



Universidad
Zaragoza

Trabajo Fin de Grado

Selección de características en deep-learning
para el análisis de la enfermedad de
Alzheimer a partir de imagen y genética

Deep feature selection for the analysis of
Alzheimer's disease from imaging-genetics

Autor/es

Javier Armunia Hinojosa

Director/es

Mónica Hernández Giménez

Elvira Mayordomo Cámara

Titulación del autor

Grado en Ingeniería Informática

Escuela de Ingeniería y Arquitectura

2021



DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD

(Este documento debe entregarse en la Secretaría de la EINA, dentro del plazo de depósito del TFG/TFM para su evaluación).

D./D^a. Javier Armunia Hinojosa ,en

aplicación de lo dispuesto en el art. 14 (Derechos de autor) del Acuerdo de 11 de septiembre de 2014, del Consejo de Gobierno, por el que se aprueba el Reglamento de los TFG y TFM de la Universidad de Zaragoza,

Declaro que el presente Trabajo de Fin de (Grado/Máster)

Grado en Ingeniería Informática (Título del Trabajo)

Selección de características en deep-learning para el análisis de la enfermedad de Alzheimer a partir de imagen y genética

es de mi autoría y es original, no habiéndose utilizado fuente sin ser citada debidamente.

Zaragoza, 20 de Enero de 2021

Fdo: Javier Armunia Hinojosa

1. RESUMEN

Selección de características en Deep-learning para el análisis de la enfermedad de Alzheimer a partir de imagen y genética

El objetivo de este trabajo es reproducir el estudio realizado en *L. Yuanyuan et al., Deep Feature Selection and Causal Analysis of Alzheimer's Disease, Frontiers in Neuroscience, 2019* [1]. El estudio consiste en el desarrollo de un sistema que integra deep learning, feature selection, inferencia causal y análisis genéticos y de imágenes para la predicción y una mejor comprensión de las causas de la enfermedad de Alzheimer.

En el estudio utilizamos imágenes cerebrales en tres dimensiones, obtenidas mediante la técnica de resonancia magnética Diffusion Tensor Imaging (DTI) y datos genéticos de los pacientes. Contamos con un dataset de imágenes DTI de 112 pacientes (51 AD y 62 CN), cada uno con imágenes en cuatro puntos temporales: baseline, 6 meses, 12 meses y 24 meses. El dataset de datos genéticos se compone del genoma de 79 pacientes, del cual tras su procesamiento hemos obtenido los 40 genes con los que hemos realizado los experimentos.

El trabajo desarrollado ha consistido, en realizar un preprocesado de las imágenes consistente en un registro lineal y no lineal a un atlas estándar y un control de calidad e imputación de los datos genéticos utilizando los datos de referencia de 1000 Genomes. A continuación se ha diseñado un modelo de red neuronal convolucional de tres dimensiones para obtener la probabilidad de que un paciente padezca Alzheimer, estimada a partir de una imagen DTI de su cerebro. Posteriormente, se ha realizado un procedimiento de interpretabilidad de redes neuronales conocido como Prediction Difference Analysis, consistente en ocultar parte de la información de la imagen de entrada para observar la variación en la predicción con el objetivo de obtener las regiones cerebrales a las que la red neuronal da más importancia a la hora de predecir la probabilidad de Alzheimer de un paciente.

Finalmente, se han realizado dos experimentos de descubrimiento causal que buscan determinar la dirección de la causalidad entre la enfermedad de Alzheimer y las regiones cerebrales, y entre las regiones cerebrales y los genes, es decir en descubrir si la dirección de la causalidad es $X \rightarrow Y$ (X causa Y) o $Y \rightarrow X$ (Y causa X). Esto se ha abordado mediante la construcción de dos redes generativas adversarias condicionales (CGAN) que aprenderán a generar la distribución de X dado Y y la distribución de Y dado X . Los modelos resultantes se han evaluado mediante un Classifier 2 Sample Test (C2ST), el cual consiste en un método de clasificación que intenta distinguir entre los ejemplos generados por una CGAN y los reales. Si el clasificador no es capaz de distinguir entre los dos conjuntos, es decir, si obtiene una precisión cercana al 50%, significa que la CGAN ha conseguido modelar de forma exitosa la distribución objetivo. Determinaremos que la dirección de causalidad correcta es la que obtenga una precisión menor en su C2ST.

Mediante el experimento de interpretabilidad se ha descubierto que el ventrículo es una región a la que la red neuronal presta especial atención al predecir Alzheimer a partir de las imágenes de DTI. Con los experimentos de búsqueda causal hemos encontrado relevancia estadística en la causalidad de todas las ROI analizadas con la dirección $ROI \rightarrow AD$ y hemos obtenido relevancia en 14 de los genes analizados, destacando el gen CD33, relacionado con el Alzheimer y el SCYL1 relacionado con la atrofia cerebral.

GLOSARIO

- AD: Alzheimer's Disease. Paciente en el grupo con enfermedad de Alzheimer.
- CN: Control Normal. Paciente en el grupo de control.
- ADNI: Alzheimer's Disease Neuroimaging Initiative. Repositorio online de datos para su uso en investigaciones relacionadas con el Alzheimer.
- DTI: Diffusion Tensor Imaging
- EMCI: Early Mild Cognitive Impairment.
- LMCI: Late Mild Cognitive Impairment.
- FC: Fully Connected.
- ROI: Region Of Interest. Región de interés.
- C2ST: Classifier 2 sample test. Test de dos muestras mediante clasificador.
- GAN: Generative Adversarial Network.
- CGAN: Conditional Generative Adversarial Network.
- CAD: Computer Aided Diagnosis.

2. TABLA DE CONTENIDOS

Resumen	2
GLOSARIO	3
Tabla de contenidos	4
Introducción	6
Objetivo del TFG	6
Introducción sobre la enfermedad de Alzheimer	6
Diagnóstico	6
ADNI	7
Aprendizaje automático en medicina	7
Interpretabilidad del aprendizaje automático	8
Resultados obtenidos	8
Método desarrollado	8
Recopilación de datos	8
Imágenes	8
Datos genéticos	9
Procesamiento de las imágenes	10
Data augmentation	10
CNN	11
Elección del optimizador	12
Arquitectura del modelo	12
Selección del learning rate	12
Selección del número de filtros y número de neuronas	13
Selección de los parámetros de regularización	13
Modelo final	14
Selección de características “profundas” de imágenes DTI	15
FPCA en imágenes	15
Procesamiento de datos genéticos	16
Redes generativas adversarias condicionales y test de dos muestras mediante clasificación para descubrimiento de causalidad	16
CGAN	16
C2ST	17
	4

Descubrimiento causal	18
Arquitectura del experimento causal	19
Resultados	20
Resultados CNN	20
Resultados de la selección de características profundas	21
Baseline	22
6 meses	22
1 año	22
2 años	22
Resultados causalidad ROI vs AD	23
ROI	23
P-valor	23
Resultados causalidad ROI vs Gen	23
Conclusiones	25
Conclusiones	25
Diagrama de Gantt	25
Líneas futuras de trabajo	26
Referencias	27

3. INTRODUCCIÓN

OBJETIVO DEL TFG

El objetivo de este TFG es reproducir el método de interpretabilidad y búsqueda de causalidad desarrollado en [1] de la forma más fiel posible. Con este objetivo, si bien no ha sido posible su reproducción exacta, se ha implementado el sistema utilizando datos obtenidos de la misma fuente y la arquitectura de los experimentos se ha implementado de la forma más fiel posible con la información disponible. Para ello se ha utilizado principalmente Python 3.6 con Tensorflow 2.1, R, herramientas de procesamiento de imágenes MRI como RNiftyReg y FSL, herramientas de procesamiento de datos genéticos como Plink, IMPUTE2, SHAPEIT y la herramienta de control de calidad de datos genéticos de McCarthy Tools.

INTRODUCCIÓN SOBRE LA ENFERMEDAD DE ALZHEIMER

La enfermedad de Alzheimer (AD) es una enfermedad neurodegenerativa que se manifiesta mediante un deterioro cognitivo progresivo y trastornos conductuales. Se caracteriza en su forma típica por una pérdida de la memoria inmediata y de otras capacidades mentales (tales como las capacidades cognitivas superiores), a medida que mueren las células nerviosas (neuronas) y se atrofian diferentes zonas del cerebro. La enfermedad suele tener una duración media aproximada, después del diagnóstico, de 10 años, aunque esto puede variar en proporción directa con la severidad de la enfermedad al momento del diagnóstico. La enfermedad de Alzheimer es la forma más común de demencia. Es una enfermedad incurable y terminal que aparece con mayor frecuencia en personas mayores de 65 años de edad, aunque también en raros casos puede desarrollarse a partir de los 40 años.

DIAGNÓSTICO

El diagnóstico y la predicción de la evolución de la enfermedad mediante evaluaciones clínicas y psicométricas son complejos [16]. El diagnóstico se basa en el estudio de la historia y la observación clínicas del paciente, la realizada por el profesional de la salud y la que es referida por los familiares. El estudio se centra en las características neurológicas y psicológicas, así como en la ausencia de condiciones alternativas (diagnóstico de exclusión). Durante las semanas o meses posteriores se realizan pruebas de memoria y de funcionamiento o evaluación intelectual. También se efectúan análisis de sangre y escáner para descartar diagnósticos alternativos. No existe un test pre-mortem para diagnosticar concluyentemente el Alzhéimer, por ello, el diagnóstico que se realiza en la práctica clínica es de probable enfermedad de Alzheimer. Se ha conseguido aproximar la certeza del diagnóstico a un 85%, pero el definitivo debe hacerse con pruebas histológicas sobre tejido cerebral, generalmente obtenidas en la autopsia post-mortem.

