

# INTELIGENCIA ARTIFICIAL PARA LA SEGMENTACIÓN Y CLASIFICACIÓN ALGORITMOS DE HIPERINTENSIDADES EN SUSTANCIA BLANCA CEREBRAL A PARTIR DE RESONANCIA MAGNÉTICA

Sucre Cando Pacheco<sup>1</sup>, Kevin Cando Garcés, Kerling Cando Garcés<sup>1</sup>

<sup>1</sup>Universidad de Guayaquil, Ecuador

Sucre.candop@ug.edu.ec, kerling.candog@ug.edu.ec, scando@fiec.espol.edu.ec

**Resumen:** Las hiperintensidades de sustancia blanca se asocian a enfermedades cerebro vasculares debido a su rol en el decaimiento cognitivo y pérdida de funciones motrices en la vejez. Estas lesiones son visibles en imágenes obtenidas de secuencias FLAIR por resonancia magnética. El proceso de segmentación automática de estas lesiones, depende de las características del equipo lo que se traduce en una falta de robustez en los algoritmos de segmentación tradicionales. En este trabajo, se tomaron un conjunto de imágenes FLAIR y T1 obtenidas por equipos de características diferentes. Se proporcionó el “ground truth” o segmentación manual de esas lesiones. El método consiste en entrenar un modelo de aprendizaje profundo usando la red U-Net 3D. La red neuronal convolucional propuesta es de 28 capas y 22 millones de parámetros entrenables. El resultado del proceso de segmentación se comprobó con el “ground truth” mediante cinco coeficientes de comparación. Los resultados muestran una efectividad del 81% en el aprendizaje usando el coeficiente de DICE para la comparación, sin embargo, el modelo identifica con precisión las lesiones con información en el eje z. El método propuesto resultó robusto y simple al momento de segmentar imágenes tridimensionales, sin necesidad de segmentarlas con un modelo bidimensional.

**Palabras Clave:** Segmentación manual, Resonancia magnética, Inteligencia artificial, Modelos de aprendizaje

## WEB TOOL FOR AGILE DEVELOPMENT OF EXPERT SYSTEM

**Abstract:** White matter hyperintensities are usually associated with cerebrovascular diseases due to the role they occupied in cognitive decay and loss of motor functions in old ages. These lesions are visible in images obtained from FLAIR sequences by Magnetic Resonance. The process of automatic segmentation is dependent on the characteristics of the machine, which translates into a lack of robustness in the traditional segmentation algorithms. The method developed in this work is based on a data set of FLAIR and T1 images obtained by machines with different characteristics. The "ground truth" consists in the manual segmentation which was also provided by the dataset. The method consisted in training a deep learning model using the U-Net 3D network. The convolutional neuronal network proposed had 28 layers and 22 million trainable parameters. Finally, the result of this segmentation process was verified with the ground truth by comparison of five coefficients. The results showed an effectiveness of 81% which was achieved using the DICE coefficient to make the comparison, however, it could also be observed that the model identifies with information in the z-axis with great precision. The proposed method was robust and simple when segmenting three-dimensional images, without having to segment the three-dimensional images to a two-dimensional model.

**Key words:** Manual segmentation, Magnetic resonance, Artificial inteligeny, Learning models.

## I. INTRODUCCIÓN

Las hiperintensidades de sustancia blanca cerebral (WMH por sus siglas en inglés), están asociadas al deterioro cognitivo [1]. Estas son relevantes porque pueden aparecer en adultos sanos sin demencia, lo cual puede tener efecto sobre las funciones motoras, de memoria o ejecución de funciones. Las enfermedades de vaso sanguíneo se refieren a un grupo de procesos patológicos los cuales afectan pequeñas arterias, arteriolas, vénulas y capilares del cerebro; las consecuencias de estas enfermedades son lesiones ubicadas en

estructuras subcorticales tales como: infartos lacunares, lesiones por WMH, microsangrados y hemorragias. Esto tiene gran importancia en enfermedades cerebro vasculares debido al rol ocupado en el decaimiento cognitivo y pérdida de funciones motrices en la vejez [2].

La cuantificación del volumen, ubicación y forma de la WMH es clave en los estudios clínicos para apoyar el diagnóstico, pronóstico y la monitorización del tratamiento de la demencia y otras enfermedades neurodegenerativas. La calificación visual de WMH tiene limitaciones importantes [3], por lo tanto, se requiere una segmentación más detallada. Se han desarrollado diversas técnicas automatizadas de segmentación, para proporcionar mediciones cuantitativas y reemplazar procedimientos dependientes del observador que consumen mucho tiempo.

Una revisión sobre las técnicas automatizadas de segmentación de WMH permite concluir la dificultad de compararlas [3]. Cada técnica de segmentación se evalúa de manera distinta (diferente número de sujetos, diferentes expertos, diferentes protocolos), utilizando diferentes criterios.

En 2015, se desarrolló un algoritmo que permite segmentar de manera automática las hiperintensidades de sustancia blanca cerebral [4]. Sin embargo, a la hora de probar con imágenes FLAIR (Fluid Attenuated Inversion Recovery) obtenidas por otros autores, los algoritmos carecen de robustez. La naturaleza de los datos, basados en otros protocolos de adquisición de imágenes, tomadas con diferentes equipos hizo que el algoritmo redujera significativamente el rendimiento al momento de hacer la segmentación automática. Se trabajó con un dataset de una sola máquina de resonancia magnética (Siemens, TrioTim, 3 Tesla, con bobina de 12 canales), con una muestra de 18 pacientes de Parkinson con media de edad = 57.44 años y 18 sujetos sanos pareados por edad. Los criterios de selección de este grupo fueron: no padecer ninguna enfermedad y no estar tomando medicación alguna.

El problema se debía a que el algoritmo desarrollado tomaba como punto de segmentación los valores observados por un umbral (threshold) de las frecuencias

de la imagen, el cual depende del conjunto de prueba a analizar, del protocolo de adquisición de la imagen y de las características del equipo: marca, modelo, potencia del campo magnético, tipo de bobina, etc. Esto evita segmentar de manera correcta las lesiones de imágenes FLAIR y T1 realizadas en otros equipos donde cambian las variables descritas anteriormente.

Las WMH han sido sistemáticamente relacionados con factores de riesgo vascular y deterioro funcional en las actividades cotidianas razones que en realidad su significación clínica es cada vez más importante [5]. Las lesiones en sustancia blanca se visualizan mejor con las secuencias FLAIR en Resonancia Magnética (MR) que tienen la ventaja de la supresión de líquido cefalorraquídeo, permitiendo que las lesiones presenten un alto contraste [3].

