# SEGMENTACIÓN DE LESIONES DE ACCIDENTE CEREBROVASCULAR UTILIZANDO REPRESENTACIONES PROFUNDAS QUE INCLUYAN MECANISMOS DE ATENCIÓN

Sebastian Florez Rojas

**Director:**

Fabio Martínez Carrillo, Ph.D

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Ingeniería de Sistemas e Informática

Programa en Ingeniería de Sistemas e Informática

Bucaramanga

2023

# SEGMENTACIÓN DE LESIONES DE ACCIDENTE CEREBROVASCULAR UTILIZANDO REPRESENTACIONES PROFUNDAS QUE INCLUYAN MECANISMOS DE ATENCIÓN

**SEBASTIAN FLÓREZ ROJAS**

**Trabajo de grado para optar por el título de:**
**Ingeniero de Sistemas**

**Director:**
**Fabio Martínez Carrillo**
**Doctor en ingeniería de sistemas y computación**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER**
**FACULTAD DE INGENIERÍAS FISICOMECÁNICAS**
**ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA**
**BUCARAMANGA**
**2023**

## AGRADECIMIENTOS

El autor expresa sus agradecimientos:

Primero que todo, a Santiago Gómez y el profesor Fabio Martínez, que con santa paciencia me acompañaron a hacer este trabajo hasta el ultimo segundo posible.
Quiero agradecer a mis compañeros de grupo, siempre dispuestos a jugar UNO cuando me sentaba a trabajar, una constante inspiración para acabar mas pronto.
Quiero agradecer a mis padres, no hubiera podido hacer nada de esto sin ellos apoyándome y manteniéndome en la casa sin costo durante todo el proceso.
Finalmente, agradezco a mi antiguo computador, que descarté a mitad de camino, pero que luchó hasta el final para correr todos los experimentos que desarrollé pero no incluí en este documento.

# CONTENIDO

## LISTA DE FIGURAS

**pág**

## LISTA DE TABLAS

**pág**

# RESUMEN

**Título:** SEGMENTACIÓN DE LESIONES DE ACCIDENTE CEREBROVASCULAR UTILIZANDO REPRESENTACIONES PROFUNDAS QUE INCLUYAN MECANISMOS DE ATENCIÓN

**Autor:** Sebastian Florez Rojas

**Palabras Clave:** mecanismos de atención, tomografía computarizada, segmentación, ACV, mapa paramétrico

**DESCRIPCIÓN:**

La tomografía computarizada (CT) es hoy en día la secuencia de imágenes diagnósticas más utilizada para el análisis de hallazgos y detección temprana de los ACV. En estos estudios las lesiones cerebrales son observadas como regiones hipo-atenuadas y su principal uso es en estudios de tamizaje para discriminar entre los posibles tipos de lesión. En la literatura se han propuesto representaciones neuronales para la localización, la delineación y caracterización de lesiones relacionadas con ACV para apoyar la delineación de estos estudios. Sin embargo, las estrategias actuales presentan limitaciones para caracterizar estas lesiones debido a su alta variabilidad en cuanto a su apariencia y geometría. Además estas arquitecturas típicamente se entrenan sobre las secuencias completas, en donde las lesiones cerebrales representan aproximadamente un 5 % de la masa cerebral. En el presente trabajo se desarrolló una arquitectura de tipo codificador-decodificador, que entrenada bajo un esquema supervisado, aprende a realizar segmentaciones a partir de delineaciones de expertos sobre estudios CT. Para enfocarse en las regiones asociadas a las lesiones, en este trabajo se realizo una red que incluye mecanismos de atención para establecer relaciones no-locales que representan la geometría de la lesión. Sumado a esto, se desarrolló una segunda fase de entrenamiento a la red para utilizar la información obtenida en la primera fase en un nuevo entrenamiento para refinar los resultados obtenidos inicialmente en la primera fase. La estrategia fue validada en el conjunto de datos ISLES 2017, obteniendo un DSC de 0.66 con 0.67 de precisión.

---

* Trabajo de Grado

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez, PhD.

# ABSTRACT

**Title:** Segmentation of ischemic strokes lesions using deep representations that include attention mechanisms

**Author:** Sebastian Florez Rojas

**Keywords:** Parametric map, attention, computed tomography, segmentation

**DESCRIPTION:**

Computed tomography (CT) is nowadays the most utilized diagnostic imaging for the early analysis and detection of stroke. From these studies, the lesions are observed as hipo-attenuated regions and its main use is in screening studies to determine the type of stroke lesion. In the state of the art there have been proposed multiple computational strategies for the localization, characterization and delineation of stroke related lesions. Nevertheless, these strategies have characterization limitations on these lesions due to their high variability on appearance and geometry. Besides, this architectures are tipically trained on complete sequences where the lesions are approximately 5 % of brain. This proposed work is aimed to develop an encoder-decoder type architecture to be trained in a supervised scheme, learning delineation from experts in CT studies. In order to focus on the regions associated with the lesions, in this work a neural network including attention mechanisms was made to stablish non-local relationships representing the geometry of the lesion. Added to this, we developed a second training phase in the architecture to use the information found in the first phase in a new training, refining the results acquired from the first phase. This strategy was validated on the ISLES 2017 dataset, obtaining a DSC of 0.66 with a precision of 0.67 in the validation dataset.

---

* Degree Work

** Faculty of Physics-Mechanics Engineering. School of Systems Engineering and Informatics. Advisor: Fabio Martínez, PhD

## INTRODUCCION

Los accidentes cerebrovasculares (ACV) son la enfermedad con la segunda mayor tasa de mortalidad en el mundo y la primer causa de discapacidad en países desarrollados [1]. Las lesiones de ACV están relacionadas con la interrupción del flujo sanguíneo, causando daños irreparables por muerte del tejido celular del cerebro [2]. La intervención temprana resulta crítica para garantizar pronósticos favorables del paciente. Para poder realizar dichas intervenciones, es necesaria la previa identificación, localización y delineación de la lesión. Dicho análisis en la práctica clínica es soportado por estudios imagenológicos como la tomografía computarizada (CT, por sus siglas en inglés) debido a su alta disponibilidad y rápida adquisición. Esta modalidad es ideal para realizar estudios de triaje, permitiendo diferenciar entre ACV de tipo isquémico y hemorrágico, además permite visualizar signos tempranos tales como: hipodensidad, pérdida de la diferenciación entre la sustancia gris y blanca, e hiperdensidad de la arteria cerebral media [3]. Sin embargo, la segmentación manual de las lesiones de ACV es una tarea tediosa que toma aproximadamente 15 minutos por caso [4]. Además, en la literatura se reporta una baja concordancia entre expertos [5].

---

[1] Gregory A Roth et al. “Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study”. En: *Journal of the American College of Cardiology* 76.25 (2020), págs. 2982-3021.

[2] Islem Rekik et al. “Medical image analysis methods in MR/CT-imaged acute-subacute ischemic stroke lesion: Segmentation, prediction and insights into dynamic evolution simulation models. A critical appraisal”. En: *NeuroImage: Clinical* 1.1 (2012), págs. 164-178.

[3] Grant Mair et al. “Sensitivity and specificity of the hyperdense artery sign for arterial obstruction in acute ischemic stroke”. En: *Stroke* 46.1 (2015), págs. 102-107.

[4] Anne L. Martel et al. “Measurement of infarct volume in stroke patients using adaptive segmentation of diffusion weighted MR images”. En: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 1999. DOI: 10.1007/10704282_3.

[5] Anders B. Neumann et al. “Interrater agreement for final infarct mri lesion delineation”. En: *Stroke* 40.12 (2009), págs. 3768-3771. DOI: 10.1161/STROKEAHA.108.545368.

Hoy en día, existen estrategias de aprendizaje profundo que pueden detectar diferencias entre regiones de tejido sano e hipoperfundido de manera relativamente fiable y en corto tiempo, logrando soportar esta compleja tarea en el diagnóstico [6] [7].

Recientemente se han propuesto diferentes estrategias para tratar de dar solución a la difícil tarea de segmentar lesiones isquémicas desde imágenes de CT. Por ejemplo, un primer grupo de trabajos utilizó redes convolucionales para modelar la información multicontexto que ofrecen las imágenes de CT perfusión (CTP). Un segundo grupo de arquitecturas han sido diseñadas para recibir datos volumétricos, siendo provechoso para obtener patrones en 3d de la lesión, pero limitando el conjunto de datos de entrenamiento. Persiguiendo un enfoque distinto hay un tercer grupo de trabajos que se han enfocado en enfrentar el desbalance de datos. Por ejemplo, en lugar de tomar las imágenes completas, se tomaron parches seleccionados de estas de manera independiente manteniendo en todo momento el balance entre parches que contienen lesión y los parches sanos. Un cuarto grupo de trabajos se inclinó por utilizar una función de pérdida en una red distinta, buscando tomar en cuenta la forma y los bordes de la lesión como una variable importante durante en entrenamiento. Un último grupo de trabajos que obtuvo resultados sobresalientes se enfocó en la creación de imágenes sintéticas a partir de secuencias CT, buscando recrear las imágenes de MRI para así obtener las ventajas que estas proveen.[8] [6]

Este trabajo presenta una arquitectura codificador-decodificador que toma imágenes de CT y sus mapas paramétricos para la segmentación de lesiones cerebrovasculares isquémicas

---

[6] Guotai Wang et al. “Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks”. En: *Medical Image Analysis* 65 (2020), pág. 101787. DOI: 10.1016/j.media.2020.101787. arXiv: 2007.03294.