Ante la imposibilidad de realizar un examen histológico en pacientes vivos, las técnicas de diagnóstico mediante imagen surgen como un importante complemento en la mejora del diagnóstico y predicción de AD. La variación de la estructura, funcionalidad y efectos patológicos del cerebro pueden ser medidos mediante imágenes. De acuerdo con la recomendación del National Institute of Neurological and Communicative Disorders and Stroke–AD and Related Disorders Association (NINCDS-ADRDA) Work Group, la clasificación clínica de AD debería explorar los marcadores por imagen: Magnetic Resonance Imaging (MRI),

Diffusion Tensor Imaging (DTI), Positron Emission Tomography (PET), amyloid-PET, tau-PET y los marcadores de fluido cerebroespinal (CSF) neuronal anormal (tau y/o $A\beta$) [16][17][18][19].

Este Trabajo Final de Grado (TFG) se centra en el estudio de marcadores DTI que es una técnica de imagen mediante resonancia magnética utilizada para describir la microestructura del tejido biológico explotando la cuantificación de la difusión del agua en los tejidos. En la ilustración 1 vemos diferentes secciones sagital, coronal y axial de una de las imágenes DTI de nuestro dataset. En la imagen se aprecia el contraste proporcionado por la direccionalidad de las fibras en las estructuras cerebrales, que facilitan la difusión del agua en estas direcciones.

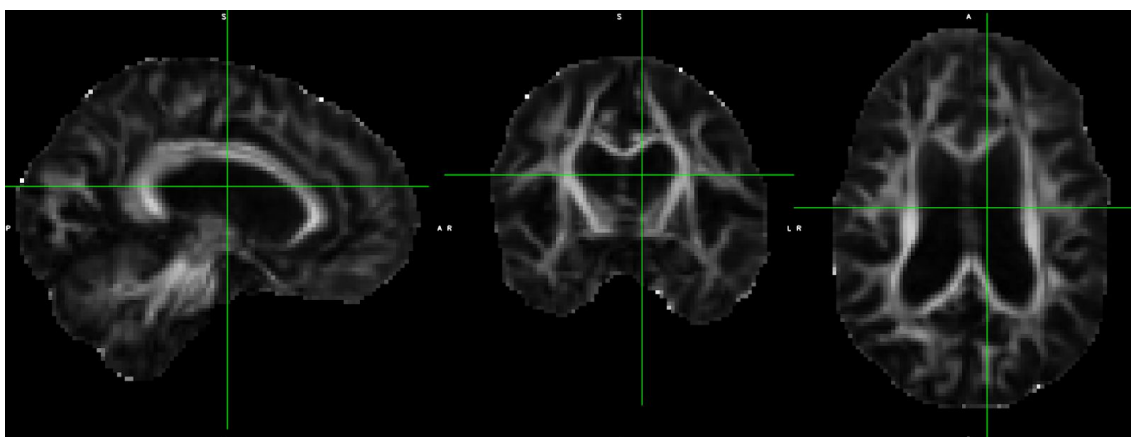


Ilustración 1. Vistas sagital, coronal y axial de una imagen de corrected FA DTI obtenida de la base de datos de ADNI.

ADNI

La iniciativa Alzheimer's Disease Neuroimaging Initiative ([ADNI](#)) es un estudio multicentro longitudinal diseñado para producir biomarcadores genéticos, bioquímicos, clínicos y de imagen para la detección temprana y seguimiento de la enfermedad de Alzheimer. ADNI comenzó en 2004, fundado como una colaboración público-privada. El estudio inicial de 5 años (ADNI-1) fue extendido 2 años más en 2009 gracias a una donación de Grand Opportunities (ADNI-GO), y posteriormente se volvió a extender en 2011 (ADNI-2) y en 2016 (ADNI-3).

En este TFG utilizamos datos de ADNI-2, que se planteó con el principal objetivo de establecer biomarcadores como predictores de la disminución cognitiva y como medidas del estado de los pacientes. Este subproyecto de ADNI duró cinco años y está compuesto por los pacientes existentes en ADNI-1 y ADNI-GO resultando en un total de 150 CN (Control Normal), 100 EMCI (Early Mild Cognitive Impairment), 150 LMCI (Late Mild Cognitive Impairment) y 150 AD (Alzheimer's Disease).

APRENDIZAJE AUTOMÁTICO EN MEDICINA

Ante el aumento de tamaño de los datasets de imágenes médicas, su análisis manual se vuelve tedioso y costoso. El diagnóstico asistido por computador (CAD) combina modelos computacionales y herramientas analíticas para datos con gran número de dimensiones surge como una gran opción para el diagnóstico y predicción de la enfermedad de Alzheimer. Los métodos más ampliamente utilizados en CAD incluyen discriminant analysis, regresión logística, random forest, redes neuronales y support vector machines. Ante la revolución producida por el deep learning en varias áreas como la conducción autónoma, el reconocimiento de imágenes, de texto o del habla y su gran potencial en el diagnóstico y

predicción de la tuberculosis [8], cancer [9], retinopatía diabética [10], enfermedades crónicas renales [11], AD y conversión desde MCI a AD [12]. Esto ha provocado un incremento en el interés de utilizar deep learning para aplicaciones médicas.

INTERPRETABILIDAD DEL APRENDIZAJE AUTOMÁTICO

A pesar de los grandes avances realizados en áreas como la visión por computador o el procesamiento de lenguaje natural, el deep learning es bien conocido como una caja negra debido a su difícil interpretabilidad para los humanos. Esta dificultad a la hora de explicar por qué un modelo ha tomado determinada decisión dificulta su aplicación en medicina. Uno de los problemas que surge de la falta de interpretabilidad es la dificultad de descubrir los posibles sesgos que sufra el modelo.

RESULTADOS OBTENIDOS

Se ha realizado un análisis de interpretabilidad en el modelo basado en redes neuronales convolucionales 3D desarrollado para obtener las regiones de interés cerebrales a las que el modelo da más importancia a la hora de predecir la enfermedad de Alzheimer obteniendo como región con mayor relevancia al ventrículo.

En el experimento de búsqueda causal realizado con el objetivo de hallar la dirección causal entre ROI y AD se ha determinado que la dirección causal es $ROI \rightarrow AD$ (ROI causa AD) y se han detectado todas las ROI analizadas como relevantes estadísticamente.

En último lugar, en el experimento de búsqueda causal entre ROI y genes se ha determinado en todos los casos la dirección $Gen \rightarrow ROI$. Destacamos la relevancia estadística obtenida en el experimento de los genes **SCYL1** relacionado con la atrofia del cerebelo y enfermedades neurodegenerativas y **CD33** relacionado con AD.

4. MÉTODO DESARROLLADO

4.1. RECOPIACIÓN DE DATOS

4.1.1. Imágenes

Todos los datos utilizados en este TFG han sido obtenidos de ADNI. Se han utilizado un total de 372 imágenes Corrected Fractional Anisotropy DTI pertenecientes a 112 pacientes divididos en dos clases: AD, Grupo de pacientes que padecen la enfermedad de Alzheimer (Alzheimer's Disease) y CN: Grupo de control (Control Normal). En la tabla 1 podemos ver la distribución de los pacientes según su sexo y grupo de estudio. Podemos apreciar que tenemos un mayor número de mujeres en el grupo de control que en el de AD, mientras que hay una cantidad similar de hombres en los dos grupos. También podemos ver que la desviación del grupo AD en los dos sexos es mayor que en CN destacando la gran diferencia en la desviación en el grupo de mujeres.

Grupo de estudio	Sexo	Número de pacientes	Edad media	Desviación estándar edad
AD	F	20	73.9	8.92
	M	30	76.6	8.21
CN	F	33	73.6	5.36
	M	29	75.4	7.13

Tabla 1: Distribución de los pacientes según grupo de estudio y sexo

Cada paciente tiene varias imágenes correspondientes a distintos puntos temporales según el momento en el que se obtuvieron las imágenes del paciente: baseline (el punto inicial del estudio para ese paciente), 6 meses, 1 año y 2 años. En la tabla 2 vemos la distribución de los pacientes agrupados por el punto temporal del estudio y por grupo de estudio. Destaca la gran cantidad de imágenes de baseline y la poca cantidad de imágenes de AD a los dos años. Podemos ver que los datos del grupo de AD están algo sesgados ya que las mujeres están infrarrepresentadas.

Punto temporal	Grupo de estudio	Número de imágenes	Edad media	Desviación estándar edad
Baseline	AD	48	75	8.67
	CN	50	72.5	6.17
6 meses	AD	31	76	8.3
	CN	37	73.7	5.94
1 año	AD	32	76.7	7.9
	CN	38	74.1	6.02
2 años	AD	6	81.8	9.12
	CN	31	75.9	5.6

Tabla 2: Distribución de las imágenes según el punto temporal y el grupo de estudio

Hay que destacar que en el artículo tomado como referencia para este TFG [1] no hay información suficiente para poder establecer el conjunto de imágenes utilizado para el estudio. Se intentó contactar con dos de los autores del trabajo sin éxito. Por ello, el conjunto de imágenes utilizado en este TFG es forzosamente diferente. En el artículo de referencia se utilizaron 100 imágenes del grupo CN y 51 de AD distribuidas entre los grupos baseline, 6 meses, 1 año y 2 años.

4.1.2. Datos genéticos

Los datos genéticos, al igual que las imágenes, han sido obtenidos de ADNI. Estos datos consisten en una codificación del genoma de los pacientes. Hemos encontrado datos de 79 de los pacientes de los que tenemos imágenes.

Grupo de estudio	Grupo de estudio	Número de imágenes
AD	F	15
	M	26
CN	F	18
	M	20

Tabla 3: Cantidad de pacientes con datos genéticos según el sexo y el grupo de estudio

En la tabla 3 apreciamos la cantidad de datos genéticos que tenemos según el sexo y el grupo de estudio del paciente. Podemos ver que existe un sesgo en la cantidad de datos según el sexo, ya que tenemos un 41.7% de mujeres, puede ser un punto a mejorar en el estudio ya que

este porcentaje no representa la distribución real de la población ya que la incidencia de la enfermedad en Europa es mayor en mujeres (13.25 por 1000 personas-año) que en hombres (7.02 por 1000 personas-año) [3].