Con fines clínicos, las imágenes FLAIR se evalúan principalmente mediante inspección visual, sin embargo, esto no es suficiente para cuantificar y caracterizar el comportamiento fisiológico de diferentes enfermedades [6]. Se conoce que este tipo de lesiones en sustancia blanca se presentan tanto en sujetos sanos a partir de los 60 años y sin ninguna manifestación clínica, o en pacientes con diversos tipos de patologías [1]. En este sentido, se han desarrollado varios métodos automatizados y semiautomatizados para segmentar y cuantificar las WMHs usando diferentes técnicas, proporcionando información sobre el tamaño, forma y la ubicación principalmente. Estos trabajos están más enfocados en los procedimientos matemáticos para la segmentación ya sea automática o semi-automática de las lesiones [2,5,7-10].

Básicamente, en este trabajo se determina si un método de aprendizaje profundo es capaz de realizar el proceso de segmentación automática de lesiones en sustancia blanca cerebral en imágenes de resonancia magnética, a saber:

- Realizar un método de preprocesado de las imágenes para aumentar el conjunto de datos de entrenamiento.
- Elaborar un modelo de aprendizaje profundo basado en la red U-Net 3D
- Validar la efectividad del método por medio de diferentes coeficientes de comparación.

Se establece una nueva metodología de segmentación de lesiones por hiperintensidades de sustancia blanca cerebral, usando la arquitectura propuesta por Çiçek [11] para aprendizaje profundo y siguiendo la metodología usada por Li et. al en [12], se logró segmentar lesiones con poca área axial pero con suficiente volumen en el eje z, en un equipo con recursos computacionales moderados.

En la sección II se muestra la metodología usada en el desarrollo del trabajo, luego en la sección III se muestran

los principales resultados. En la sección IV se escribe la discusión de resultados. Finalmente, se muestran las conclusiones y las referencias usadas en el trabajo.

## II. DESARROLLO

Una red neuronal está compuesta de muchas neuronas artificiales. Las redes neuronales se inspiraron originalmente en el objetivo de modelar los sistemas neuronales biológicos, pero desde entonces se ha ramificado en diferentes direcciones y se ha convertido en una herramienta de simulación en ingeniería con el fin de lograr buenos resultados en las tareas de aprendizaje automático.

### Redes Neuronales Convolucionales

Las redes neuronales convolucionales, CNNs, o ConvNets, son similares a las redes neuronales regulares. Están formadas por neuronas con pesos que aprenden de los datos. Tienen una función de pérdida en la última capa completamente conectada. Una red neuronal regular recibe datos de entrada como un único vector y pasa a través de una serie de capas ocultas. Cada una de estas consiste en un conjunto de neuronas, cada una totalmente conectada a todas las otras neuronas en la capa anterior. En una capa, cada neurona es completamente independiente y no comparten ninguna conexión. La última capa totalmente conectada, también llamada capa de salida, contiene puntuaciones por clase en el caso del problema de clasificación de imágenes. En general, hay tres capas principales en una ConvNet simple: de convolución, de agrupamiento y la capa completamente conectada.

La principal diferencia en una CNN radica en que generalmente toma imágenes como entrada, esto permite codificar algunas propiedades en la red, reduciendo así el número de parámetros. En el caso de los datos de imágenes reales, los CNN funcionan mejor que los Perceptrones de Capas Múltiples (MLP por sus siglas en inglés). Hay diferentes motivos para esto: para alimentar una imagen a un MLP, se convierte a la matriz de entrada en un vector simple sin estructura espacial. Entonces, las CNN se crean, para elucidar los patrones en datos multidimensionales. A diferencia de las MLP, las CNN entienden el hecho de que los píxeles de la imagen que están más cerca unos de otros están más relacionados que los píxeles que están más separados:

Las CNN difieren de las MLP en los tipos de capas ocultas que se pueden incluir en el modelo. Una ConvNet organiza sus neuronas en tres dimensiones: ancho, alto y profundidad. Cada capa transforma su volumen de entrada 3D en un volumen de salida 3D de neuronas usando funciones de activación.

El agrupamiento de datos es una técnica común usada en machine learning para reducir la dimensionalidad. Si se tiene un conjunto de datos de imágenes que tengan un número de píxeles de  $28 \times 28$ , esto se representará como una matriz de 1 dimensión de 784 píxeles. Al agrupar los datos, se pierde información sobre la estructura 2D de la imagen; Sin embargo, los datos se simplifican. De tal forma, que todos los datos de entrenamiento pueden estar contenidos en una matriz de la forma  $(60,000, 784)$ , donde la primera dimensión representa el número de imágenes de entrenamiento y la segunda representa el número de píxeles en cada imagen.

### Algoritmo

La propagación hacia atrás (backpropagation) consiste en:

- Realizar de una operación de avance
- Comparar la salida del modelo con la salida deseada
- Cálculo del error
- Ejecutar de la operación de acondicionamiento de la red (feedforward) hacia atrás (backpropagation) para distribuir el error a cada uno de los pesos.
- Actualizar los pesos, para obtener un mejor modelo
- Continuar hasta que se tenga un modelo óptimo.

Para la segmentación de las lesiones en sustancia blanca cerebral a partir de imágenes cerebrales FLAIR obtenidas por resonancia magnética e imágenes estructurales potenciadas en T1, se usaron imágenes adquiridas en tres escáneres diferentes de tres centros de estudios diferentes ubicados en Holanda, Ámsterdam y Singapur. El conjunto de datos constaba de un total de 60 sujetos con imágenes de resonancia magnética FLAIR y T1 de tres escáneres diferentes junto con sus máscaras binarias, las cuales consisten en mapas binarios de datos que contienen la información de las lesiones segmentadas por manualmente.

Como se muestra en la Tabla I, existe una gran diferencia en las configuraciones de adquisición; en los tamaños de vóxel particulares de las imágenes capturadas difieren significativamente entre los tres escáneres. Para cada sujeto, una imagen 3D ponderada en T1 y una división múltiple en 2D

Tabla I. Características del dataset de imágenes

Dataset	Tamaño del Voxel (m <sup>3</sup> )	Maquina Usada	Dimensiones de la imagen FLAIR	Dimensiones de la imagen T1
Utrecht	0.96x0.95x3.00	3T Philips Achieva	240x240x48	240x240x48
Singapore	1.00x1.00x3.00	3T Siemens Trio Tim	252x232x48	256x232x48
GE3T	0.98x0.98x1.20	3T GE Signa	132x256x83	132x256x83

El estándar de referencia manual está definido en la imagen FLAIR, se generó una versión en 2D de varias secciones de la imagen T1 volviendo a muestrear la imagen 3D ponderada en T1 para que coincida con la del FLAIR. El conjunto de datos a usados consta de 60 sujetos con imágenes de resonancia magnética FLAIR

y T1 de tres escáneres diferentes. Logrando así que en la etapa de entrenamiento se usa el método de dejar un paciente afuera y entrenar con el resto.