[7] Liangliang Liu et al. “Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease”. En: *Medical Image Analysis* 65 (2020). DOI: 10.1016/j.media.2020.101791.

[8] Pengbo Liu. “Stroke lesion segmentation with 2D novel CNN pipeline and novel loss function”. En: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11383 LNCS (2019), págs. 253-262. DOI: 10.1007/978-3-030-11723-8_25.

con el apoyo de mecanismos de atención cruzada. Además, la arquitectura propuesta es complementada con mecanismos de supervisión profunda. La metodología realizada consta de un proceso de dos pasos en el cual se realiza una segmentación inicial a través de una red que realiza un entrenamiento de segmentación completo como primer paso. De los mecanismos de atención implementados en esta arquitectura se extraen una serie de datos relevantes que tras un preprocesamiento son incluidos en la entrada de una segunda red similar a la primera. Una vez más, esta red es entrenada para realizar una segmentación de la lesión, buscando que al haber recibido en la entrada los datos de la red anterior pueda obtener un resultado mas refinado.

# 1. FUNDAMENTOS Y TRABAJO PREVIO

## 1.1. Accidente cerebrovascular isquémico

Un accidente cerebrovascular se presenta cuando se interrumpe el flujo de sangre en el cerebro [9]. El tipo más común de estos son los accidentes isquémicos, los cuales son causados por un bloqueo en los vasos sanguíneos, usualmente por un coágulo. En la figura 1 se ilustra una lesión isquémica, así como su observación a través de diferentes modalidades imagenológicas.

En este sentido, una lesión isquémica no produce sangrado, pero impide o dificulta el paso de sangre a una sección del cerebro causando una variedad de síntomas dependiendo del área afectada. Al bloquearse el flujo de sangre en el cerebro se empieza a generar la muerte de ciertos tejidos, causando daños permanentes [9]. Además, las lesiones isquémicas están compuestas por regiones denominadas núcleo y penumbra. El tejido que se encuentra en el núcleo ha muerto y no se puede recuperar mientras que la penumbra es el tejido que está en riesgo y que podría recuperarse si se realiza el tratamiento de manera oportuna[10].

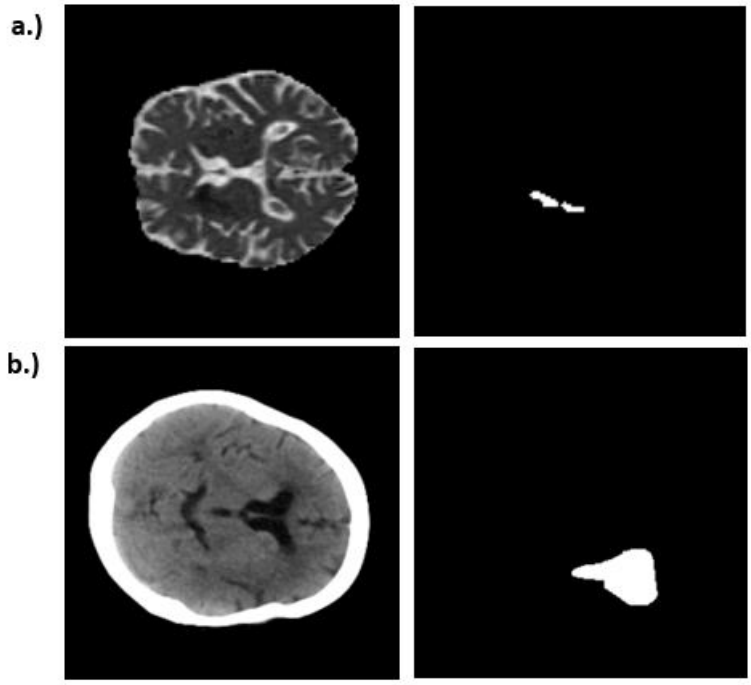
En la rutina clínica, las imágenes son la alternativa diagnóstica más eficaz para la observación, localización y caracterización de las lesiones relacionadas con los accidentes cerebrovasculares. Particularmente el CT suele ser el primer procedimiento realizado para las labores de triaje ya que permite distinguir de manera oportuna si la lesión es isquémica o hemorrágica. La adquisición de secuencias imagenológicas utilizando este procedimiento es relativamente rápida, los scanner utilizados para su adquisición se encuentran en la mayoría de los hospitales y, además, se cuenta con la posibilidad de complementarse con un estudio

---

[9] Brandi R French, Raja S Boddepalli y Raghav Govindarajan. “Acute ischemic stroke: current status and future directions”. En: *Missouri medicine* 113.6 (2016), pág. 480.

[10] Gregory W Albers et al. “Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging”. En: *New England Journal of Medicine* 378.8 (2018), págs. 708-718.

**Figura 1.** Accidente cerebrovascular isquémico, mostrado en: (primera fila) la modalidad ADC (primera columna) y su respectiva segmentación (segunda columna), y en (segunda fila) la modalidad CT (primera columna) y su respectiva segmentación (segunda fila).

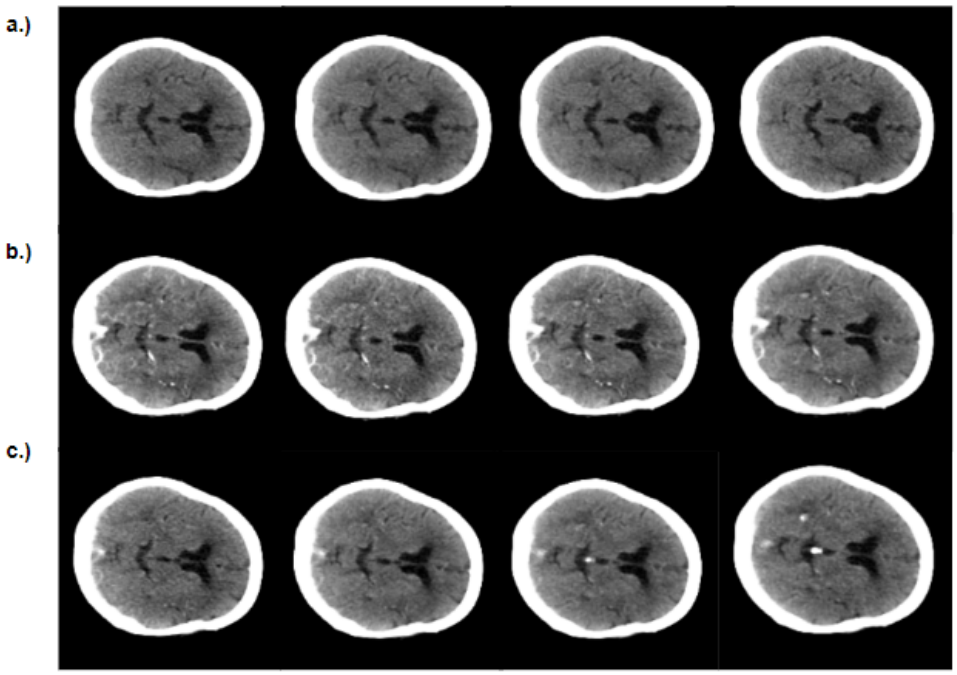



de perfusión. Estas características son primordiales en la caracterización de lesiones, debido a la dinámica de la enfermedad, donde el tiempo oportuno en el diagnóstico es clave para tener un pronóstico favorable del paciente. En las secuencias CT, el tejido afectado por la lesión puede identificarse como hipodenso en sus etapas tempranas, es decir tonalidades más oscuras con respecto al resto del tejido. A pesar de que esta modalidad imagenológica es la más utilizada en el escenario clínico, el contraste y observación del tejido afectado es muy sutil, siendo casi imperceptible para lesiones agudas o tempranas. Es por ello, que los análisis observacionales son en muchas ocasiones reforzados con otros procedimientos clínicos o alternativas de contraste.

