Daniel E Wollman and Isak Prohovnik. Sensitivity and specificity of neuroimaging for the diagnosis of alzheimer's disease. *Dialogues in clinical neuroscience*, 5(1):89, 2003.
- [63] Thomas G Beach, Sarah E Monsell, Leslie E Phillips, and Walter Kukull. Accuracy of the clinical diagnosis of alzheimer disease at national institute on aging alzheimer disease centers, 2005–2010. *Journal of neuropathology and experimental neurology*, 71(4):266–273, 2012.
- [64] Marwan N Sabbagh, Lih-Fen Lue, Daniel Fayard, and Jiong Shi. Increasing precision of clinical diagnosis of alzheimer's disease using a combined algorithm incorporating clinical and novel biomarker data. *Neurology and therapy*, 6(1):83–95, 2017.
- [65] Jennifer M Harris, Jennifer C Thompson, Claire Gall, Anna MT Richardson, David Neary, Daniel du Plessis, Piayali Pal, David MA Mann, Julie S Snowden, and Matthew Jones. Do nia-aa criteria distinguish alzheimer's disease from frontotemporal dementia? *Alzheimer's & Dementia*, 11(2):207–215, 2015.
- [66] Lisa M Bloudek, D Eldon Spackman, Michael Blankenburg, and Sean D Sullivan. Review and meta-analysis of biomarkers and diagnostic imaging in alzheimer's disease. *Journal of Alzheimer's Disease*, 26(4):627–645, 2011.
- [67] W. Neuroimaging. Spm - statistical parametric mapping. <https://www.fil.ion.ucl.ac.uk/spm/>, 2014. Consultado: 5 Dic. 2019.
- [68] Pierrick Coup . Mri denoising software. <https://sites.google.com/site/pierrickcoupe/softwares/denoising-for-medical-imaging/mri-denoising/mri-denoising-software>, 2019. Consultado: 9 Dic. 2019.
- [69] H kon Gudbjartsson and Samuel Patz. The rician distribution of noisy mri data. *Magnetic resonance in medicine*, 34(6):910–914, 1995.
- [70] Stephen M Smith. Fast robust automated brain extraction. *Human brain mapping*, 17(3):143–155, 2002.

- [71] Yongyue Zhang, Michael Brady, and Stephen Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, 20(1):45–57, 2001.
- [72] Vladimir S Fonov, Alan C Evans, Robert C McKinstry, CR Almlí, and DL Collins. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, (47):S102, 2009.
- [73] Vladimir Fonov, Alan C Evans, Kelly Botteron, C Robert Almlí, Robert C McKinstry, D Louis Collins, Brain Development Cooperative Group, et al. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage*, 54(1):313–327, 2011.
- [74] D Louis Collins, Alex P Zijdenbos, Wim FC Baaré, and Alan C Evans. Animal+ insect: improved cortical structure segmentation. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 210–223. Springer, 1999.
- [75] Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002.
- [76] Mark Jenkinson and Stephen Smith. A global optimisation method for robust affine registration of brain images. *Medical image analysis*, 5(2):143–156, 2001.
- [77] Douglas N Greve and Bruce Fischl. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48(1):63–72, 2009.
- [78] Fetal-Neonatal Neuroimaging and Developmental Science Center (FNNDSC). med2image 1.1.2. <https://pypi.org/project/med2image/1.1.2/>, 2018. Consultado: 16 Dic. 2019.
- [79] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [80] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [81] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [82] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [83] Max Ferguson, Ronay Ak, Yung-Tsun Tina Lee, and Kincho H Law. Automatic localization of casting defects with convolutional neural networks. In *2017 IEEE international conference on big data (big data)*, pages 1726–1735. IEEE, 2017.

A Apéndice

En este apéndice se incluyen los dos modelos que mejores resultados obtuvieron. En la sección A.1 se encuentra el modelo ganador basados en la red de tipo VGG16, y en la sección A.2 el ganador inspirado en AlexNet.