Las librerías para el análisis y segmentación de las lesiones, se implementaron en Python 2.7 y Anaconda como package manager, se muestran en la Tabla II.

Tabla II Varias librerías usadas en la implementación del programa

Librería	Versión	Descripción
Numpy	1.15	Presentada por Walt et, al en, [5] permite hacer cálculos complejos usando estructuras numéricas de otros ambientes de programación más simples.
Tensorflow	1.9.0	Es una librería <i>open source</i> permite hacer cálculos computacionales complejos mediante diagramas de flujo de datos, usados para el aprendizaje profundo como lo muestra Abadi et. al en [13]
Tensorflow-gpu	1.9.	Parecida a Tensorflow esta se diferencia de que usa el GPU, para realizar los cálculos y el aprendizaje.
Keras[16],	2.2.2	Es un API de alto nivel para redes neuronales, escrita en Python y capaz de correr usando Tensorflow, CNTK o Threano. Introducida por Chollet [14], es un <i>framework</i> que permite simplificar las operaciones del aprendizaje profundo agregando una capa de abstracción entre el modelo y el procesamiento.
Nibabel	2.3.0	Esta librería permite hacer procesamiento de imágenes tridimensionales de resonancia magnética Brett [15]
Nilearn	0.4.2	Tal como lo muestra Estève esta librería permite hacer minado de datos en imágenes cerebrales.

### Módulos del programa

El programa desarrollado es robusto a los cambios de los parámetros de configuración que pudieran darse en los equipos de resonancia magnética. Se puede lograr probando con diferentes datasets teniendo en cuenta que se requieren una muestra más numerosa para hacer el algoritmo más robusto.

El programa se divide en tres módulos principales:

- Módulo de Preprocesamiento

- Módulo de Segmentación

- Módulo de Postprocesamiento

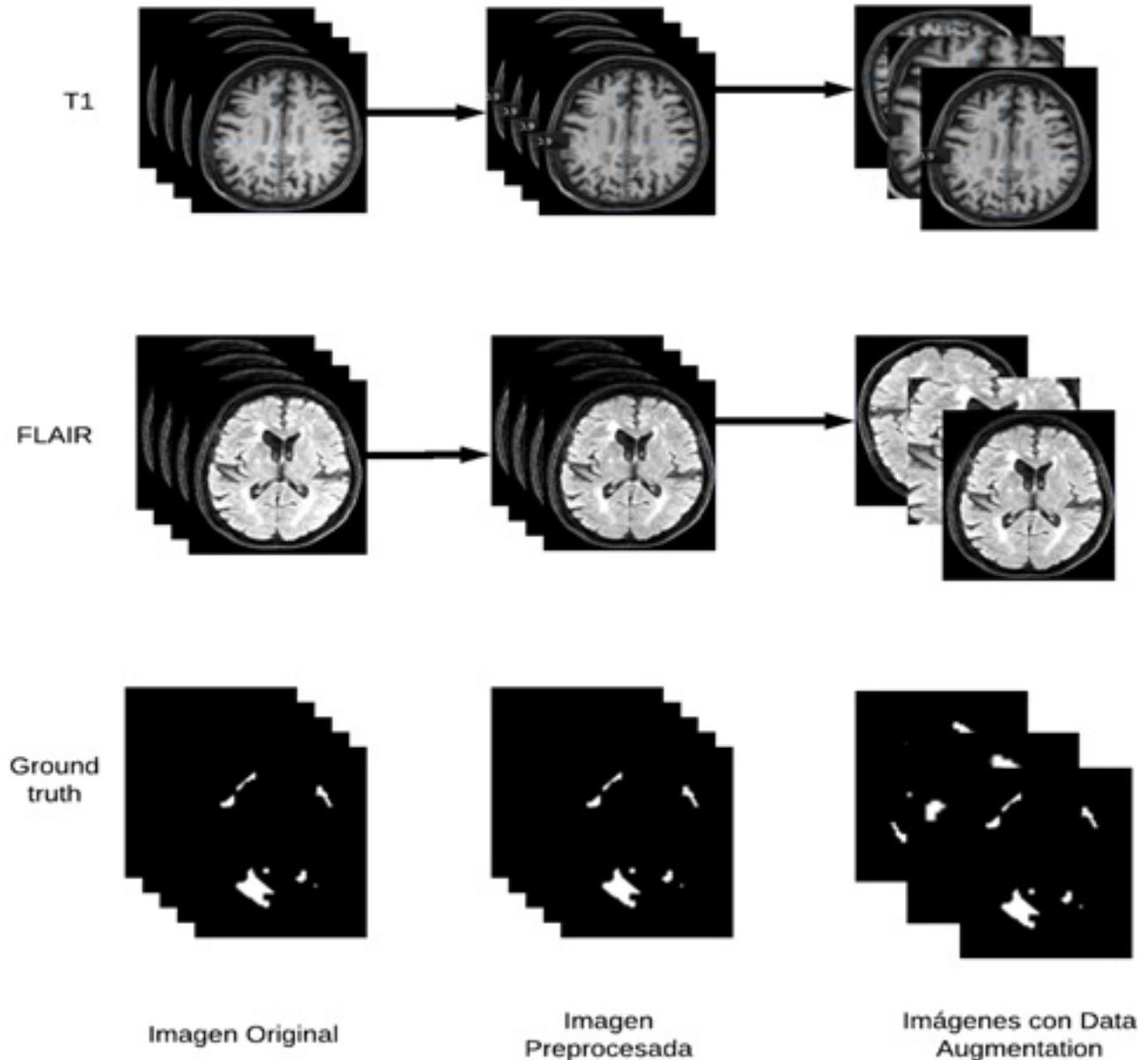
### Módulo de Preprocesado

Las imágenes requieren de un tratamiento previo antes de ser usadas para entrenar a la red neuronal. Si estas se ingresan tal como se obtienen de los escáneres de resonancia magnética se tendría una cantidad baja de detección de lesiones cerebrales. Hace pocos años se

usaban técnicas de preprocesado para la segmentación. Existían técnicas como la que muestra [4], donde se usa segmentación por umbral.