4.2. PROCESAMIENTO DE LAS IMÁGENES

Una práctica común en el análisis de grupos mediante técnicas de Anatomía Computacional que se ha trasladado a los sistemas de aprendizaje automático y deep-learning es el registro de las imágenes a un atlas común. En este TFG se ha utilizado el atlas MNI152_T1_2mm_brain incluido en la librería de procesado de imágenes cerebrales [FSL](#). Las imágenes se han alineado mediante un algoritmo de registro afín y otro de registro no-rígido utilizando la librería [RNiftyReg](#). Como resultado de este procesado las imágenes finales tienen tamaño el tamaño del atlas (109x91x91) y las estructuras cerebrales se encuentran alineadas en un sistema de coordenadas no-rígido común.

En la ilustración 2 podemos ver el atlas de referencia utilizado para realizar el registro lineal y no lineal. Esto hace que las regiones cerebrales, que en cada persona pueden ser algo distintas en forma y tamaño, queden en la misma posición, facilitando su análisis comparativo.

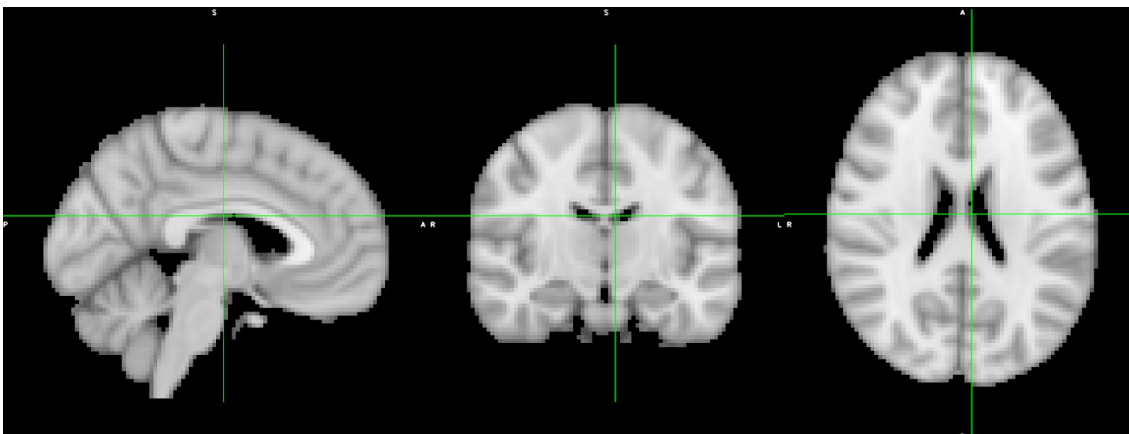


Ilustración 2: Atlas MNI152_T1_2mm_brain

4.3. DATA AUGMENTATION

Las redes neuronales son conocidas por su gran necesidad de datos. Debido a su gran capacidad para aproximar funciones complicadas, es decir, a su gran varianza, son muy propensas al overfitting. La mejor solución para este problema es aumentar la cantidad de datos de entrenamiento del modelo. Aunque esto es un problema sobre todo para datos médicos ya que son complicados de obtener, suelen estar protegidos y su clasificación para poder utilizar en modelos supervisados debe ser realizada normalmente por expertos en la materia. Ante la dificultad de aumentar el número de datos surgen dos alternativas. La primera es transfer learning, consistente en utilizar un modelo entrenado en una tarea similar a la tarea objetivo y reentrenar las capas finales del modelo para adaptarlo a nuestras necesidades. La segunda alternativa es el aumento de datos (data augmentation) que trata de realizar modificaciones en los datos disponibles tales como rotaciones, escalados, shifts o filtros gaussianos.

Para sortear la limitación provocada por el escaso número de imágenes disponibles, se han utilizado técnicas de data augmentation. La primera técnica utilizada es la aplicación de filtros

Gaussianos para distorsionar la imagen con el objetivo de imitar posibles variaciones en la resolución de las distintas imágenes debidos a cambios en los parámetros de adquisición de los diferentes scanners. Se han utilizado filtros de tamaños 3x3x3, 5x5x5 y 7x7x7 con parámetros de desviación estándar 0.7, 0.7 y 0.6 respectivamente. La segunda técnica utilizada es la traslación de las imágenes en cada una de las dimensiones en ± 1 vóxel. Esta técnica pretende imitar las posibles variaciones de las imágenes debidas a pequeños errores en el proceso de registro. La última de las técnicas consiste en voltear horizontalmente las imágenes, ya que el cerebro se puede considerar una estructura casi simétrica.

Mediante estas técnicas hemos conseguido aumentar 20 veces el número de imágenes del que disponemos ((1 imagen original + 3 imágenes con filtro Gaussiano + 6 imágenes trasladadas) x 2 imágenes volteadas). Además, para el entrenamiento de la red neuronal se han utilizado imágenes de otros puntos temporales distintos a los estudiados en el trabajo original [1]. De esta forma, el conjunto de datos se ha incrementado con 99 imágenes adicionales, de las cuales 37 son de AD y 62 de CN.

Además, para abordar el problema de la diferencia en número de imágenes de cada grupo de estudio (153 imágenes de AD vs 219 imágenes de CN), se han duplicado de forma aleatoria algunas imágenes del grupo con menor número de muestras.

Este aumento de los datos ha sido implementado durante el entrenamiento de la red neuronal. A cada imagen con la que se alimenta la red, se le aplica aleatoriamente una de las siguientes transformaciones con probabilidades:

- 10% Imagen original
- 60% Imagen trasladada (10% para cada traslación distinta)
- 30% Imagen con filtro Gaussiano (10% para cada filtro)

Finalmente, con probabilidad 50% se voltea la imagen.

Este aumento de los datos se realiza de forma online para que no sea necesario mantener el dataset completo de todas las transformaciones en la memoria principal. Esto sería impracticable debido a su gran tamaño ya que el dataset ocupa 2.5GB, siendo necesarios 50GB de RAM para almacenar el dataset aumentado completo. Con este tipo de aumento tampoco es necesario guardar el dataset transformado en disco. Las transformaciones se han implementado con operaciones de TensorFlow para ejecutarlas en paralelo para cada imagen de forma que pudiéramos aumentar la velocidad del procesado cada imagen.

4.4. CNN

Para la arquitectura de la red neuronal nos hemos inspirado en la utilizada en el estudio original, aunque su reproducción exacta ha sido imposible con la información disponible. En el trabajo original se utilizó una red neuronal convolucional basada en VGG16, con 5 capas convolucionales y 3 fully connected. Además se menciona la utilización de una capa global average pooling entre las capas convolucionales y las fully connected. El tamaño de los filtros convolucionales, el stride y el padding de cada capa convolucional se recopila en la Ilustración 3.

Conv layer	Filter size	Stride in (X, Y) direction	Stride in Z direction
Conv 1	11 × 11 × 11	4	4
Conv 2	5 × 5 × 5	1	1
Conv 3	3 × 3 × 3	1	1
Conv 4	3 × 3 × 3	1	1
Conv 5	3 × 3 × 3	1	1

3D, three-dimensional.

Ilustración 3: Tabla 1 del artículo original, contiene la información sobre el número de capas convolucionales, tamaño y stride de los filtros

En este TFG se ha diseñado una arquitectura utilizando las restricciones que conocemos del trabajo original y el resto de la arquitectura se ha diseñado de forma propia. El número de filtros en cada capa convolucional, el de neuronas en las capas fully connected, y el número de capas adicionales (como capas de batch normalization o max pooling) han sido seleccionadas mediante validación cruzada.

A continuación detallamos los parámetros de configuración más relevantes de la arquitectura utilizada en este TFG:

4.4.1. Elección del optimizador

Para la elección del optimizador se realizaron experimentos con RMSProp y Adam. Finalmente, se seleccionó Adam, ya que se comprobó que conseguía converger más rápido en este problema.

4.4.2. Arquitectura del modelo

Tras varias pruebas, se diseñó un modelo que utiliza batch normalization tras cada capa convolucional, una capa de max pooling en la última capa convolucional y regularización de cada capa convolucional mediante dropout.

4.4.3. Selección del learning rate

Para seleccionar el learning rate (LR) se utilizó la técnica de cyclical learning rates [13], consistente en entrenar el modelo con un learning rate que va aumentando a cada iteración. El LR seleccionado es aquel que muestra una caída en el loss más pronunciada, como puede apreciarse en las ilustraciones 4 y 5. Las gráficas muestran que este valor es un punto alrededor de 10^{-6} . Finalmente se ha seleccionado un LR de $3 * 10^{-6}$ ya que es el que mejor ha funcionado en los experimentos realizados.

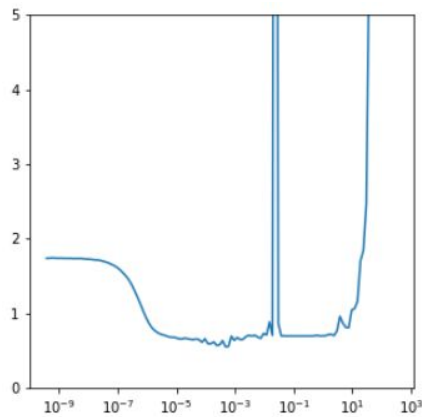


Ilustración 4: Grafica loss vs learning rate

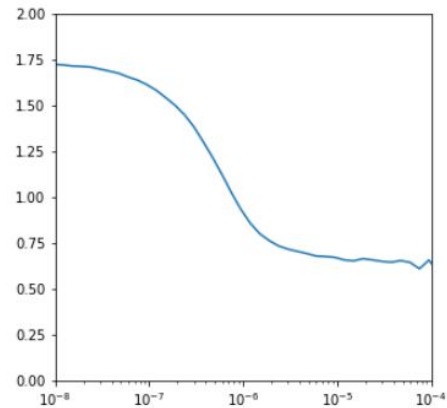


Ilustración 5: Zoom entre 10^{-8} y 10^{-4} de la grafica loss vs lr

4.4.4. Selección del número de filtros y número de neuronas

Se decidió utilizar el mismo número de filtros en todas las capas convolucionales y el mismo número de neuronas en todas las capas FC, ya que suele funcionar bien en la mayoría de los casos quedando solamente un hiper-parámetro que ajustar. Se ha realizado la selección de este hiper-parámetro mediante k-fold cross validation. Se han generado 3 folds con un 20% de los datos para el conjunto de validación, seleccionando el modelo con mejor loss de entre 16 posibles.