### 1.2. Tomografía computarizada de perfusión (CTP)

La tomografía de perfusión se diferencia de la tomografía común en la inserción de un líquido de contraste en el torrente sanguíneo del paciente, ayudando así a tener un mayor contraste de la lesión y en consecuencia una mejor caracterización del tejido afectado. Tras ser inyectado este agente de contraste, se realiza una serie de tomografías que permiten observar la circulación de un líquido a través del tejido cerebral a lo largo del tiempo. Es decir una tomografía con contraste o CTP por sus siglas en inglés se extiende a lo largo del tiempo. En otras palabras, mientras un CT es una imagen 3D que muestra el estado del cerebro en un instante del tiempo, un CTP es una imagen 4D compuesta por una serie de imágenes 3D que muestran el flujo de sangre a través del cerebro como describe la figura 2

**Figura 2.** Ejemplo de slices de perfusión. Todos corresponden a una misma sección del cerebro. a.) Primeros frames de la secuencia b.) Frames de la mitad de la secuencia c.) Frames finales de la secuencia



Como resultado se obtienen volúmenes que se propagan a través del tiempo, lo cual resulta difícil de analizar mientras el agente de contraste se propaga. La concentración del agente de contraste $C_x(t)$, en una región de interés $x$, a través del tiempo $t$ puede ser definida, de acuerdo a tres funciones [11]:

- **Función de transporte ($h(t)$)**. Es una función de densidad de probabilidad que estima el tránsito del agente en un instante $t$, sobre una región de interés $x$ y depende de la estructura vascular.
- **Función residual ($R(t) = 1 - \int_0^t h(t)$).** Es el porcentaje de contraste que aún está en la arteria. En este sentido, esta función empieza en el máximo una vez llega el contraste a la arteria e inicia a decrementar con respecto a la función de transporte $h(t)$.
- **Función de entrada arterial - AIF - ($C_a(t)$)**. Estima la concentración inyectada del agente de contraste en una región de interés $x$, en un tiempo específico $t$

Desde estas tres funciones es posible entonces modelar el flujo del contraste a través del tiempo, como:

$$C_{VOI}(t) = \frac{\rho}{kh} CBF \cdot (\mathrm{C}_a(\mathrm{t}) \bigotimes \mathrm{R}(\mathrm{t}))$$

en el que $\rho$ es la densidad del tejido cerebral, CBF es el flujo de sangre cerebral y $kh$ es un factor que toma en cuenta la diferencia de hematocrito (porcentaje de glóbulos rojos en la sangre) entre la sangre del tejido y la arteria. A partir de esta definición de contraste se pueden resumir en una serie de mapas paramétricos de las secuencias CTP $3D + t$ que definen propiedades dinámicas del agente de contraste durante el paso por los vasos sanguíneos del paciente.

---

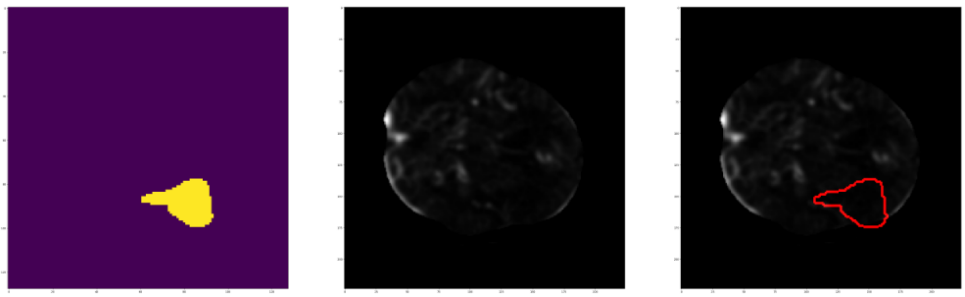
[11] Fernando Calamante et al. “Measuring cerebral blood flow using magnetic resonance imaging techniques”. En: *Journal of cerebral blood flow & metabolism* 19.7 (1999), págs. 701-735.

### 1.3. Mapas paramétricos

Los mapas paramétricos son imágenes 3D obtenidas a partir de procesos computacionales que muestran características específicas dentro del estudio realizado. Cabe resaltar que estos parámetros son dados por algunos fabricantes, según las características físicas de la máquina de adquisición e incorporando algunas variables simuladas. Algunos de estos son:

- **CBF (cerebral blood flow)**. Este parámetro representa el volumen de sangre que recorre el tejido en base de mililitros de sangre que atraviesan cien gramos de tejido. Su valor a lo largo del tiempo se obtiene deconvolucionando la función agente de contraste (C(t)). Los valores cuantitativos del CBF (qCBF) varían dependiendo del método utilizado para estimarlos, pero rondan los 22.72 ± 5.45 ml/100g-min en la materia blanca y los 50.46 ± 18.08 ml/100g-min en la materia gris al ser calculados con spin echo EPI[12]. Un ejemplo de este mapa paramétrico puede observarse en la figura 3.

**Figura 3.** Ejemplo de CBF, seguido de la lesión correspondiente al slice y el contorno de la lesión sobre el CBF



- **CBV (cerebral-blood-volume).** Este parámetro representa el volumen de la sangre en una zona del parénquima en una base de mililitros de sangre en cien gramos de tejido.

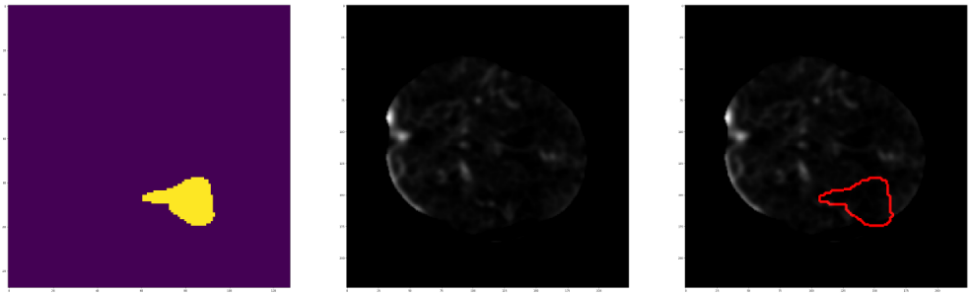
[12]Sandra and Shin, Wanyong and Mouannes, Jessy and Sawlani, Rahul and Ali, Saad and Raizer, Jeffrey and Futterer, Stephen Carroll, Timothy J and Horowitz, “Quantification of cerebral perfusion using the “bookend technique”: an evaluation in CNS tumors”. En: *Magnetic resonance imaging* 26.10 (2008), págs. 1352-1359.

El CBV se calcula con la siguiente fórmula:

$$CBV = \frac{k_h}{\rho} * \frac{\int C_{voi}(t)dt}{\int C_a(t)dt}$$

Al calcular los valores cuantitativos del CBV (qCBV) hay diferencias dependiendo del método utilizado, pero los valores suelen rondar los 1.68 ± 0.41 ml/100g para la materia blanca y los 2.41 ± 0.36 ml/100g para la materia gris al ser calculados con estudios de imagen eco planares (spin echo EPI)[12]. Un ejemplo de este mapa paramétrico puede observarse en la figura 4.

**Figura 4.** Ejemplo de CBV, seguido de la lesión correspondiente al slice y el contorno de la lesión sobre el CBV



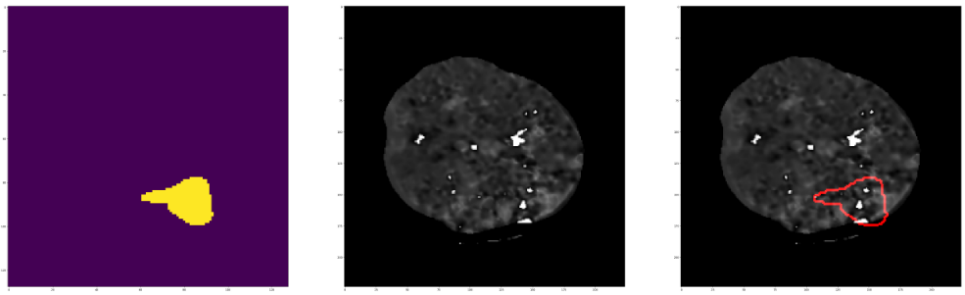
- **MTT (Mean-transit-time)**. Este mapa, obtenido a partir del CBF y el CBV, cuantifica el tiempo promedio que la sangre permanece en un volumen determinado de tejido. El teorema que une este mapa paramétrico con el CBF y el CBV es llamado el principio del volumen central, aproximando la medida como $MTT = \frac{CBV}{CBF}$ [13]. Otro método para obtener el MTT se basa en la función de transporte expuesta anteriormente, a partir de la ecuación:

$$MTT = \frac{\int t*h(t)dt}{\int h(t)dt}$$

[13]Ellen G Hoeffner et al. “Cerebral perfusion CT: technique and clinical applications”. En: *Radiology* 231.3 (2004), págs. 632-644.