Las CNN por naturaleza son buenas para identificar y clasificar características de la imagen sin procesar. Sin

embargo, tal como lo muestra la Figura 1, se requiere que se potencie la imagen T1, la imagen FLAIR y que se haga una corrección de bias para potenciar la sustancia blanca cerebral, esto fue realizado previamente junto con el dataset de entrenamiento.



**Figura 1** Preprocesamiento de las imágenes antes de ser usadas para el entrenamiento de la red.

### Imagen original

Las imágenes originales, no poseen el mismo tamaño. Esto implica más cortes axiales, sagitales o coronales dependiendo del caso. Para agrupar las imágenes e ingresarlas a un módulo de aprendizaje autónomo, deben tener las mismas dimensiones caso contrario no se puede entrenar la red.

Se tomaron 6 grupos de imágenes por sujeto, las imágenes FLAIR y T1 originales, el ground truth proveniente del conjunto de datos. Adicionalmente las imágenes FLAIR y T1, a las cuales solo se les removió el cráneo usando la función `Bet2` de la librería `FSL`[17], tratando de alterar en lo menos posible la imagen original.

### Preprocesamiento

El suavizado se realizó pasando un kernel Gaussiano por la imagen preprocesada para reducir el posible ruido de la imagen. Una imagen con una alta resolución puede ser difícil de procesar usando Deep Learning, debido a que el rendimiento está fuertemente ligado al hardware computacional. Por esto se decidió usar imágenes de 128x128 de tal manera que el volumen tridimensional sea mucho más corto de analizar.

Se usó la función de resize de OpenCV [18], usando una interpolación cubica, reduciendo la imagen de entrada.

Luego de esto las imágenes se concatenan una tras otra con la finalidad de hacer un stream continuo de datos para el modelo de aprendizaje el cual en el caso de la imagen T1 es convertido a float32 y en el caso de la FLAIR se deja igual.

### Data augmentation

El aumento del dataset, se hizo por cada imagen dentro del conjunto de imágenes, para cada uno se hicieron tres transformaciones: rotación, shearing y zoom. Tal como se muestra en la Figura 2.

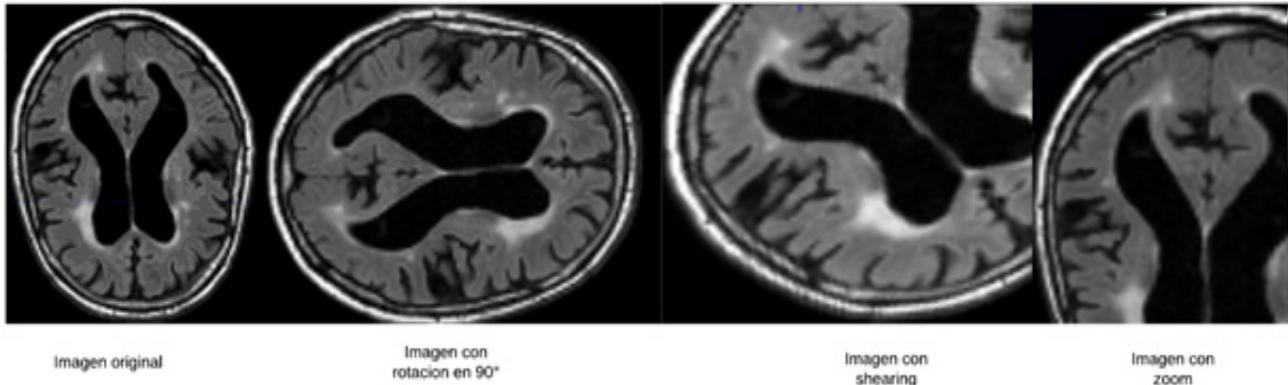


Figura 2 Imágenes después de hacer el aumento de la data

La última transformación fue hacer el zoom de la imagen ya redimensionada, logrando una imagen con más detalle en la sección a segmentar. Este proceso de preprocesamiento se repite para cada imagen que se quiera agregar al sistema, incluida la máscara del ground truth.

### Módulo de Segmentación de WMH

Teniendo las imágenes preprocesadas en un espacio estándar, se tiene que realizar la segmentación de las lesiones de sustancia blanca. Con esta finalidad se hace el siguiente proceso:

Usando el método de `train_split` proporcionado en [19], se estima el tamaño de la muestra de prueba y entrenamiento. Para esto se sigue la metodología provista en [19], en el cual se deja un paciente afuera como conjunto de prueba y se entrena con todo el dataset sobrante. Esto implica repetir el proceso de aprendizaje  $n$  veces, donde  $n$  significa el número de sujetos. Para cada iteración se toma  $n-1$  sujetos para aprendizaje y 1 sujeto para prueba. Se usó una configuración de 99% del dataset para aprendizaje y 1% para pruebas siendo este proceso rotativo.

Para la segmentación automática, se usó una red neuronal convolucional, de tal manera que se pueda extraer la mayor cantidad de información de las imágenes FLAIR y T1. El conjunto de entrada se encuentra concatenado haciendo un arreglo de imágenes tridimensionales, el

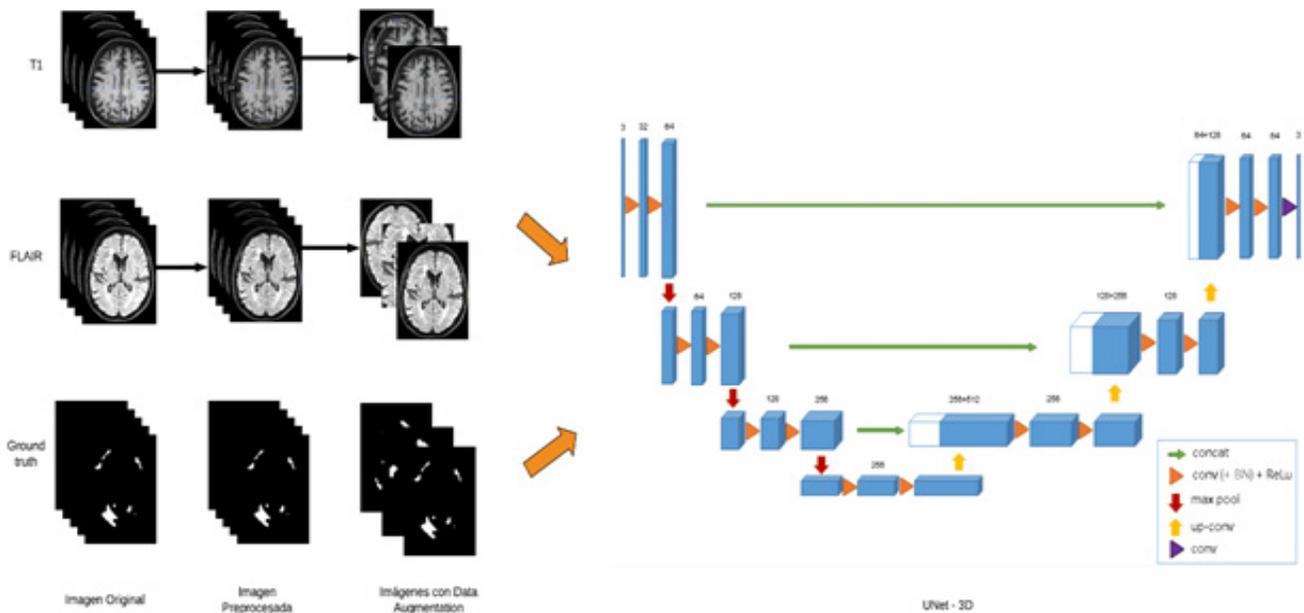
cual incluye las imágenes sin aumentar y las imágenes aumentadas; cada una con su respectiva máscara de segmentación.