4.4.5. Selección de los parámetros de regularización

Finalmente, se han seleccionado los parámetros de regularización para reducir el overfitting. Se ha utilizado como regularización para las capas convolucionales batch normalization, para las capas FC, se utiliza regularización L2 para la activación y dropout.

Se mantienen los parámetros por defecto de batch normalization. Para la selección de los parámetros de dropout y de regularización de capas FC, se utiliza nuevamente k-fold cross validation.

4.4.6. Modelo final

- Capa convolucional 3D con tamaño de filtro de 11 x 11 x 11, 64 filtros y stride = 4
- Capa de batch normalization
- Capa de activación ReLU
- Capa convolucional 3D con tamaño de filtro de 5 x 5 x 5, 128 filtros y stride = 1
- Capa de batch normalization
- Capa de activación ReLU
- Capa convolucional 3D con tamaño de filtro de 3 x 3 x 3, 256 filtros y stride = 1
- Capa de batch normalization
- Capa de activación ReLU
- Capa convolucional 3D con tamaño de filtro de 3 x 3 x 3, 512 filtros y stride = 1.
- Capa de batch normalization
- Capa de activación ReLU
- Capa de max pooling 3D
- Capa convolucional 3D con tamaño de filtro de 3 x 3 x 3, 512 filtros y stride = 1
- Capa de batch normalization
- Capa de activación ReLU
- Capa de global average pooling 3D
- Capa fully connected con 512 neuronas
- Capa dropout con dropout rate = 0.1
- Capa de activación ReLU
- Capa fully connected con 512 neuronas
- Capa dropout con dropout rate = 0.1
- Capa de activación ReLU
- Capa fully connected con 512 neuronas
- Capa dropout con dropout rate = 0.1
- Capa de activación ReLU
- Capa fully connected con 1 neurona con activación logística

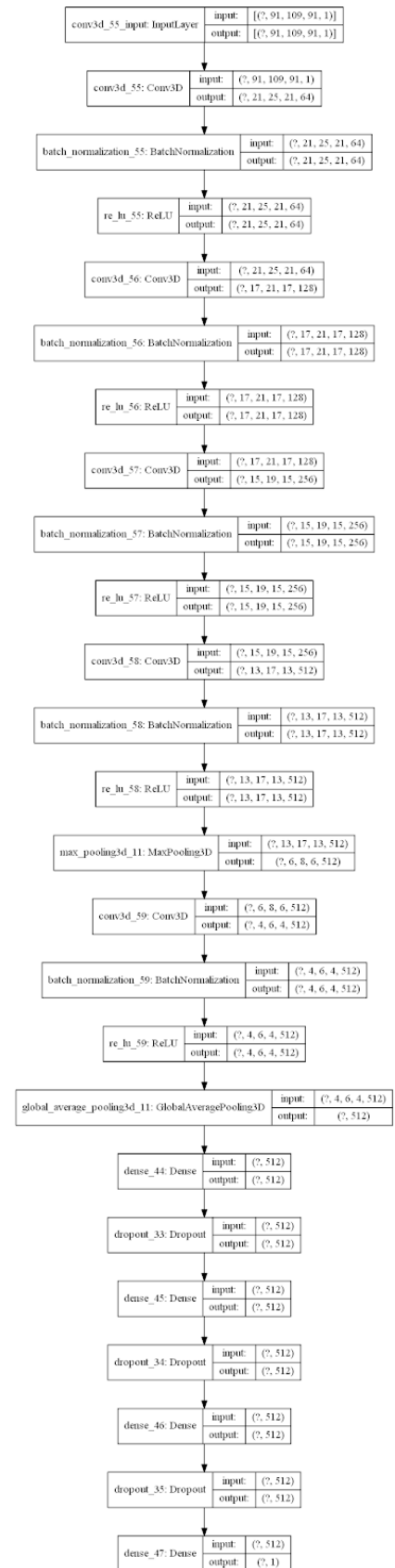


Ilustración 6: Arquitectura de la CNN

4.5. SELECCIÓN DE CARACTERÍSTICAS “PROFUNDAS” DE IMÁGENES DTI

Para seleccionar características relevantes en imágenes DTI para la clasificación de AD se ha utilizado la técnica denominada Prediction Difference Analysis [14]. Esta técnica consiste en estimar la importancia de los píxeles de entrada calculando el efecto de eliminar información de la imagen en la probabilidad de predicción. Así se puede visualizar la respuesta de la CNN a una entrada específica.

Para ello, se ha utilizado una ventana móvil de tamaño 3x3x3 para ocultar la información de cada vóxel de la imagen. Cada vóxel de la ventana es sustituido por un valor aleatorio de una distribución normal multivariable con media 0 y desviación estándar 1. La nueva imagen es utilizada como entrada para el modelo de CNN previamente entrenado para obtener la probabilidad p_1 de predicción de AD. Siendo p_0 la probabilidad de AD de la imagen original, se calcula la importancia relativa mediante la fórmula:

$$d = \log \left(\frac{\frac{p_0}{1-p_0}}{\frac{p_1}{1-p_1}} \right) \quad (1)$$

En el trabajo de referencia [1] se generó una matriz de relevancia del mismo tamaño que la imagen original. Esto requiere realizar una predicción para cada voxel de cada imagen, lo que costaría aproximadamente 16 días de cómputo, lo que se consideró excesivo para la obtención de resultados en este TFG dada la extensión de los experimentos a desarrollar. Por ello, se ha decidió realizar esta técnica con un stride de 3, es decir, moviendo la ventana de 3 en 3 vóxeles, en vez de uno en uno. Con ello, se redujo el número de operaciones a realizar en 3 por cada dimensión, por lo que se consiguió una obtención de resultados 27 veces más rápido: $\lfloor 91/3 \rfloor * \lfloor 109/3 \rfloor * \lfloor 91/3 \rfloor = 33.400$ operaciones en vez de 902.629, permitiendo realizar el cálculo en unas horas, a costa de obtener una matriz de relevancia con menor resolución aunque suficiente para la selección de características.

En la matriz de relevancia, un valor positivo indica que el vóxel contribuye positivamente a la clasificación de AD, mientras que un valor negativo indica que el vóxel contribuye negativamente a esta. Se han seleccionado como regiones relevantes las que muestran una variación en la predicción cuyo valor esté en el percentil décimo ($10^{\text{º}}$) superior.

4.6. FPCA EN IMÁGENES

Functional Principal Component Analysis es un método estadístico para la investigación de los modos dominantes de datos funcionales. Los datos funcionales son datos que representan información sobre curvas superficies o cualquier otra magnitud que varíe sobre un continuo. Generalmente cada elemento es considerado una función. El continuo sobre el que están definidas estas funciones suele ser el tiempo pero también puede ser localización espacial, longitud de onda, probabilidad, etc. Así, pues, FPCA puede considerarse una adaptación de Principal Component Analysis (PCA) para conseguir una reducción de la dimensionalidad en datos funcionales.

En este TFG hemos utilizado FPCA para la reducción de la dimensionalidad de las regiones de interés de las imágenes con las que vamos a realizar el experimento causal. Para ello hemos

utilizado la librería FPCA3D de R desarrollada por los autores del estudio original. Los datos se han reducido manteniendo una variabilidad del 95%.

4.7. PROCESAMIENTO DE DATOS GENÉTICOS

Para los experimentos involucrando los datos genéticos de ADNI, nos hemos quedado solamente con los datos correspondientes a los pacientes de los que tenemos imágenes DTI. Para la utilización de los datos genéticos, hemos realizado en primer lugar un control de calidad con la herramienta [Plink](#), la cual hemos utilizado mediante la librería de R [plinkQC](#). Al igual que en el estudio original, se han realizado los siguientes pasos:

A nivel de individuo:

- Se han eliminado los individuos con información de género discordante.
- Se han eliminado los individuos con falta de información mayor del 10%.
- Se han eliminado los individuos con ratio de heterocigosidad mayor de tres desviaciones estándar.
- Se han eliminado los individuos con identity by descent (IBD) > 0.185.

A nivel de SNP:

- Se han eliminado los SNP con falta de información mayor del 10%.
- Se han eliminado los SNP con P-valor para el test de equilibrio de Hardy-Weinberg (HWE) menor de 1E-6.
- Se han eliminado los SNP sin polimorfismo.

A continuación se ha utilizado la herramienta de control de calidad preimputación de McCarthy Groups con los datos de referencia 1000G [15]. El siguiente paso a realizar ha sido la imputación de los datos faltantes utilizando las herramientas [SHAPEIT](#) y [IMPUTE2](#). Finalmente, hemos obtenido los datos de los genes que nos interesan de los datos del genoma completo. En el estudio original [1] indican que han utilizado 61 genes. No obstante, sólo hemos encontrado 42, de los cuales hemos utilizado solo los que tienen al menos 5 SNP tras el control de calidad, por lo que finalmente hemos utilizado 40 genes.