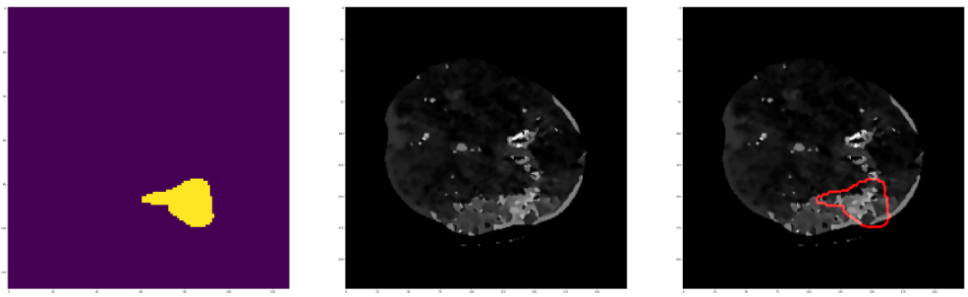
Un ejemplo de este mapa paramétrico puede observarse en la figura 5.

**Figura 5.** Ejemplo de MTT, seguido de la lesión correspondiente al slice y el contorno de la lesión sobre el MTT



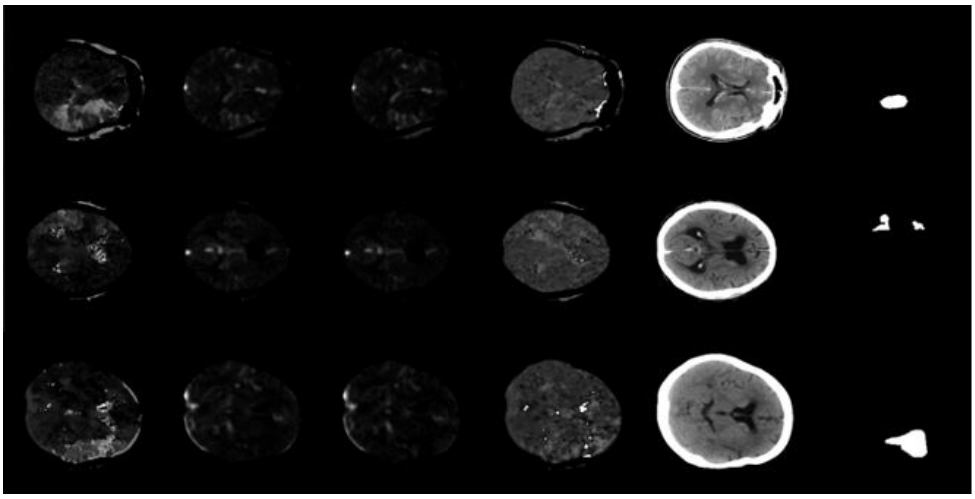
- **Tmax (Time-to-maximum).** El Tmax es el tiempo necesario para alcanzar el punto máximo de la función residuo (R(t)). El Tmax puede ser afectado por una variedad de condiciones experimentales, pero a su vez combina elementos fisiológicos importantes como la dispersión, la demora o el MTT[14]. Un ejemplo de este mapa paramétrico puede observarse en la figura 6.

**Figura 6.** Ejemplo de TMAX, seguido de la lesión correspondiente al slice y el contorno de la lesión sobre el TMAX



[14]Søren and Desmond, Patricia M and Østergaard, Leif and Davis, Stephen M and Connelly, Alan Calamante, Fernando and Christensen. "The physiological significance of the time-to-maximum (Tmax) parameter in perfusion MRI". En: *Stroke* 41.6 (2010), págs. 1169-1174.

**Figura 7.** Mapas paramétricos de estudios en tomografía computarizada de pacientes diferentes. De izquierda a derecha: Tmax, CBV, CBF , MTT, CT y mascara de segmentación de la lesión



### 1.4. Arquitecturas codificador-decodificador para la segmentación

Hoy en día, las arquitecturas más utilizadas para la tarea de segmentación de imágenes médicas son las codificador-decodificador. Estas arquitecturas están compuestas por dos redes que procesan las imágenes médicas y producen un mapa de probabilidad para estimar la máscara de segmentación de los objetos de interés. En términos generales, el codificador toma una imagen de entrada y opera con ella generando vectores embebidos mientras retiene la información que la red considera relevante. El decodificador toma el vector embebido que la primera genera y reconstruye una imagen a partir de este. Al llegar a la salida de la segunda red se obtiene una imagen con el mismo tamaño que la imagen original que entró al codificador.

Un ejemplo de estas arquitecturas es la U-Net, la cual incorpora conexiones de salto desde el codificador al decodificador en niveles de procesamiento con igual dimensionalidad espacial. Al anexar la información correspondiente proveniente del codificador se está otorgando un apoyo al decodificador para preservar información global, relacionada con la forma del objeto

de interés.

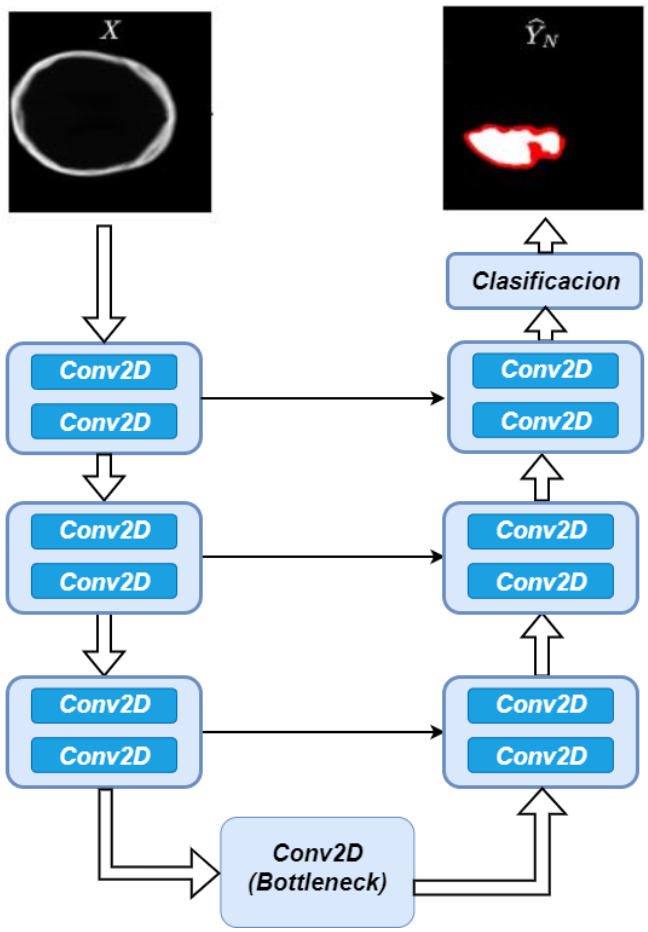



**Figura 8.** Arquitectura U-Net. Cada caja representa una capa de la red, conteniendo dos capas convolucionales que operan con la imagen progresivamente. Al pasar de una capa a otra, disminuyen las dimensiones de la imagen hasta llegar al bottleneck. Posteriormente se realiza un proceso de reconstrucción, aumentando las dimensiones de la imagen hasta alcanzar el tamaño de la imagen de entrada. Las flechas que conectan la red de izquierda a derecha corresponden a las conexiones de salto que transportan la imagen que será concatenada al inicio de la capa objetivo.

Además, en el proceso de upsampling se define un mecanismo de deconvolución que aumenta el tamaño del vector embebido mientras se conserva el contexto de la misma para propagarlo a las capas superiores. Algunos métodos de upsampling son 'vecino más cercano' el cual toma el valor más cercano al punto a evaluar o la interpolación bilinear que deduce el valor a partir de todos los píxeles cercanos.

## 1.5. Mecanismos de atención

Los mecanismos de atención son una estrategia relativamente reciente, la cual fue implementada para mejorar los resultados obtenidos en distintos aspectos del entrenamiento de redes neuronales. En las redes de visión por computador, su funcionamiento se basa en tomar una imagen y relacionar los elementos presentes en esta entre sí para determinar zonas de mayor relevancia para la tarea que se está llevando a cabo.
Por regla general, los mecanismos de atención constan de tres elementos importantes: key, query y value. Estos tres son extraídos de los vectores embebidos utilizados a través de procesos convolucionales y al operarse entre sí dan como salida un vector embebido que representa las relaciones entre sus partes[15]. Existen dos tipos principales de mecanismos de atención: la auto atención y la atención cruzada.