### III.RESULTADOS

Con la finalidad de tener resultados más robustos al momento de encontrar y clasificar lesiones, se hizo un shuffle del dataset de entrada, de tal manera que las imágenes combinadas junto con sus máscaras sean organizadas de manera aleatoria.

Debido a la limitante computacional, se hizo que todas las imágenes del dataset de entrenamiento y pruebas también tuvieran una dimensión extra para trabajar con la arquitectura presentada. Esto implica usar tensores de dimensiones 128x128x48x1. Sin embargo, la arquitectura tiene un límite de 18 capas y 22 millones de parámetros entrenables, tal que se hace una reducción en el eje  $z$  de tal manera que los tensores a usar en la red son de la forma 128x128x16x1. Teniendo así solo 16 cortes en el eje axial.

Se construyó la red neuronal usando la arquitectura presentada en [11] donde se hace una representación de U-Net 3D. En la Figura 3, se puede apreciar que la red consiste en una parte convolucional hacia abajo (lado izquierdo) y una parte convolucional hacia arriba (lado derecho).



**Figura 3. Arquitectura de la red U-Net 3D**

El lado izquierdo tiene como objetivo extraer las características para ir clasificando a cada voxel en regiones WMH y no WMH. Y el lado derecho tiene como objetivo localizar las regiones WMH de manera más precisa.

La parte convolucional de abajo consiste en dos capas de convolución  $3 \times 3$ , cada una seguida por una unidad lineal rectificadora (ReLU) y una capa de agrupación máxima  $2 \times 2$  para muestreo descendente. Para las dos primeras capas convolucionales, se utilizó un núcleo de tamaño  $5 \times 5$  para manejar las diferentes transformaciones.

Cada paso en la parte convolucional hacia arriba involucra un muestreo ascendente de características seguidas de una capa convolucional  $3 \times 3$  que reduce el número de canales de atributos y dos  $3 \times 3$  capas de convolución seguidas de una capa ReLU. La concatenación se realizó entre convolucionales hacia abajo y hacia arriba usando la línea verde. La inicialización aleatoria se utilizó para inicializar los pesos del modelo.

También el modelo se entrena usando dice loss como función de pérdida, y como parámetro de métricas el factor f1, el cual por cada iteración va indicando cuanto es el valor de la función de costo (en este caso dice loss), cada epoch a usar se repite varias veces siguiendo la metodología usada en [12] la cual consiste en dejar un paciente afuera de la parte de aprendizaje.

Para este entrenamiento se usó 50 epochs y un batch\_size de 1 por cada iteración teniendo un total de 60 iteraciones. A esto se le agregó un modelo de EarlyStopping de tal manera que si no se mejora los valores de la función de pérdida el modelo decide si sigue entrenando o no. En el caso que ya no se desee seguir entrenando se agregó un ModelCheckpoint para que se guarden los pesos de los

arcos dado un EarlyStopping.

Dentro del aprendizaje se entrenó a la red en diferentes momentos con diferentes datasets de entrenamiento para todos se usó el ground truth respectivo este fue orden de entrenamiento:

- Imágenes FLAIR preprocesadas
- Imágenes T1 preprocesadas
- Imágenes T1 sin cráneo sin aumento de datos.
- Imágenes FLAIR sin cráneo con aumento de datos
- Imágenes T1 sin cráneo con aumento de datos

El postproceso incluye los aspectos:

Recorte o relleno de la segmentación obtenida con respecto al tamaño original

Debido a que se necesita tener las mismas características que la imagen segmentada, se debe cortar al ground truth obtenido para poder tener congruencia de manera axial.

### Rotación de las imágenes

Con el fin de tener las imágenes luego de ser segmentadas en el mismo espacio se le realiza una transposición a la imagen para alternar los ejes y que quede en el formato nifti estándar.

### Guardado de las imágenes

Debido a que las imágenes segmentadas tienen solo 16 cortes axiales, no se puede mezclar con el dataset original, por esto se debe tomar a la imagen original, ground truth y guardarlas luego de haber hecho el corte axial respectivo. Teniendo así un nuevo dataset, el cual forma un subconjunto del dataset original. Los resultados indican que el aprendizaje acertado al momento de hacer

las clasificaciones de las lesiones, verificado por cada conjunto de prueba. cada centro presenta imágenes con características diferentes. De la Tabla III, se extrae que el promedio de las lesiones es aceptable.

El coeficiente DICE se tiene un valor cercano a 0.78 lo indica un 78% de exactitud, para los 3 centros estudiados. Sin embargo, la distancia de Hausdorff muestra una mejor exactitud en el modelo a usar, se puede decir que el resultado parece ser bueno, pero va a depender de la maquina a usar.

### Métricas de evaluación y puntaje

Para evaluar la similitud entre el resultado obtenido y el resultado deseado se usan cinco indicadores diferentes que fueron utilizados para comparar y clasificar los métodos por diferentes equipos; esas métricas evalúan la segmentación rendimiento en diferentes aspectos.

De acuerdo con [12], se asume que dado un mapa de segmentación (ground truth) G y un mapa de segmentación P generado por un algoritmo, las cinco

métricas de evaluación se definen de la siguiente manera: coeficiente de similitud de DICE (DSC), distancia de Hausdorff (percentil 95), diferencia de volumen promedio (en porcentaje), sensibilidad para lesiones individuales (Recall).

### Resultados obtenidos en el dataset de prueba

Los hiper parámetros U-Net 3D se configuraron: el tamaño de lote para computar la pérdida de aprendizaje se estableció en 1; la tasa de aprendizaje en 0,00002; el número de épocas en 50. El número de modelos en el conjunto se estableció en 3.