4.8. REDES GENERATIVAS ADVERSARIAS CONDICIONALES Y TEST DE DOS MUESTRAS MEDIANTE CLASIFICACIÓN PARA DESCUBRIMIENTO DE CAUSALIDAD

4.8.1. CGAN

Una Generative Adversarial Network (GAN) [20] es una arquitectura de redes neuronales utilizada para generar ejemplos de una distribución de probabilidad. Esta arquitectura consiste de un modelo generador G, encargado de capturar la distribución estadística de los datos y un modelo discriminador D, que estima la probabilidad de que una muestra provenga de los datos de entrenamiento en vez de los datos generados por G. El entrenamiento de este tipo de redes consiste en un juego minimax en el cual el generador busca engañar cada vez mejor al discriminador y el discriminador intenta distinguir cada vez mejor qué ejemplos son reales y cuales generados. Este tipo de redes tiene como objetivo optimizar la función de pérdida:

$$\min_g \max_d \mathbb{E}_{x \sim P(X)} [\log(d(x))] + \mathbb{E}_{z \sim P(Z)} [\log(1 - d(g(z)))]$$

En una GAN no se puede controlar la clase de los datos generados, ya que el input del generador es un vector Z de ruido aleatorio.

Una Conditional GAN (CGAN) es una arquitectura de redes neuronales cuyo propósito es condicionar la generación de los datos de forma que se generen ejemplos de la clase objetivo. Esto se consigue introduciendo como input del generador, además del vector de ruido Z , la etiqueta de la clase objetivo. También es necesario introducir como input al discriminador la etiqueta de las muestras. La ilustración 7 muestra un esquema de CGAN. En la imagen podemos observar cómo introducimos en el discriminador el ejemplo X o el ejemplo generado $G(z|y)$ y su etiqueta Y . También vemos como al generador le introducimos el vector de ruido Z y la etiqueta del ejemplo que queremos obtener.

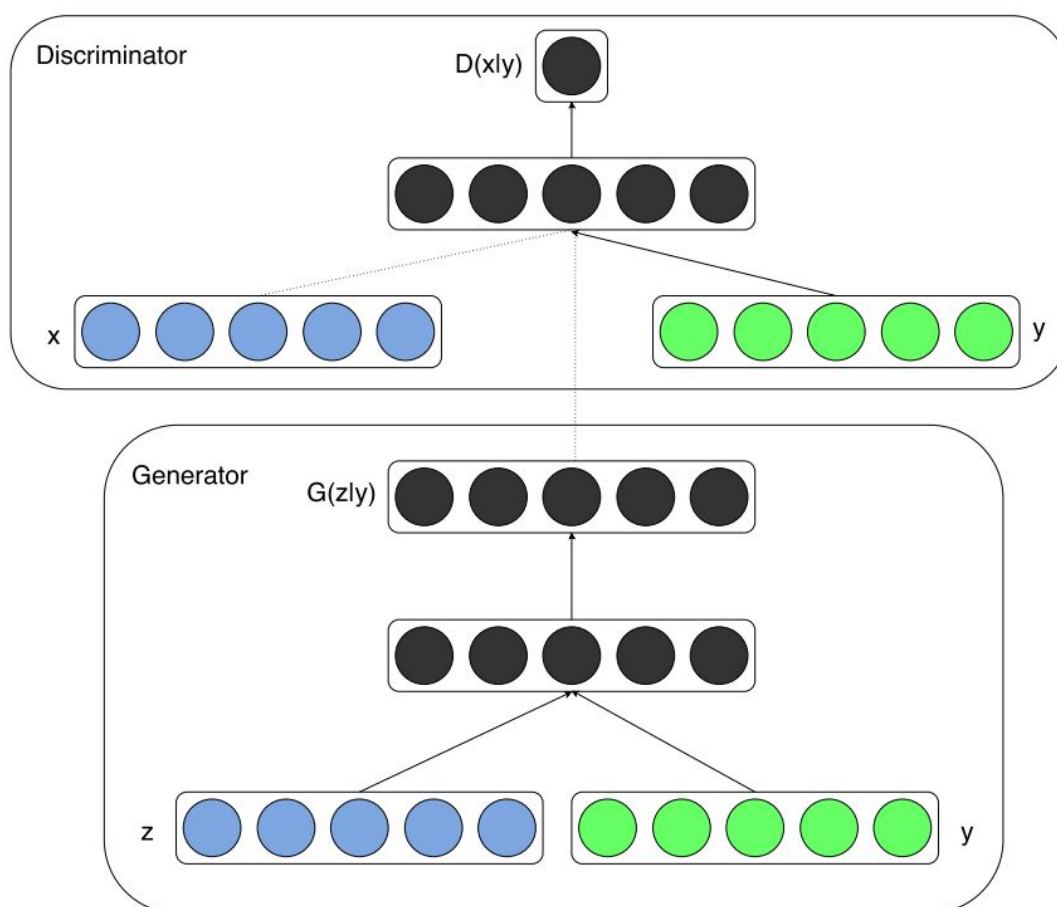


Ilustración 7: Red Generativa Adversaria Condicional

4.8.2. C2ST

Un 2 Sample Test es un test utilizado para discernir si dos muestras $S^P \sim P^n$ y $S^Q \sim Q^m$ se han obtenido de la misma distribución. Una forma de abordar este problema es mediante el uso de un clasificador binario. Este consiste en construir un dataset emparejando los n elementos en S_p con una etiqueta positiva y los m elementos en S_Q con una etiqueta negativa. Si la

hipótesis nula $P = Q$ es cierta, entonces la precisión del clasificador debería permanecer alrededor del 50%. En el caso contrario, si el clasificador obtiene una precisión mayor significa que P es distinto de Q .

Este tipo de test puede darnos una forma de obtener una evaluación de la calidad del modelo asociado a las GAN, ya que si un clasificador no es capaz de distinguir entre los ejemplos de la distribución real y los generados por la GAN significa que la GAN está generando buenos resultados. Esta medida de evaluación será la precisión del clasificador: a mayor precisión, mayor diferencia existe entre los conjuntos.

4.8.3. Descubrimiento causal

En el descubrimiento causal se estudia la estructura subyacente a un set de d variables aleatorias X_1, \dots, X_d . Asumimos que las variables aleatorias X_1, \dots, X_d comparten una estructura causal descrita por una colección de Ecuaciones Estructurales (SE). Específicamente asumimos que la variable aleatoria X_i toma valores de la forma descrita por la SE $X_i = g_i(P_a(X_i, G), N_i)$, para todo $i = 1, \dots, d$. G es un Grafo Dirigido Acíclico (DAG) con vértices asociados a cada una de las variables aleatorias X_1, \dots, X_d . $P_a(X_i, G)$ denota el set de variables aleatorias que son padres de X_i en el grafo G y N_i es una variable independiente de ruido aleatorio que sigue la distribución de probabilidad $P(N_i)$. Entonces decimos que $X_i \rightarrow X_j$ si $X_i \in P_a(X_j)$, es decir, que X_i causa X_j si X_i está entre los padres de X_j ya que un cambio en X_i **causará** un cambio en X_j , como está descrito en la i -ésima SE.

El objetivo del descubrimiento causal es inferir el grafo causal G dada una muestra de $P(X_1, \dots, X_d)$. Para simplificar, nos concentramos en el descubrimiento de las relaciones causales entre dos variables X e Y . Dada la muestra $D = \{(x_i, y_i)\}_{i=1}^n \sim P_n(X, Y)$, nuestro objetivo es concluir si “ X causa Y ” o “ Y causa X ”. Este problema se llama *descubrimiento de causa-efecto*. En el caso $X \rightarrow Y$, podemos escribir la relación causa efecto como:

$$x \sim P(X), \quad n \sim P(N), \quad y \leftarrow g(x, n)$$

Las CGAN surgen como alternativa para simular las SE debido a su gran parecido, dado que el generador de una CGAN produce $y = g(n, x)$, siendo n una variable de ruido aleatorio y x la condición de entrada del generador y la variable y es la variable sintetizada por el generador. Además las CGAN respetan la independencia entre la causa X y el ruido N ya que $n \sim P(N)$ es independiente del resto de las variables.

El procedimiento para el descubrimiento de causa efecto propuesto consiste en entrenar dos CGANs: una con un generador g_y que genera el dataset $D_{X \rightarrow Y} = \{(x_i, g_y(x_i, z_i))\}_{i=1}^n$ y un generador g_x que genera el dataset $D_{Y \rightarrow X} = \{(g_x(y_i, z_i), y_i)\}_{i=1}^n$. Finalmente, escogemos preferiblemente la dirección causal $X \rightarrow Y$ si el resultado del C2ST entre la muestra real D y $D_{X \rightarrow Y}$ es menor que el resultado del C2ST entre D y $D_{Y \rightarrow X}$. Utilizamos el principio de la navaja de Occam, la dirección más simple es la verdadera dirección causal.