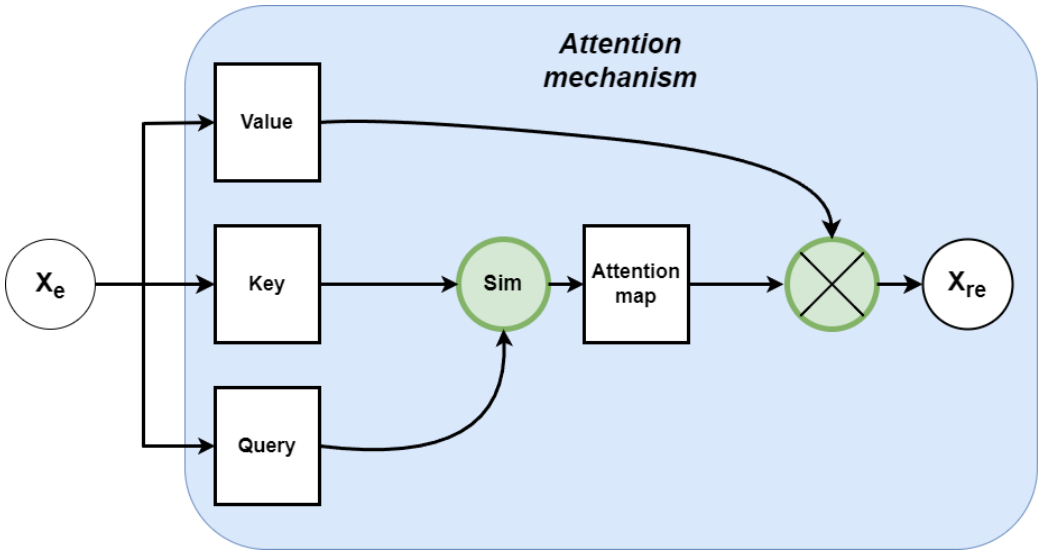
**1.5.1. Auto atención** La auto atención se basa en tomar el vector embebido generado en el codificador y relacionar distintos elementos de este entre si. El key, query y value son todos extraídos de un solo vector a través de procesos convolucionales[15].
Esto es similar a elegir cada píxel de una imagen individualmente y determinar la importancia de la relación que tiene este con todos los otros píxeles de la imagen. Al repetir este proceso para cada píxel, se obtiene una nueva imagen que resalta los puntos de mayor importancia que se detectaron en el proceso. Para mejor entendimiento, se gráfico el proceso en la figura 9.

**1.5.2. Atención cruzada** La atención cruzada se basa en tomar más de un vector embebido y cruzar la información de estos entre sí. Los elementos son los mismos, pero el value y el key provienen del codificador mientras que el query se extrae del decodificador. En este

---

[15]Ashish Vaswani et al. “Attention is all you need”. En: *Advances in neural information processing systems* 30 (2017).

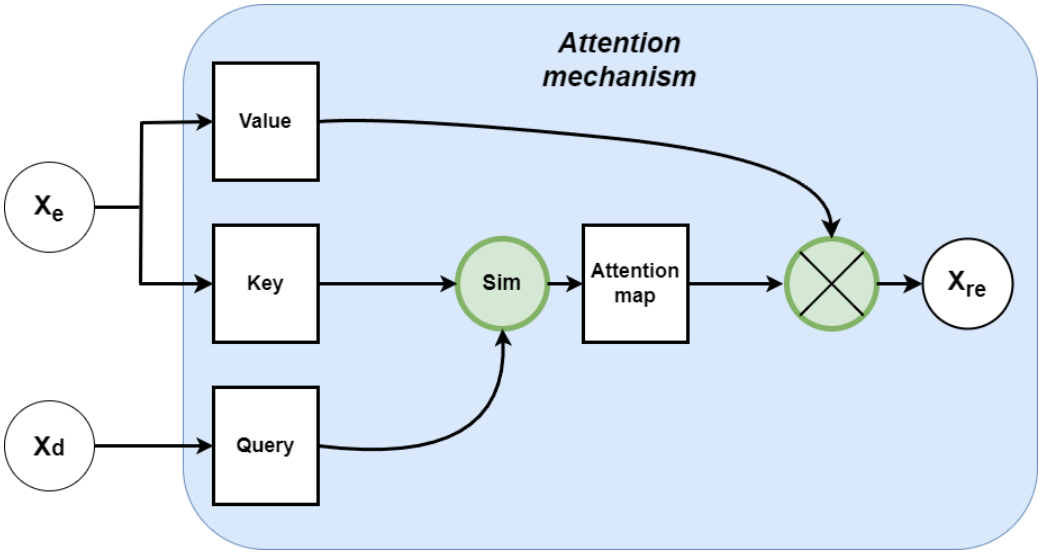
**Figura 9.** Esquema de auto atención




caso, puede entenderse como utilizar una imagen en un punto de la red para determinar los puntos más importantes de otra imagen proveniente de la misma red.

Existen varios tipos de atención cruzada debido a que, al poseer más de un vector como entrada estos tienen que ser unificados a través de diversos métodos. Por ejemplo, un esquema de atención aditiva sumaría los valores del key y del query extraidos de dos matrices con dimensiones idénticas para unificarlos en una sola matriz, mientras que un esquema de atención multiplicativo obtendría el producto punto entre ambas matrices[15]. Para mejor entendimiento, se gráfico el proceso en la figura 10.

**Figura 10.** Esquema de atención cruzada



## 2. ESTRATEGIAS COMPUTACIONALES PARA SOPORTAR LA SEGMENTACION DE LESIONES EN CT

Recientemente, en el estado del arte se han propuesto diversas estrategias, principalmente relacionadas con representaciones profundas para soportar las tareas de segmentación de lesiones. Por ejemplo, la arquitectura U-Net 3D presentada por Tureckova *et al.* ha sido adaptada para integrar el contexto de la lesión[16]. Además de poseer una visión en 3D del problema, se aplicaron convoluciones dilatadas 2x2 que presentaron mejores resultados que las convoluciones sin dilatar. Los resultados obtenidos indican que la capa de entrada es la mas importante a la hora de analizar el contexto de la imagen completa, pero salvo por la aplicación de las convoluciones dilatadas en este punto se carece de un método para extraer esta información.

Similarmente, Dolz *et al.* propuso una variación de la arquitectura U-Net para tomar modalidades de imágenes de manera separada con múltiples codificadores [17], aprendiendo características especializadas intrínsecas en cada modalidad (ver figura 11). Los codificadores están densamente conectados e incorporan módulos de incepción de dos bloques convolucionales con diferentes grados de dilatación. Aunque las convoluciones dilatadas permiten incluir más contexto local en el cómputo, estas operaciones no tienen en cuenta información no-local de las imágenes médicas. Sumado a esto, no se solventa el problema del desbalance de datos, que podría agravar la dificultad que la red tiene para aprender al estar involucrando varias redes codificadoras adicionales a entrenar en su interior sin una ayuda

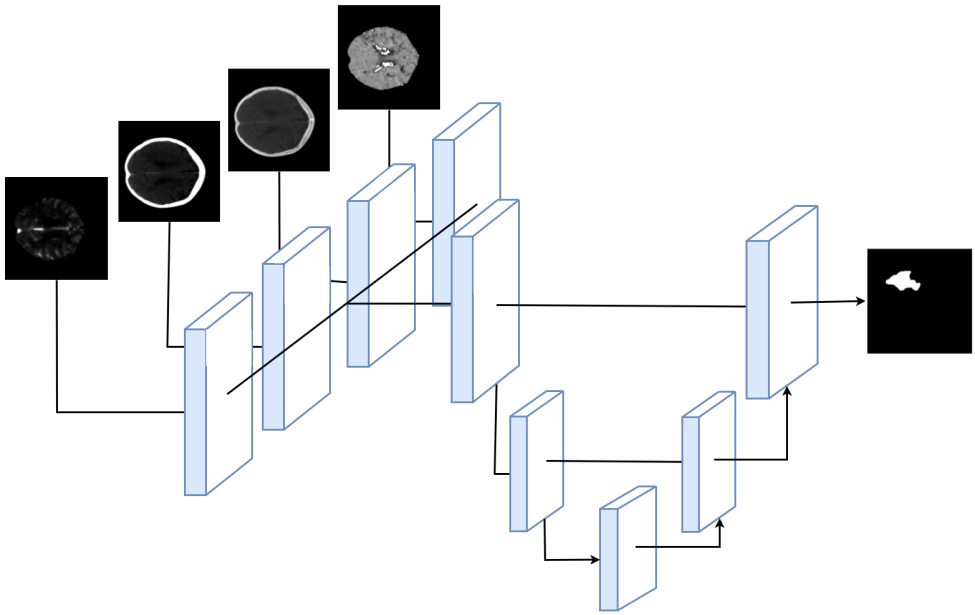
---

[16] Alzbeta Tureckova y Antonio J Rodríguez-Sánchez. “ISLES challenge: U-shaped convolution neural network with dilated convolution for 3D stroke lesion segmentation”. En: *International MICCAI Brainlesion Workshop*. Springer. 2018, págs. 319-327.