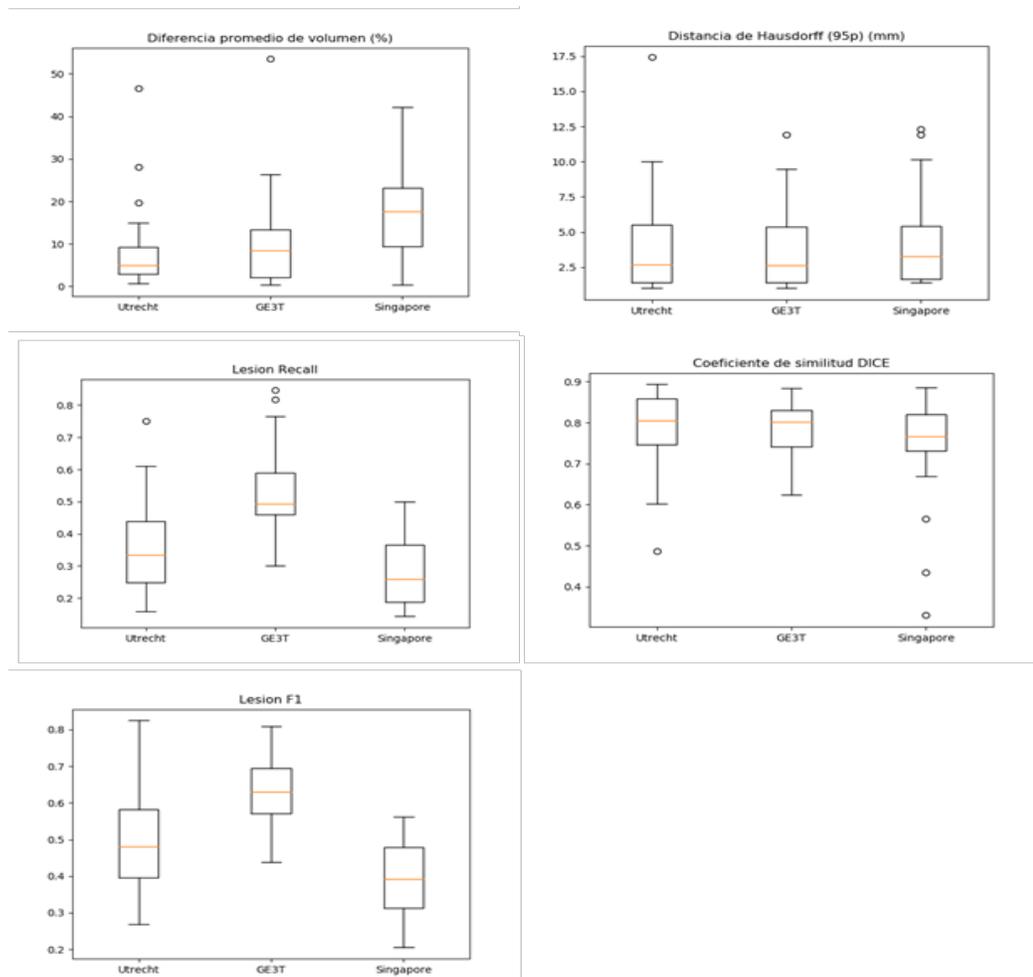
Con esos valores de aprendizaje se realizó el proceso de entrenamiento para esto, la Tabla 3 muestra el resultado de los 5 coeficientes utilizados para realizar la verificación de la efectividad, de estos los valores de H95 y AVD son mejores a medida que el valor es más pequeño, siendo el mínimo 0. Para los otros coeficientes se tiene que entre más alto el valor mejor la segmentación siendo el máximo 1.

**Tabla III Resumen de los resultados obtenidos por el dataset de entrenamiento**

Centro	DSC	H95	AVD	Recall	F1
Utrecht (n = 20)	0,784	4,419	9,087	0,364	0,499
Singapore (n = 20)	0,733	4,552	17,921	0,281	0,396
GE3T (n = 20)	0,782	3,948	11,351	0,541	0,632
Promedio	0,767	4,306	12,786	0,395	0,509

De la Figura 4, al hacer el diagrama por cada uno de los centros para cada parámetro medido, en los primeros gráficos de la parte superior una tendencia aparece, en la diferencia promedio de volumen se tiene que el centro de Singapore presenta una desviación estándar mucho más grande que el centro de Utrecht, sin embargo, en el gráfico de la distancia de Hausdorff, se tiene que son casi equivalentes.

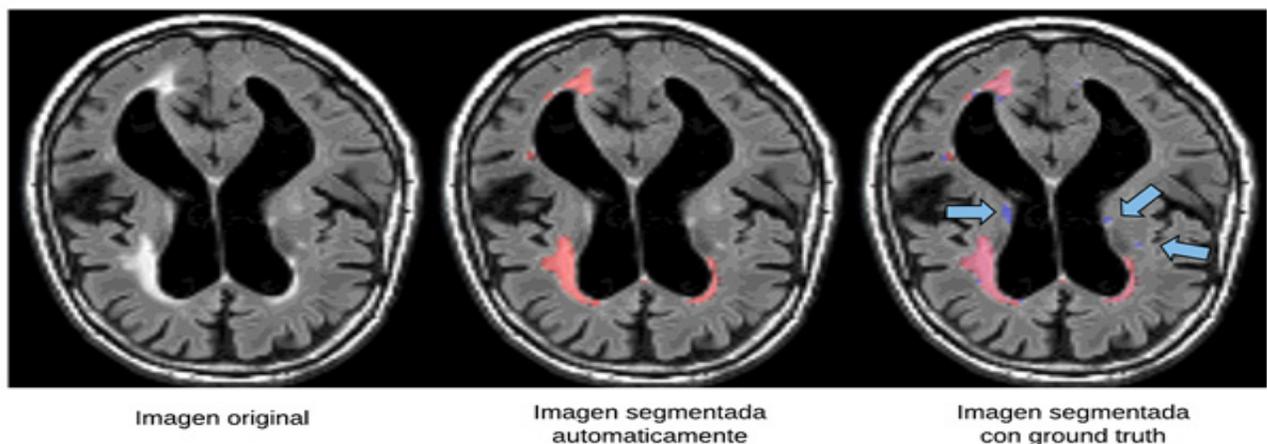
Esto implica que las imágenes generadas por el método de segmentación generan un espacio volumétrico similar al del ground truth, con la excepción que al sacar la distancia entre volúmenes este fluctúa, lo cual puede ser resultado de las lesiones encontradas en dichos sujetos de prueba tal como lo muestra la Figura 4, las lesiones más pequeñas pueden estar ocurriendo como falsos negativos.