A.1 Modelos de tipo VGG16

```
1 import tensorflow
2 from tensorflow.keras.models import Model, Sequential
3 from tensorflow.keras.optimizers import Adam
4 from tensorflow.keras.callbacks import EarlyStopping,
5     ModelCheckpoint,
6     ReduceLROnPlateau
7 from tensorflow.keras.layers import Input, Dense, Activation,
8     BatchNormalization, Flatten, Conv2D
9 from tensorflow.keras.layers import MaxPooling2D, Dropout
10 from tensorflow.keras.applications.vgg16 import VGG16
11
12 #MODELO GANADOR VGG16
13
14 model_vgg = VGG16(include_top=False, input_shape=(233,233,3))
15 model = Sequential()
16 model.add(model_vgg)
17 model.add(Flatten())
18 #Fully-connected layers
19 model.add(Dense(84, activation='relu'))
20 model.add(Dropout(0.5))
21 model.add(Dense(84, activation='relu'))
22 model.add(Dropout(0.5))
```

```

23 model.add(Dense(32, activation='relu'))
24 model.add(Dropout(0.5))
25
26 #Single-node layer with sigmoid function for two-class
27 # categorization
28 model.add(Dense(units=1, activation = 'sigmoid'))
29
30 model.compile(optimizer="SGD", loss="binary_crossentropy", metrics=["
31 accuracy"])
32
33 checkpoint_acc = tensorflow.keras.callbacks.ModelCheckpoint(
34     './x/LogsVGG/weights04bis.{epoch:02d}-{val_loss:.2f}.hdf5',
35     monitor='val_accuracy',
36     verbose=1,
37     save_best_only=True,
38     mode='max')
39
40 early_stop = tensorflow.keras.callbacks.EarlyStopping(monitor='
41 val_accuracy',
42 patience=10)
43
44 reduce_lr = tensorflow.keras.callbacks.ReduceLROnPlateau(
45     monitor="val_accuracy",
46     factor=0.2,
47     patience=4,
48     verbose=0,
49     mode="auto",
50     min_delta=0,
51     min_lr=0)
52
53 model.fit_generator(

```

```
52     train_generator_x,
53     steps_per_epoch = train_generator_x.samples // batch_size,
54     epochs = num_epochs,
55     validation_data = validation_generator_x,
56     validation_steps = validation_generator_x.samples //
57     batch_size,
58     callbacks=[checkpoint_acc, early_stop, reduce_lr])
```

A.2 Modelos de tipo AlexNet

```
1
2 #MODELO GANADOR ALEXNET
3
4 model = Sequential()
5
6 model.add(Conv2D(filters=6, kernel_size=(3,3), strides=(1, 1),
7 padding='same',
8 activation='relu', input_shape=(233, 233, 3)))
9
10 model.add(MaxPooling2D(pool_size =(2,2), strides=(1,1), padding='
11 valid'))
12
13 model.add(Conv2D(filters=16, kernel_size=(3,3), strides=(1, 1),
14 padding='valid',
15 activation='relu'))
16
17 model.add(MaxPooling2D(pool_size=(2, 2), strides=(2, 2), padding='
18 valid'))
19
20 model.add(Conv2D(120, kernel_size=(3, 3), strides=(1, 1), activation
21 = 'relu',
22 padding='valid'))
```

```

18
19 model.add(Flatten())
20
21 model.add(Dense(84, activation='relu'))
22 model.add(Dropout(0.5))
23 model.add(Dense(84, activation='relu'))
24 model.add(Dropout(0.5))
25 model.add(Dense(32, activation='relu'))
26 model.add(Dropout(0.5))
27 model.add(Dense(1, activation = 'sigmoid'))
28
29
30 model.compile(optimizer="SGD", loss="binary_crossentropy", metrics=["accuracy"])
31 print(model3.summary())
32
33 checkpoint_acc = tensorflow.keras.callbacks.ModelCheckpoint(
34     './x/Logs/weights23.{epoch:02d}-{val_loss:.2f}.hdf5',
35     monitor='val_accuracy',
36     verbose=1,
37     save_best_only=True,
38     mode='max')
39
40 early_stop = tensorflow.keras.callbacks.EarlyStopping(monitor='val_accuracy',
41 patience=10)
42
43 reduce_lr = tensorflow.keras.callbacks.ReduceLROnPlateau(
44     monitor="val_accuracy",
45     factor=0.2,
46     patience=4,
47     verbose=0,

```

```
48     mode="auto",
49     min_delta=0,
50     min_lr=0)
51
52 model.fit_generator(
53     train_generator_x,
54     steps_per_epoch = train_generator_x.samples // batch_size,
55     epochs = num_epochs,
56     validation_data = validation_generator_x,
57     validation_steps = validation_generator_x.samples //
58     batch_size,
59     callbacks=[csv_logger, checkpoint_acc, early_stop, reduce_lr
60   ])
```