[17] Jose Dolz, Ismail Ben Ayed y Christian Desrosiers. “Dense multi-path U-Net for ischemic stroke lesion segmentation in multiple image modalities”. En: *International MICCAI Brainlesion Workshop*. Springer. 2018, págs. 271-282.

para distinguir el tejido sano del hipoatenuado.

**Figura 11.** Ejemplo de red con codificadores individuales para las entradas de la red



Otro grupo de trabajos han intentado resolver el gran desbalance entre tejido sano, y el tejido relacionado con la lesión, para abordar de mejor forma los principios de discriminación. Por ejemplo, Clerigues *et al.* propuso una arquitectura codificador-decodificador asimétrica residual para generar predicciones de lesiones a nivel de parches [18]. La red está entrenada con una función de pérdida híbrida (cross-entropy y dice generalizado) en pequeños grupos de parches extraídos de imágenes que fueron cuidadosamente diseñados para incluir información estratificada. Aunque entrenar con unidades de datos más pequeñas aumenta dramáticamente la cantidad de muestras disponibles, pierde la información espacial provista por una imagen médica completa ya que asume parches de la misma imagen como ejemplos independientes.

---

[18]Albert Clèrigues et al. “Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks”. En: *Computers in biology and medicine* 115 (2019), pág. 103487.

Similarmente, Kervadec *et al.* propuso una función de pérdida para los contornos de las lesiones expresados como la suma de las funciones lineares de las salidas de probabilidad obtenidas por la función de activación softmax regional de las salidas de una red U-Net. Aunque este método representa un avance para el uso de datos altamente desbalanceados, para el problema de la segmentación es igualmente importante desarrollar una red adecuada y capaz de aprovechar la totalidad de los datos dados.

En otros trabajos, Kuang *et al.* propuso una red de aprendizaje multitarea (EIS-Net) para segmentar accidentes cerebrovasculares tempranos. También, se propuso en la literatura la arquitectura EIS-Net que se compone de una red neuronal que se ajusta mediante una representación de pérdida por tripletes (T-CNN)[19]. En el decodificador de esta red, un módulo de compuerta de atención multinivel (MAGM) preserva las activaciones esenciales para segmentar la lesión temprana (EI). El MAGM toma las características de múltiples resoluciones espaciales y aplica deconvoluciones para igualar el tamaño con las características de mayor resolución espacial. La red recibe tres entradas distintas en su triplete inicial, el NCCT original, el NCCT reflejado y un respectivo atlas. Esto permite a la red estudiar con especial detenimiento la información provista por el NCCT, pero se desaprovecha la oportunidad de utilizar los mapas paramétricos de cada caso y la información que estos proveen.

Otros trabajos aprovechan la existencia de imágenes de resonancia magnética (MRI) en dataset públicos para transferir representaciones de la lesión, con mayor contraste, hacia los estudios de CT. Posteriormente, la red de segmentación es entrenada para segmentar las lesiones a partir de datos sintéticos. Por ejemplo, Liu *et al.* [20] propuso una arquitectura convolucional que consiste en un generador, un discriminador y una red de segmentación que estima los bordes de lesiones de accidentes isquémicos. El generador sintetiza imágenes

---

[19]Hulin Kuang et al. “EIS-Net: Segmenting early infarct and scoring ASPECTS simultaneously on non-contrast CT of patients with acute ischemic stroke”. En: *Medical Image Analysis* 70 (2021), pág. 101984.

[20]Pengbo Liu. “Stroke lesion segmentation with 2D novel CNN pipeline and novel loss function”. En: *International MICCAI Brainlesion Workshop*. Springer. 2018, págs. 253-262.

DWI a partir de secuencias de CTP, mientras que el discriminador distingue entre ejemplos generados y reales. La red de segmentación predice el resultado de la lesión a partir de la imagen DWI sintética producida por el generador. Con un objetivo similar está la red presentada por Wang *et al.* [21] la cual genera un DWI sintético para posteriormente usarlo en la segmentación. Hay varias diferencias importantes entre estas aproximaciones, las dos principales son la diferencia en los datos que utilizan como entrada para la red que genera los DWI y la diferencia entre la naturaleza de ambas redes. Por un lado Wang utiliza una red U-Net para este proceso debido al buen rendimiento que ha tenido en el estado del arte. Por su parte, Liu optó por utilizar una red GAN (Generative adversarial networks) la cual es una herramienta de aprendizaje no supervisado enfocado al reconocimiento de patrones para la generación de imágenes.

Ambas implementaciones obtuvieron resultados positivos y muestran un camino a futuro con bastante potencial, pero la principal limitación a la hora de entrenar estas redes es la necesidad de tener datos tanto de CT como DWI de cada paciente. La obtención de ambos estudios para una misma lesión es bastante rara por cuestiones de disponibilidad del equipo, además de la necesidad de realizar ambos estudios con rapidez para poder ser comparables y las diferencias presentes por realizar el estudio en máquinas distintas.

De estos inconvenientes el más importante es la diferencia entre los datos de cada estudio. Por ejemplo, ambos procedimientos generan imágenes 3D las cuales son posteriormente vistas como una serie que posee una cantidad variable de slices 2D. Esto significa que pueden no ser fácilmente comparables entre ambos ya que no corresponden a la misma localización ni información. Esto puede ser mitigado con la utilización de un proceso llamado registro que es utilizado para establecer correspondencia entre imágenes, sin embargo, tiene un alto costo computacional y un tiempo de ejecución considerable para obtener un resultado

---

[21] Guotai Wang et al. “Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks”. En: *Medical Image Analysis* 65 (2020), pág. 101787.

preciso[22].

[22]Guha Balakrishnan et al. “Voxelmorph: a learning framework for deformable medical image registration”. En: *IEEE transactions on medical imaging* 38.8 (2019), págs. 1788-1800.

## 3. PROBLEMA DE INVESTIGACION

El accidente cerebrovascular es una enfermedad multifactorial cuyo tratamiento oportuno es fundamental para reducir el daño neurológico. Hoy en día, las imágenes de CT son la principal herramienta diagnóstica y de triaje para llevar a cabo la cuantificación del área hipoperfundida y determinar el tratamiento del paciente. Sin embargo, en la etapa aguda estas secuencias tienen baja sensibilidad para discriminar tejido afectado, siendo la caracterización de la lesión una tarea altamente variable y compleja, incluso para expertos radiólogos neurointervencionistas. Recientemente, la cuantificación de las lesiones de accidente cerebrovascular se ha tratado de soportar con herramientas de aprendizaje profundo. Aun así, debido a la alta variabilidad que se puede presentar en los datos, al igual que el alto desbalance que se observa entre el tejido sano y el tejido lesionado en los estudios, el problema sigue siendo abierto.

De hecho, los esquemas convolucionales codificador-decodificador son hoy en día las principales herramientas para recuperar lesiones, observadas en estudios CT. Sin embargo el entrenamiento y ajuste de estas representaciones se ve abocado a un problema relacionado con el desbalance de los datos del tejido de la lesión con respecto al tejido sano. De hecho, la proyección hacia un vector embebido en estas arquitecturas hace que se pierdan detalles relacionados con la lesión. Además, las funciones de pérdida diseñadas para mitigar este desbalance de datos no han mostrado resultados completamente satisfactorios. Es así, que en este trabajo se plantea hacer un seguimiento de la lesión en cada una de las proyecciones en el codificador, implementando mecanismos que preserven la lesión en cada etapa.

# 4. OBJETIVOS

## 4.1. Objetivo General

Desarrollar una arquitectura codificador-decodificador que integre mecanismos de atención para la segmentación de lesiones de accidente cerebrovascular en estudios de tomografía computarizada

## 4.2. Objetivos Específicos

- Seleccionar un conjunto de datos que incluyan estudios CT con anotaciones de referencia realizadas por expertos.
- Implementar una red codificador-decodificador para la segmentación de CT.
- Desarrollar mecanismos de atención, en una arquitectura codificador-decodificador, que permitan preservar información durante decodificación de la representación.
- Evaluar la capacidad de la arquitectura desarrollada con respecto a la capacidad de segmentar lesiones delineadas por expertos.