**Figura 4** Diagramas de caja de los diferentes centros, en donde se mide un coeficiente diferente por gráfico

Se puede apreciar también de la Figura 4, como el coeficiente de similitud DICE, es bastante homogéneo entre los diferentes centros de estudio, así también como su desviación estándar, lo cual nos indica que nuestro

modelo presenta una exactitud del casi 81%. Esta principal distinción se da al comparar las máscaras de manera manual como se da en la Figura 5.



**Figura 5.** Comparación entre la imagen original, la máscara obtenida y el ground truth

En la Figura 5 el principal inconveniente del método es detectar las lesiones pequeñas en espacio plano como lo muestran los análisis de diagrama de bloque, debido a que muchas lesiones pueden tener un volumen pequeño o representan un pequeño espacio en z, tal que el modelo 3D no lo detecte como una lesión. Esto se puede visualizar en la Figura 4, en el gráfico del coeficiente F1, en donde se ve que la dispersión de datos para el dataset de Singapore es bajo indicando que probablemente las lesiones en ese conjunto de datos sean mucho menores o con menor volumen que las lesiones observadas en Utrecht.

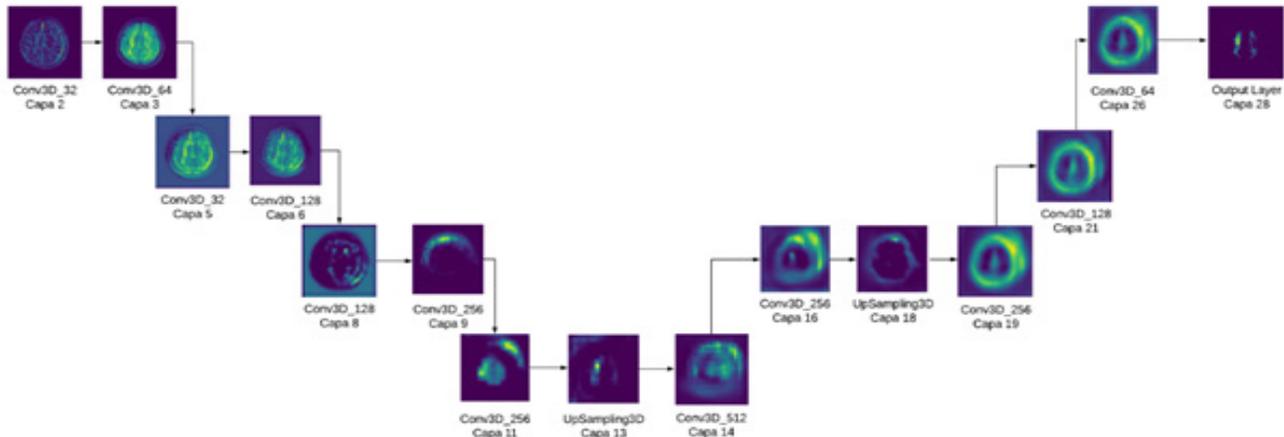


Figura 6. Transformaciones de una imagen dentro de la red.

El procesamiento computacional, se hizo mediante la adquisición del hardware necesario para poder correr la simulación y el proceso de aprendizaje. Para este trabajo se usa una máquina con las características: Memoria 15,3 GB, Procesador Intel, 2.80GHzx8, disco 109,4Gb. El tiempo total de aprendizaje fue de 2 días y 10 horas usando los 4 GB de video.

Debido a esta restricción se limitó el modelo de la red y del dataset de entrada a usar tensores de  $128 \times 128 \times 16 \times 1$ , de un máximo teórico de  $240 \times 240 \times 48 \times 1$ . Sin embargo, para hacer un modelo de aprendizaje con dichas imágenes de entradas se necesitaría una instancia de cómputo mayor. Existen instancias de alquiler, tales como p3.8xlarge de AWS con un valor de 12,24 \$USD [20].

El número total de parámetros entrenables fue de 22398465, lo cual es bastante fuerte de entrenar, sin embargo, con las restricciones el proyecto es considerablemente viable debido a que se usa una máquina de gama media para poder hacer los cálculos.

## I.V DISCUSIÓN

U-Net es una arquitectura de red neuronal convolucional presentada en [11] que ha permitido a muchos investigadores hacer modelos de aprendizaje en imágenes de resonancia magnética, se han encontrado [21] limitantes para las arquitecturas que analizan imágenes en

## Análisis de una imagen a través de la red

La resolución de las imágenes es clave al momento de hacer la segmentación debido al nivel de detalle, esto también aplica al eje z. Es en esta zona, donde actúa la red para detectar lesiones. La Figura 6 muestra como se transforma la imagen a lo largo de la arquitectura, dentro de las convoluciones se hacen resaltar las lesiones con más volumen en el eje axial, una ventaja al momento de clasificarlas. Así mismo esto mejora si la lesión tiene volumen en el eje z.

3D, debido a que no son muy buenos para detectar lesiones pequeñas como las hiperintensidades de sustancia blanca cerebral, sin embargo depende del modelo de aprendizaje usado en [12] muestra, una metodología de aprendizaje que se puede usar para potenciar los resultados. Por esto en el trabajo actual, se usó una metodología de aprendizaje adaptativo en el cual se entrena varias veces a la red de tal manera que aprenda a identificar las posibles lesiones.

U-Net3D tiene la principal ventaja de detectar las características tridimensionales, aunque se puede usar una arquitectura 2D y realizar cortes axiales de las imágenes, como lo han realizado diferentes autores [8,22]. El algoritmo desarrollado detecta de manera correcta lesiones pequeñas en el eje axial pero que tienen volumen en el eje z, esto significa que las lesiones encontradas fueron segmentadas correctamente debido a que el nivel volumétrico permite saber la forma que tienen las lesiones antes de incurrir en la identificación axial.

Se demostró en esta investigación que se presentan problemas o falsos negativos cuando las lesiones son pequeñas en el eje axial pero no tienen mucha profundidad. Así mismo se presentan falsos positivos cuando las lesiones tienen bastante información en el eje z, lo que hace que el sistema prediga la forma que la lesión va a tener en las capas superiores y eso puede ser confundido fácilmente con la intensidad normal de una imagen

Durante el proceso de escoger la arquitectura correcta se probó con diferentes configuraciones siendo las más relevantes las presentadas en [11,12]. Sin embargo la estructura 2D mostró inconvenientes de memoria, por esto se hizo un mal proceso de aprendizaje ya que hubo que restringir los hiperparámetros para hacer funcionar la arquitectura. Se demostró que la muestra en 2D también introducía mucho ruido a la imagen reconstruida e hizo dudar en la diferencia entre las lesiones y el líquido cefalorraquídeo.