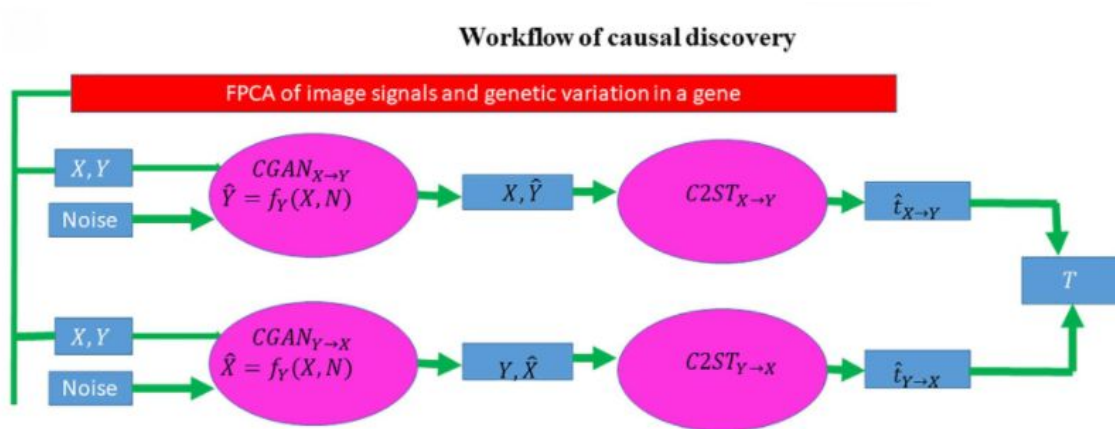


Ilustración 8: Diagrama del procedimiento de descubrimiento causal realizado

En la ilustración 8 vemos un diagrama en el que se representa el procedimiento realizado que se resume en los siguientes pasos:

1. Utilizamos una CGAN $X \rightarrow Y$ para generar el dataset $D_{X \rightarrow Y} = \{(x_i, g_y(x_i, z_i))\}_{i=1}^n$
2. Utilizamos una CGAN $Y \rightarrow X$ para generar el dataset $D_{Y \rightarrow X} = \{(g_x(y_i, z_i), y_i)\}_{i=1}^n$
3. Dividimos los datasets en conjuntos de entrenamiento y test.
4. Clasificamos dos datasets $D_u = D_y = \{y_i, i = 1, \dots, n\}$ versus $D_v = D_{X \rightarrow Y} = \{\hat{y}_i, i = 1, \dots, n\}$ and calculate the two-sample statistic $\hat{t}_{X \rightarrow Y}$.
5. Clasificamos dos datasets $D_u = D_x = \{x, i = 1, \dots, n\}$ versus $D_v = D_{Y \rightarrow X} = \{\hat{x}_i, i = 1, \dots, n\}$ and calculate the two-sample statistic $\hat{t}_{Y \rightarrow X}$.
6. Calculamos el test estadístico $T = \hat{t}_{X \rightarrow Y} - \hat{t}_{Y \rightarrow X}$. Bajo la hipótesis nula de no relación causal o test inconcluso, T se distribuye asintóticamente como $N(0, \sigma)$, donde $\sigma^2 = \frac{0.5}{n_{test}} - 2cov(\hat{t}_{X \rightarrow Y}, \hat{t}_{Y \rightarrow X})$ y n_{test} es el número de sujetos en el conjunto de test.

4.8.4. Arquitectura del experimento causal

Para los experimentos causales se ha utilizado una arquitectura basada en la utilizada en Lopez-Paz, David, and Maxime Oquab. *Revisiting classifier two-sample tests. arXiv preprint arXiv:1610.06545v4 (2018)*. [2], la cual se compone de un generador y un discriminador con 2 capas fully connected seguidas cada una de una capa de dropout y con activación ReLU. Hemos reemplazado las capas de dropout por batch normalization y la activación ReLU por leaky ReLU ya que funciona mejor en redes generativas. Para solucionar el problema de que el discriminador se entrene demasiado rápido, produciendo que el feedback que recibe el generador no sea útil, se actualizan los pesos del discriminador una vez cada dos épocas, por lo que el discriminador entrena la mitad que el generador, equilibrando su aprendizaje y permitiendo que el generador reciba un feedback útil. El output del discriminador tiene una activación lineal en vez de logística ya que da un feedback más útil al generador al no estar limitado entre 0 y 1. La CGAN se ha entrenado durante 1000 épocas en el mismo ordenador que la CNN.

En la ilustración 9 podemos ver los diagramas de la arquitectura del generador y del discriminador utilizados en uno de los experimentos. En este experimento generamos la variable \hat{X} con tamaño 47, dando como entrada al generador la variable Y , que actuará como condición, de tamaño 50 y un vector de ruido aleatorio N con tamaño 50. El discriminador recibirá como entrada X (del dataset de entrenamiento) o \hat{X} (de los ejemplos generados) e Y obteniendo como resultado la predicción del discriminador.

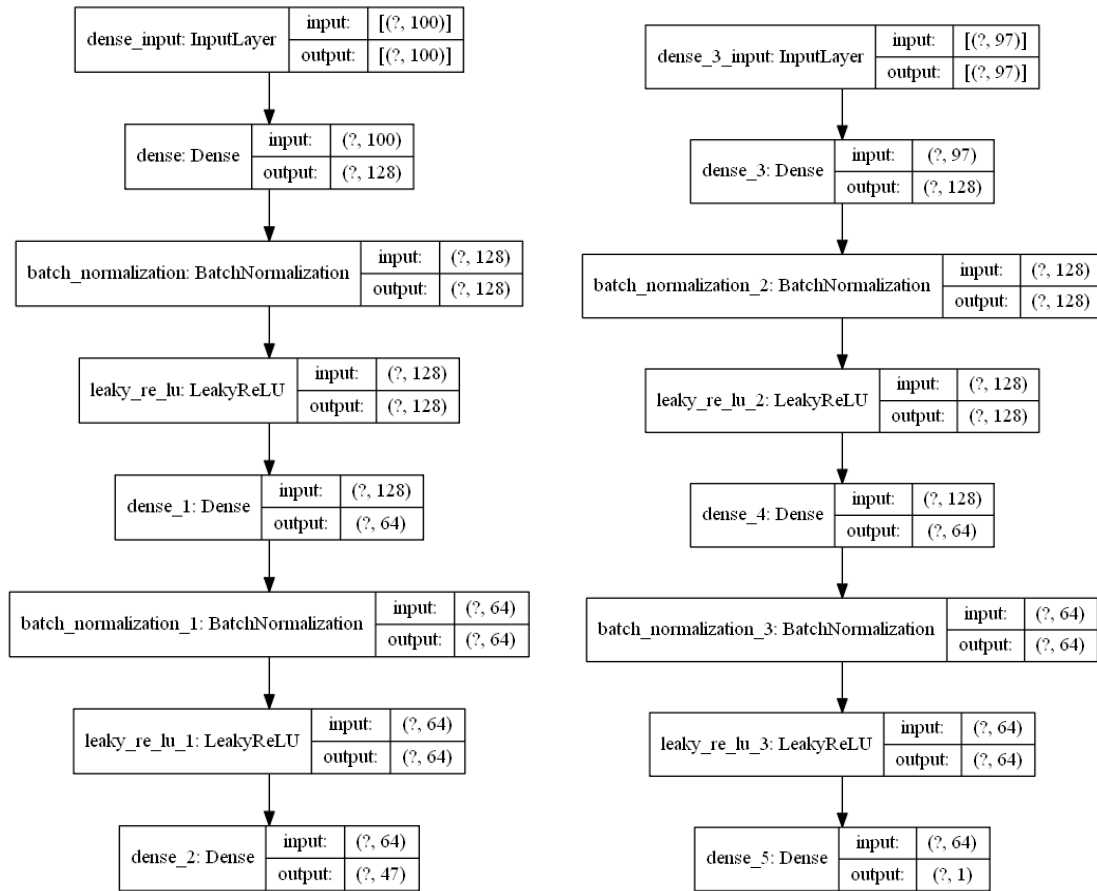


Ilustración 9: Arquitectura del generador y discriminador respectivamente

El siguiente paso en el experimento causal es la evaluación de la generación de las dos direcciones causales a evaluar. Esto lo hacemos mediante un clasificador basado en un modelo KNN con $k = \sqrt{n_{train}}$ siendo n_{train} el número de muestras en el conjunto de entrenamiento. Hemos utilizado un 70% de las muestras disponibles para entrenamiento y 30% para test del KNN.

Obtenemos la evaluación final del experimento $T = t_{X \rightarrow Y} - t_{Y \rightarrow X}$ y realizamos el contraste de hipótesis para comprobar si podemos rechazar la hipótesis nula de no relevancia de la causalidad entre X e Y .

4.9. RESULTADOS

4.9.1. Resultados CNN

El entrenamiento de la red neuronal ha sido realizado en un ordenador con 32 GB de memoria RAM, un procesador Ryzen 7 3700X con 16 hilos de procesamiento y una tarjeta gráfica NVIDIA Geforce 2700 SUPER con 8GB de VRAM. El framework utilizado para el desarrollo de la red neuronal ha sido Tensorflow 2.1 con Python 3.6.

Se ha entrenado la red neuronal un total de 400 épocas utilizando un tamaño de batch de 64. El tiempo dedicado a este entrenamiento ha sido de 1 hora y 22 minutos. La red neuronal final ha obtenido una precisión de 79.72%, una sensibilidad de 80% y una especificidad 79.41% en test. Debido a la baja cantidad de datos de entrenamiento disponibles se ha intentado regularizar todo lo posible el modelo utilizando batch normalization y dropout.