# 5. ENFOQUE PROPUESTO

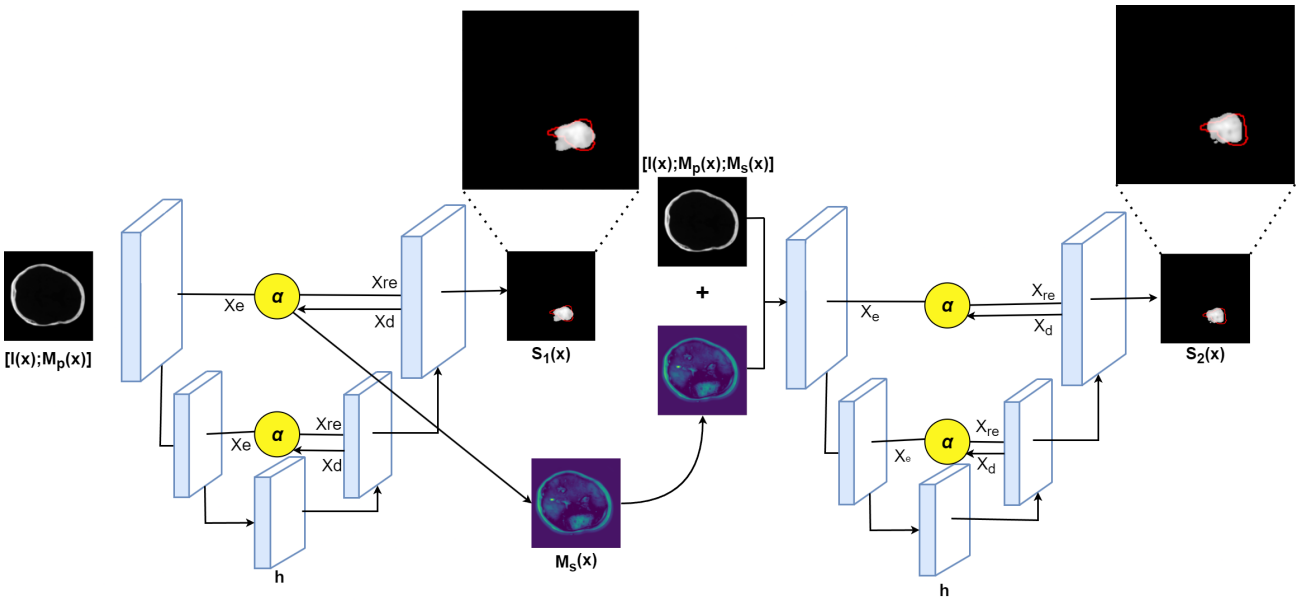
Este trabajo introduce una metodología para la estimación de lesiones de ACV isquémico en secuencias CT, siguiendo dos etapas consecutivas de entrenamiento. En las dos etapas se ajusta una representación de tipo codificador-decodificador, que incluye múltiples mecanismos de atención aditiva cruzada. En la primera etapa, se obtiene una segmentación en baja escala utilizando un entrenamiento supervisado entre la lesión anotada y un conjunto de secuencias CT multiparamétricas. En una segunda etapa de refinamiento, los detalles de la segmentación de la lesión son resaltados al incluir como entrada las respuestas de atención obtenidas en la primera etapa (mapas sintéticos). El enfoque multi-etapa permite realizar estimaciones más precisas de las lesiones isquémicas sobre secuencias CT, y además aprovecha la información resultante de los módulos de atención establecidos en la arquitectura. En la Figura 12 se representa el enfoque propuesto. Parte del contenido de esta sección fue publicado en la conferencia internacional ”SPIE Medical Imaging”[23].

## 5.1. Arquitectura codificador-decodificador con mecanismos de atención

Como base de representación, en este trabajo se utilizó una arquitectura U-Net modificada para la segmentación de lesiones de ACV isquémico desde secuencias multiparamétricas de CT. En nuestro caso particular, se ingresa una imagen CT ($\mathbf{I}(\mathbf{x})$), junto con sus mapas paramétricos ($\mathbf{M_p}(\mathbf{x})$), en un esquema simple de concatenación ($\{\mathbf{I}(\mathbf{x}); \mathbf{M_p}(\mathbf{x})\}$). En la salida, el resultado es una máscara binaria ($\mathbf{S}(\mathbf{x})$), que contiene respuestas binarias con la máscara. Como un típico autoencoder, el objetivo de esta arquitectura es aprender un vector embebido

[23] Gómez, Santiago and Florez, Sebastian and Mantilla, Daniel and Camacho, Paul and Tarazona, Nick and Martínez, Fabio. “An attentional unet with an auxiliary class learning to support acute ischemic stroke segmentation on CT”. En: *Medical Imaging 2023: Image Processing*. Ed. por Olivier Colliot e Ivana Išgum. Vol. 12464. International Society for Optics y Photonics. SPIE, 2023, 124640S. DOI: 10.1117/12.2654269.

**Figura 12.** Metodología propuesta multi-etapa para la estimación de lesiones de ACV isquémico sobre secuencias multiparamétricas CT.



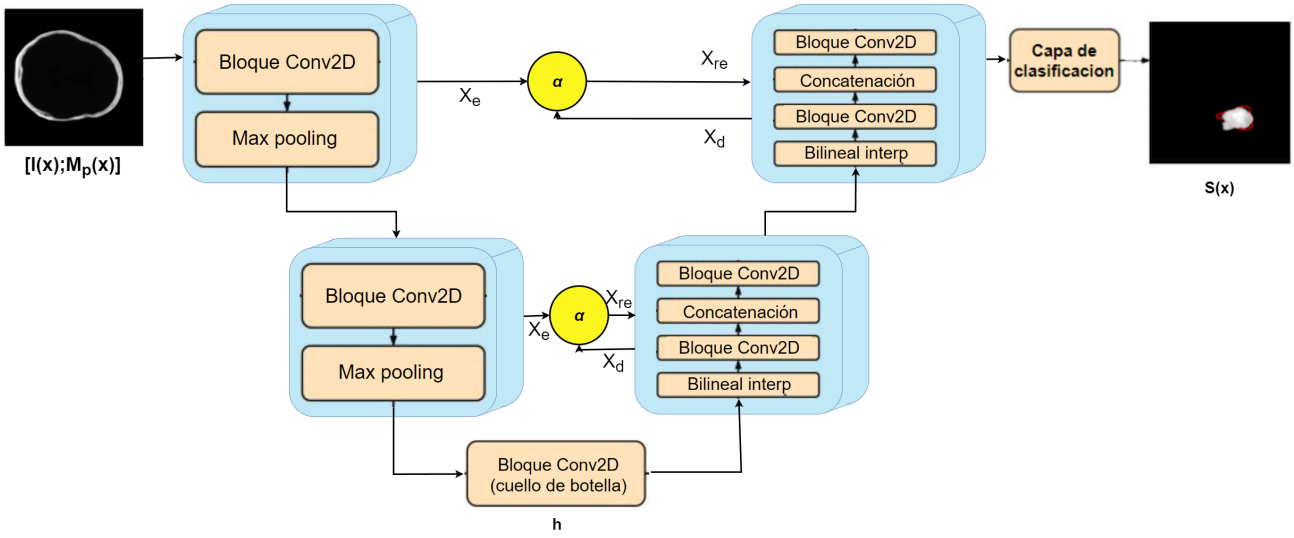

($\mathbf{h} \rightarrow \mathbf{S}(\mathbf{x})$) que codifique la información de entrada, pero que además permita la generación de la segmentación $\mathbf{S}(\mathbf{x})$.

De forma general, la arquitectura U-Net [24] es la arquitectura codificador-decodificador más utilizada para llevar a cabo la tarea de segmentación de lesiones isquémicas. Su diseño que incluye conexiones de salto entre el codificador y decodificador, favoreciendo la integración de información a la representación de la red decodificadora y alivia el problema de desvanecimiento del gradiente. Particularmente en cada nivel de procesamiento del decodificador, las respuestas convolucionales del decodificador ($X_d^L$) son complementadas con las respuestas del codificador ($X_e^L$) en el mismo nivel ($L$) de procesamiento ($\left[X_e^L; X_d^L\right]$). Sin embargo, la incorporación de características débilmente correlacionadas puede ser perjudicial para la

[24] Olaf Ronneberger, Philipp Fischer y Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". En: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, págs. 234-241.

representación del decodificador y afectar la segmentación de las lesiones isquémicas. Por esta razón, incluimos mecanismos de atención aditiva cruzada en las conexiones de salto como se ilustra en la Figura 13.