#### IV. CONCLUSIONES

- La metodología aquí planteada puede ser usada en una máquina con recursos limitados obteniendo una eficacia del 81%.

- El sistema propuesto mostró un buen desempeño en diferentes escáneres y protocolos usados.

- No se hizo un análisis exhaustivo de las diferencias entre las arquitecturas 3D y 2D, lo que implica que se necesitaría profundizar en el estudio para verificar con mayor detalle porqué un sistema actuó mejor que el otro.

- El método empleado muestra la alta eficacia de las redes neuronales convolucionales al momento de realizar la segmentación automática, de tal manera que muestra su potencial para pruebas reales.

Este trabajo queda abierto a investigaciones futuras, para resolver la identificación de lesiones pequeñas de poca profundidad, así mismo, optimizar la asignación del uso del GPU el rendimiento puede ser mucho mayor.

Debido que existen diferentes máquinas de resonancia magnética en el mercado, realizar un algoritmo válido para todas es problemático. Una restricción del problema fue perder información de capas axiales y tomar solo las 16 capas principales. Se debería probar la eficacia del algoritmo en un recurso computacional mayor que permita tomar todas las imágenes 3D y procesarlas.

#### V. REFERENCIAS

[1]. Lampe, L., Kharabian-Masouleh, S., Kynast, J., Arelin, K., Steele, C. J., Löffler, M., Bazin, P.-L. (2017). Lesion location matters: The relationships between white matter hyperintensities on cognition in the healthy elderly. *Journal of Cerebral Blood Flow & Metabolism*, 0271678X17740501.

[2]. Pantoni, L. (2010). Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *The Lancet Neurology*, 9(7), 689-701. doi: 10.1016/S1474-4422(10)70104-6

[3]. Caligiuri, M. E., Perrotta, P., Augimeri, A., Rocca, F., Quattrone, A., & Cherubini, A. (2015). Automatic Detection of White Matter Hyperintensities in Healthy Aging and Pathology Using Magnetic Resonance

Imaging: A Review. *Neuroinformatics*, 13(3), 261-276. doi: 10.1007/s12021-015-9260-y

[4]. Chancay, O., Haro, T., Yapur, M., Alvarado, R., Pastor, M., & Loayza, F. (2015). Nuevo Biomarcador en la Enfermedad de Parkinson Mediante el Análisis y Cuantificación de Lesiones Cerebrales en Secuencias Flair Obtenidas por Resonancia Magnética (ACL-Tool). *Revista Tecnológica-ESPOL*, 28(5).

[5]. Wardlaw, J. M., Smith, E. E., Biessels, G. J., Cordonnier, C., Fazekas, F., Frayne, R., Dichgans, M. (2013). Neuroimaging neurodegeneration. *The Lancet Neurology*, 12(8), 822-838. doi: 10.1016/S1474-4422(13)70124-8. standards for research into small vessel disease and its contribution to ageing and connections. Paper presented at the Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on.

[6]. Maniega, S. M., Hernández, M. C. V., Clayden, J. D., Royle, N. A., Murray, C., Morris, Z., . . . Bastin, M. E. (2015). White matter hyperintensities and normal-appearing white matter integrity in the aging brain. *Neurobiology of aging*, 36(2), 909-918.

[7]. Dalca, A. V., Balakrishnan, G., Guttag, J., & Sabuncu, M. R. (2018). Unsupervised Learning for Fast Probabilistic Diffeomorphic Registration. *arXiv preprint arXiv:1805.04605*.

[8]. Orbes-Arteaga, M., Sørensen, L., Modat, M., Cardoso, M. J., Ourselin, S., Nielsen, M., & Pai, A. (2018). Simultaneous synthesis of FLAIR and segmentation of white matter hypointensities from T1 MRIs.

[9]. Pandey, R. K., Vasan, A., & Ramakrishnan, A. (2018). Segmentation of Liver Lesions with Reduced Complexity Deep Models. *arXiv preprint arXiv:1805.09233*.

[10]. Respino, M., Kuceyeski, A., Hoptman, M., Victoria, L., Scult, M., Chowdhury, N., Gunning, F. (2018). T111. Changes in Connectivity Induced by White Matter Hyperintensities Predict Dysexecutive Behaviors in Late-Life Depression. *Biological Psychiatry*, 83(9), S171.

[11]. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation. Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention.

[12]. Li, H., Jiang, G., Wang, R., Zhang, J., Wang, Z., Zheng, W.-S., & Menze, B. (2018). Fully Convolutional Network Ensembles for White Matter Hyperintensities Segmentation in MR Images. *arXiv preprint arXiv:1802.05203*.

[13]. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., . . . Isard, M. (2016). Tensorflow: a system for large-scale machine learning. Paper presented at the OSDI.

[14]. Chollet, F. (2015). Keras: Deep learning library for theano and tensorflow. URL: <https://keras.io/k>, 7(8).

- [15]. Brett, M., Hanke, M., Cipollini, B., Côté, M.-A., Markiewicz, C., Gerhard, S., Kastman, E. (2016). nibabel: 2.1.0. Zenodo.
- [16]. Estève, L. (2015). Big data in practice: the example of nilearn for mining brain imaging data. Paper presented at the Scipy 2015.
- [17]. Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Flitney, D. E. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23, S208-S219.
- [18]. Bradski, G., & Kaehler, A. (2000). OpenCV. Dr. Dobb's journal of software tools.
- [19]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct), 2825-2830.
- [20]. Amazon, A. (2014). Amazon.
- [21]. Jin, D., Xu, Z., Harrison, A. P., & Mollura, D. J. (2018). White matter hyperintensity segmentation from T1 and FLAIR images using fully convolutional neural networks enhanced with residual.
- [22]. Li, H., Zhang, J., Muehlau, M., Kirschke, J., & Menze, B. (2018). Multi-Scale Convolutional-Stack Aggregation for Robust White Matter Hyperintensities Segmentation. arXiv preprint arXiv:1807.05153.
- [23]. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95.
- [24]. Jones, E., Oliphant, T., & Peterson, P. (2014). {SciPy}: open source scientific tools for {Python}.
- [25]. Lowekamp, B. C., Chen, D. T., Ibáñez, L., & Blezek, D. (2013). The design of SimpleITK. *Frontiers in neuroinformatics*, 7, 45.
- [26]. Moreso, V. G. (2015). Adaptación de algoritmos de aprendizaje automático para su ejecución sobre GPUs.
- [27]. Walt, S. v. d., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2), 22-30.