4.9.2. Resultados de la selección de características profundas

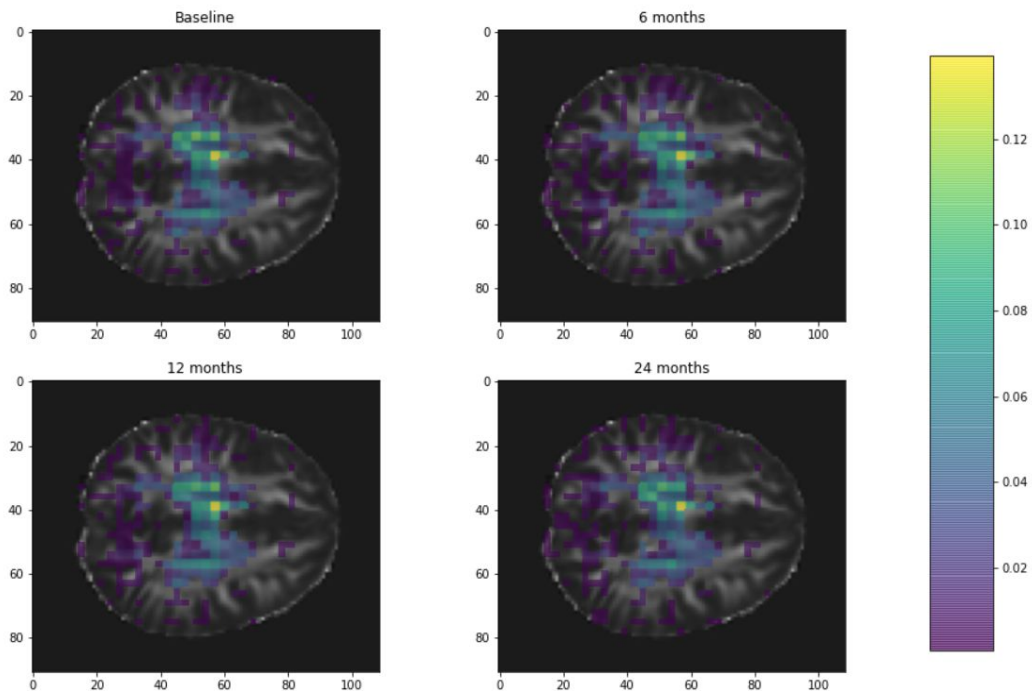


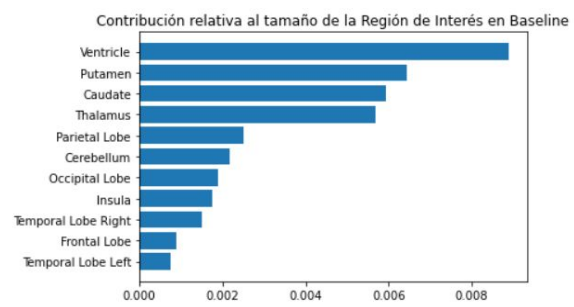
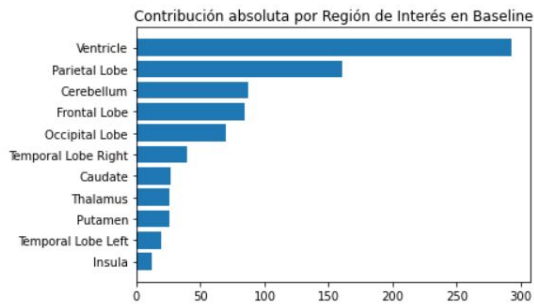
Ilustración 10: Imagen del mapa de relevancia de los pacientes en los distintos puntos temporales.

Presentamos en la ilustración 10 una imagen con un corte axial de un cerebro donde se muestra el mapa de relevancia obtenido de cada uno de los puntos temporales del estudio. Se puede observar que no hay diferencias significativas en esta sección dependiendo del punto temporal. La zona más relevante es la central, donde se encuentra el ventrículo.

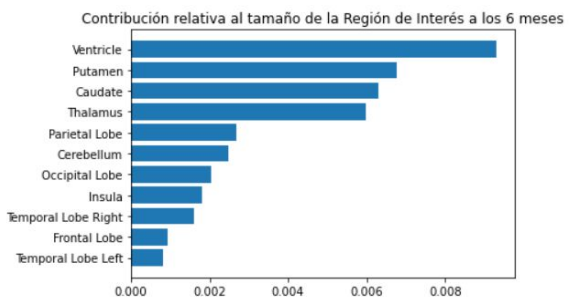
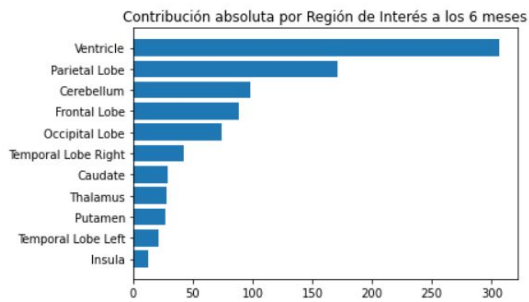
donde podemos ver un punto mucho más relevante que los demás en todas las imágenes. Seguramente este punto tan relevante sea debido al sobreentrenamiento del modelo.

Los resultados de relevancia de las distintas regiones de interés según su grupo temporal son los siguientes:

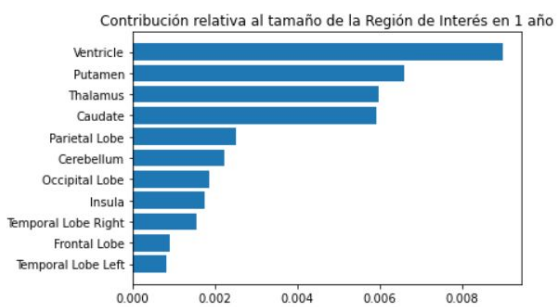
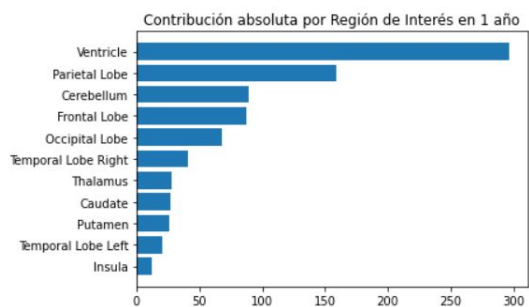
Baseline



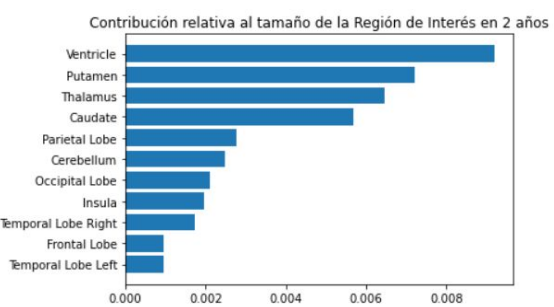
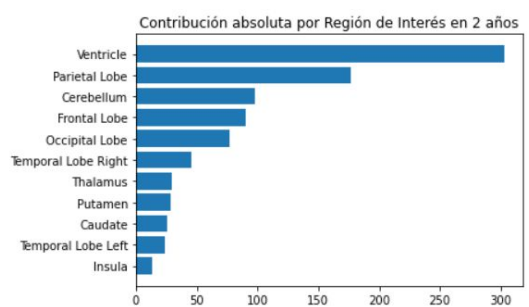
6 meses



1 año



2 años



Las métricas absolutas de las gráficas anteriores han sido obtenidas realizando un sumatorio del valor de relevancia de cada vóxel de cada región de interés, las medidas relativas se han obtenido dividiendo la métrica absoluta entre el número de vóxeles de la región de interés.

Podemos observar que, como ya podíamos intuir en la ilustración 10, la ROI con mayor relevancia, tanto absoluta como relativa, es el ventrículo. Los siguientes puestos cambian según si miramos a los gráficos absolutos o relativos, en los gráficos absolutos, los primeros puestos en relevancia los obtienen zonas grandes del cerebro como el lóbulo parietal, el frontal que al ponderar según el tamaño caen en relevancia. En los gráficos relativos podemos

ver en los primeros puestos al putamen, el tálamo y el núcleo caudado, esto son regiones más pequeñas pero con mayor densidad de relevancia.

En conclusión, las regiones con las que hemos obtenido mayor relevancia para la predicción de la enfermedad del Alzheimer son los ventrículos, las regiones cercanas a estos como el putamen, el tálamo y el núcleo caudado han obtenido también una gran relevancia relativa.

Podemos encontrar en la literatura trabajos que mencionan al ventrículo como una región que suele verse muy afectada por la enfermedad de Alzheimer [4]. El agrandamiento del ventrículo es un biomarcador estructural útil en el diagnóstico del Alzheimer [5].

4.9.3. Resultados causalidad ROI vs AD

ROI	P-valor
Ínsula	0
Putamen	0
Núcleo caudado	6.86e-12
Tálamo	2.43e-11
Ventrículo	7.6e-7
Lóbulo parietal	2.66e-6
Cerebelo	9.13e-6
Lóbulo temporal izquierdo	9.91e-5
Lóbulo occipital	2.34e-4
Lóbulo frontal	2.01e-3
Lóbulo temporal derecho	2.019e-2

Tabla 4: Resultados del experimento causal ROI vs AD

En la tabla 4 vemos los resultados obtenidos en el experimento causal, podemos apreciar que se ha obtenido un P-valor inferior a 0.05 en todas las regiones, estos valores se han obtenido con la dirección causal ROI → AD. Podemos observar que en general, las ROI con menor p-valor coinciden con las que mayor relevancia relativa han obtenido en el experimento de selección de características profundas.

4.9.4. Resultados causalidad ROI vs Gen

Para la realización del experimento de descubrimiento causal de ROI vs Gen, hemos utilizado los datos de las ROI con dimensionalidad reducida mediante FPCA y la información de los genes reducida mediante PCA.

Para reproducir los resultados obtenidos en el estudio original y debido al elevado coste computacional de las pruebas, sólomente hemos realizado experimentos causales de ROI vs genes con las ROI utilizadas con el estudio original: el lóbulo temporal izquierdo y el lóbulo temporal derecho. Para la información genética hemos utilizado los genes mencionados en el material suplementario del estudio original. Para este experimento hemos utilizado la misma arquitectura para las CGAN y C2ST que en el estudio anterior.

En la tabla 5 se incluyen los resultados de los experimentos en los que se ha obtenido un p-valor inferior a 0.05 para alguna de las ROI. Se ha encontrado causalidad en 14 de los 40 genes analizados. La dirección causal resultante en todos los casos ha sido $Gen \rightarrow ROI$.

De entre los genes con los que hemos obtenido relevancia estadística destacamos el **SCYL1**, relacionado con enfermedades neurodegenerativas [6] y con atrofia del cerebelo [7] y el **CD33** [8], relacionado con el Alzheimer. Todos los genes obtenidos fueron identificados como relevantes en el estudio original.

Gen	P-Valor Lóbulo Temporal Izquierdo	P-Valor Lóbulo Temporal Derecho
CARS	0	4.247820e-01
RP11.394D2.1	3.057332e-12	1.861066e-01
FOLR2	5.108880e-04	1.220506e-08
RP11.749H17.2	3.126150e-05	4.250513e-01
NBR1	3.803908e-01	5.687734e-05
ZBTB1	3.200175e-01	3.803190e-03
YTHDC2	4.654660e-03	1.639215e-01
FRMD6	1.337385e-02	4.410689e-02
SCYL1	2.501113e-01	1.388172e-02
SPATA21	3.277120e-01	2.826043e-02
RP11.953B20.2	2.880412e-01	2.897384e-02
LIPE-AS1	3.011426e-02	1.335094e-01
LINC00926	6.133416e-02	4.559638e-02
CD33	9.329396e-02	4.837104e-02

Tabla 5: Resultados del experimento con p-valor < 0.05 causal ROI vs Genes

5. CONCLUSIONES

CONCLUSIONES

El objetivo principal de este TFG ha consistido en desarrollar un sistema de interpretación del aprendizaje profundo aplicado al diagnóstico de la enfermedad de Alzheimer y a la búsqueda de relación causal entre AD, ROI y Genética.

Durante la realización del TFG ha sido necesario el estudio de una gran variedad de campos. Para la realización de la CNN ha sido necesario el estudio de esta arquitectura y su implementación mediante Tensorflow, a continuación se realizó una investigación y aprendizaje sobre el aumento de datos y cómo implementarlo en la CNN de forma online. Para terminar con los conocimientos sobre deep learning se estudió e investigó sobre GANs y sus diferentes implementaciones como CGAN, Cycle-GAN o Pix2Pix y sobre técnicas para mejorar su rendimiento. Además también fue necesario el estudio de campos más relacionados con el campo de la bioinformática como el procesamiento de imágenes DTI o el procesamiento de datos genéticos, lo que implicó aprender a utilizar una gran variedad de herramientas. Finalmente fue necesario también estudiar sobre inferencia causal de forma que fuera posible entender el estudio causal realizado en el estudio original y poder reproducirlo.

Los resultados obtenidos en el experimento de interpretabilidad de la CNN indican que el ventrículo es un biomarcador importante para el diagnóstico de AD, aunque para poder asegurar que los resultados no son debidos al overfitting, sería necesario realizar el estudio con una mayor cantidad de datos. En el experimento causal ROI vs AD nuestros resultados indican que podría existir una dirección causal $ROI \rightarrow AD$ en todas las ROI analizadas con un p-valor inferior a 0.05 al igual que sucedía en el estudio original. En el experimento causal ROI vs Genes hemos obtenido como resultado la relevancia estadística de 14 genes con dirección causal $ROI \leftarrow Genes$, destacando los genes CD33 que está relacionado con AD y el SCYL1, relacionado con la atrofia del cerebelo y enfermedades neurodegenerativas.

El avance en el campo de la inferencia causal podría beneficiar en gran medida a las aplicaciones médicas ya que una de las grandes dificultades en este campo es determinar la causalidad entre dos eventos.

El código desarrollado en el TFG puede consultarse en <https://github.com/JArmunia/TFG> .

DIAGRAMA DE GANTT

En la ilustración 10 se muestra un diagrama de Gantt con las tareas desarrolladas en este TFG.

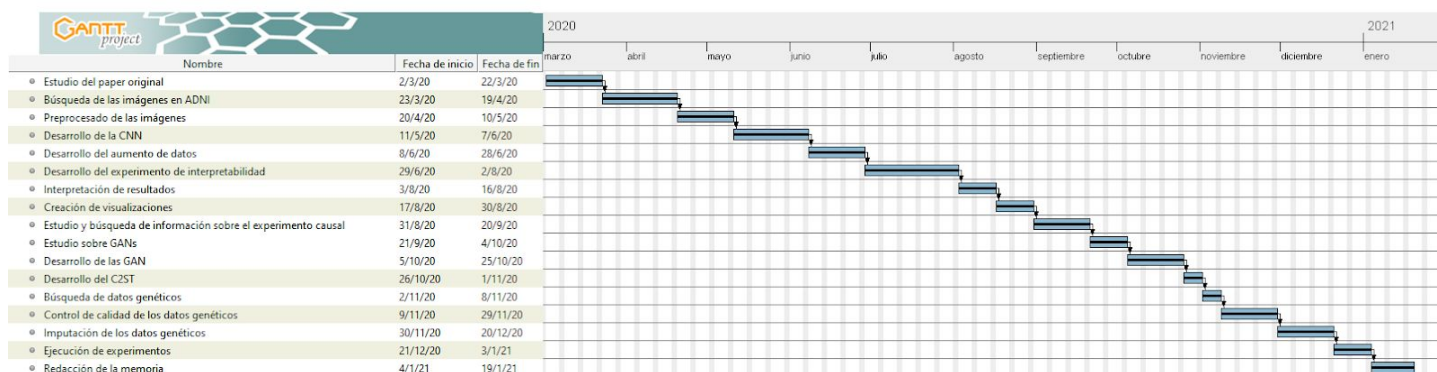


Ilustración 10: Diagrama de Gantt del TFG.

LÍNEAS FUTURAS DE TRABAJO

Como líneas futuras de trabajo, sería de especial interés poder aumentar la cantidad de datos utilizados para obtener información más generalizable, dado que con una muestra pequeña, los resultados obtenidos son difícilmente transferibles a la población general. Con el fin de mejorar la calidad de los resultados sería beneficioso realizar un ajuste preciso de los modelos para cada problema en específico.

La utilización de un clasificador para evaluar GANs mediante C2ST es una gran forma de evaluar el rendimiento de los modelos generadores, pero sería necesario descubrir una forma de distinguir si el clasificador del C2ST devuelve como salida una predicción baja debido a el buen funcionamiento del generador o al mal funcionamiento del clasificador.

Sería interesante extender la búsqueda de causalidad a otras combinaciones de biomarcadores de forma que se pudiera generar un grafo causal con aplicaciones médicas o incluso extender este sistema a aplicaciones no relacionadas con la medicina. Adicionalmente se podrían comparar los resultados obtenidos en DTI a otras técnicas de imagen como MRI o PET.

El campo de la inferencia causal es prometedor pero actualmente es un campo nuevo que en ocasiones incluso resulta polémico. Los avances en este campo podrían dar lugar a nuevas formas de hallar causalidad en aplicaciones médicas.

6. REFERENCIAS

- [1] Liu Yuanyuan, Li Zhouxuan, Ge Qiyang, Lin Nan, Xiong Momiao. Deep Feature Selection and Causal Analysis of Alzheimer's Disease. *Frontiers in Neuroscience*, página 1198, (2019).
- [2] Lopez-Paz, David, and Maxime Oquab. Revisiting classifier two-sample tests. *arXiv preprint arXiv:1610.06545v4*, (2018).
- [3] H. Niu et al., Prevalencia e incidencia de la enfermedad de Alzheimer en Europa: metaanálisis, *Elsevier revista de Neurología*, Vol 32, Num 8, páginas 523-532, Octubre (2017).
- [4] Ferrarini, L., Palm, W. M., Olofsen, H., van Buchem, M. A., Reiber, J. H., and Admiraal-Behloul, F. Shape differences of the brain ventricles in Alzheimer's disease. *Neuroimage* 32, 1060–1069, (2006).
- [5] Anandh, K. R., Sujatha, C. M., and Ramakrishnan, S. Segmentation of ventricles in Alzheimer mr images using anisotropic diffusion filtering and level set method. *Biomed. Sci. Instrum.* 50, 307–313, (2014).
- [6] Schmidt, W. M., Kraus, C., Höger, H., Hochmeister, S., Oberndorfer, F., Branka, M., et al. Mutation in the *Scyl1* gene encoding amino-terminal kinase-like protein causes a recessive form of spinocerebellar neurodegeneration. *EMBO Rep.* 8, 691–697. 2007
- [7] Schmidt, W. M., Rutledge, S. L., Schüle, R., Mayerhofer, B., Züchner, S., Boltshauser, E., et al. Disruptive *SCYL1* mutations underlie a syndrome characterized by recurrent episodes of liver failure, peripheral neuropathy, cerebellar atrophy, and ataxia. *Am. J. Hum. Genet.* 97, 855–861. doi: 10.1016/j.ajhg.2015.10.011. (2015)
- [8] Heo, S. J., Kim, Y., Yun, S., Lim, S. S., Kim, J., Nam, C. M., et al. Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in annual workers' health examination data. *Int. J. Environ. Res. Public Health* 16:E250. (2019).
- [9] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118. (2017).
- [10] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316, 2402–2410. (2016).
- [11] Ravizza, S., Huschto, T., Adamov, A., Böhm, L., Büsser, A., Flöther, F. F., et al. Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data. *Nat. Med.* 25, 57–59. (2019).
- [12] Spasov, S., Passamonti, L., Duggento, A., Liò, P., and Toschi, N. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *Neuroimage* 189, 276–287. (2019).
- [13] Leslie N. Smith. Cyclical Learning Rates for Training Neural Networks. *arXiv:1506.01186*. (2017).
- [14] Zintgraf, L. M., Cohen, T. S., Adel, T., and Welling, M. Visualizing deep neural network decisions: prediction difference analysis. *arXiv*. (2017).

- [15] A global reference for human genetic variation, The 1000 Genomes Project Consortium, *Nature* 526, 68-74, (01 October 2015).
- [16] Leandrou, S., Petroudi, S., Kyriacou, P. A., Reyes-Aldasoro, C. C., and Pattichis, C. S. Quantitative MRI brain studies in mild cognitive impairment and Alzheimer's disease: a methodological review. *IEEE Rev. Biomed. Eng.* 11, 97–111. (2018).
- [17] Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol.* 6, 734–746. (2007).
- [18] Liu, X., Chen, K., Wu, T., Weidman, D., Lure, F., and Li, J. Use of multimodality imaging and artificial intelligence for diagnosis and prognosis of early stages of Alzheimer's disease. *Transl. Res.* 194, 56–67. (2018a).
- [19] Liu, X., Hou, D., Lin, F., Luo, J., Xie, J., Wang, Y., et al. The role of neurovascular unit damage in the occurrence and development of Alzheimer's disease. *Rev. Neurosci.* 30, 477–484. (2018b).
- [20] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. Generative adversarial nets. *Adv. Neural Inform. Process. Syst.* 2, 2672–2680. (2014).