**Figura 13.** Arquitectura U-Net enriquecida con mecanismos de atención aditiva cruzada, centrada en los límites de las lesiones de ACV isquémico. Corresponde a las redes individuales presentes en la fase 1 y 2 del pipeline

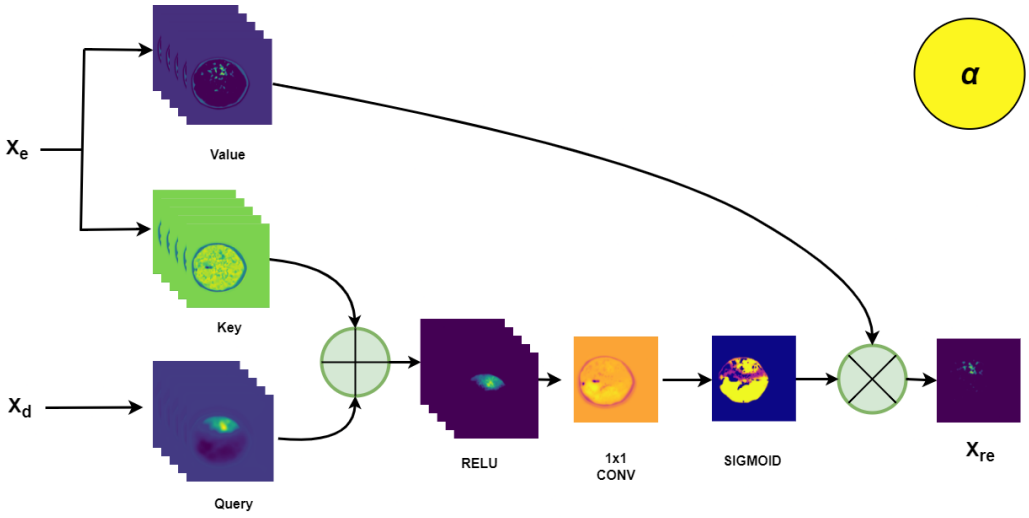



Los mecanismos de atención aditiva cruzada permiten regularizar y seleccionar las características del codificador ($X_e$) según la representación del decodificador ($X_d$) en un mismo nivel de procesamiento. Particularmente, el cálculo de similitud sigue una alineación aditiva entre las proyecciones lineales de codificador y decodificador, calculadas a partir de convoluciones. Posteriormente, se extraen las similitudes positivas por medio de una función de activación ReLU. Seguidamente, se lleva a cabo el cálculo de un único mapa de atención por medio de la reducción de dimensionalidad de las similitudes positivas usando una capa convolucional $1 \times 1$ y una función de activación Sigmoid. Por último, calculamos las características refinadas del codificador ($X_{re}$) por medio de una multiplicación elemento a elemento entre las características del codificador y el mapa de atención. El proceso explicado puede ser resumido con la fórmula:

$$X_{re} = X_e \cdot \text{Sigmoid}(W_{re}^T \text{ReLU}(W_e^T X_e + W_d^T X_d)).$$

El cálculo de las características refinadas del codificador es descrito por la Figura 14

**Figura 14.** Esquema de atención cruzada. De la entrada del codificador se obtiene el key y value mientras que el query viene del decodificador. El key y query se unen en un proceso aditivo y posteriormente el value se une por un proceso multiplicativo.




## 5.2. Refinamiento multinivel

Para mejorar las predicciones de la U-Net enriquecida con mecanismos de atención implementamos un entrenamiento de refinamiento multinivel (RM). Este proceso incluye varias medidas para minimizar la propagación de errores y mejorar la calidad de la segmentación. En primer lugar, se llevó a cabo una supervisión profunda de la representación del decodificador para minimizar la propagación de errores y evitar el desvanecimiento del gradiente [25].

[25] Renjie Li et al. "A Comprehensive Review on Deep Supervision: Theories and Applications". En: *arXiv preprint arXiv:2207.02376* (2022).

Esta supervisión profunda consiste en aplicar una interpolación bilineal a las características de salida de cada nivel del decodificador, seguida de una capa de clasificación para realizar la estimación en baja escala de la lesión. Estas estimaciones permiten calcular funciones de pérdida complementarias que miden la calidad de la estimación en cada nivel. Además, para abordar el fuerte desequilibrio de clases y dar más importancia a los píxeles límite de la lesión, se añadieron mapas de peso de clase (**C**) al cálculo de la función de pérdida final. Estos mapas de peso de clase se construyen a partir de las delineaciones manuales de las lesiones isquémicas como referencia y asignan un peso específico a los píxeles de la misma clase para resaltar su importancia. Consecuentemente, se utilizan las estimaciones de lesión junto con sus correspondientes mapas de peso de clase para calcular la señal de pérdida final de la siguiente manera: $\mathcal{L} = \mathbf{WCY}\log(\hat{\mathbf{Y}})$ donde $\{\mathbf{Y}, \hat{\mathbf{Y}}, \mathbf{C}\} \in \mathbb{R}^{O\times H\times W}$, **Y** corresponde a la salida esperada, $\hat{\mathbf{Y}}$ es la salida obtenida y los pesos del output model son $\mathbf{W} \in \mathbb{R}^{|W|}$.

### 5.3. Primera fase de entrenamiento: Generación de mapas sintéticos

La arquitectura propuesta que incluye mecanismos de atención es entrenada y ajustada en dos fases seriales. En esta fase, uno de los principales propósitos en esta fase de entrenamiento es obtener mapas paramétricos sintéticos, capturados a partir de un entrenamiento supervisado de segmentación. Es así, que la arquitectura U-net con atención definida en este trabajo es entrenada para generar máscaras de segmentación para lesiones de ACV y ajustar los mapas de atención de la representación profunda. Estos mapas sintéticos pueden ser claves para codificar relaciones texturales que más aportan en las lesiones de ACV. En este caso, podrían funcionar como mapas paramétricos complementarios que mejoren la estimación de la lesión, junto con los mapas resultantes del proceso de perfusión. Además en escenarios sin perfusión, estos estudios pueden complementar los estudios de CT.
Para ello, en la entrada de la arquitectura se dispuso de un CT ($\mathbf{I}(\mathbf{x})$) sin contraste y cuatro mapas paramétricos ($\mathbf{M_p}(\mathbf{x})$, conformado por CBV, CBF, MTT, TMAX). Estas secuencias son concatenadas en la entrada como múltiples canales de un tensor de entrada ($[\mathbf{I}(\mathbf{x}); \mathbf{M_p}(\mathbf{x})]$).

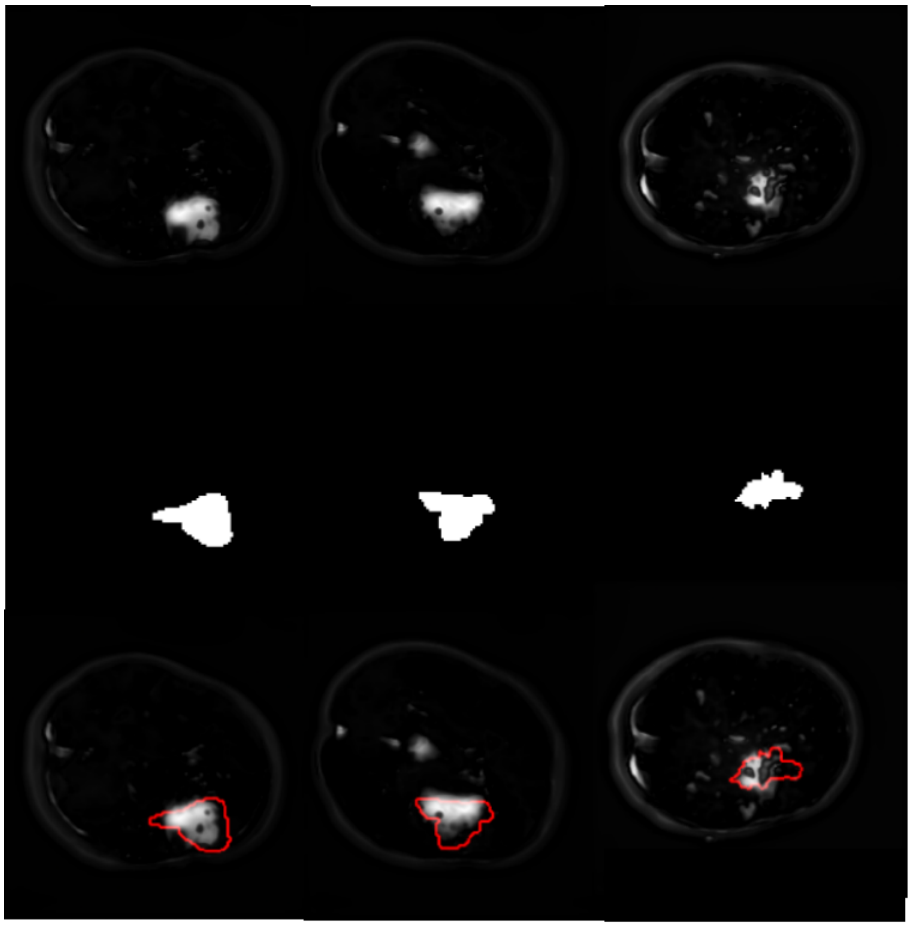
La representación inicial permite mantener una correlación espacial entre los diferentes canales, y las respuestas asociadas con la lesión. Entonces, esta entrada es convolucionada en el codificador hasta llegar a un espacio latente compacto que mantiene las relaciones con mayor asociación con la lesión entre los mapas paramétricos. Luego, desde este espacio embebido (**h**) es decodificada la información, siguiendo un re-escalamiento de esta y generando información entre las escalas, reforzado por los mecanismos de atención. Estos mecanismos de atención cruzada aditiva permiten codificar las mejores relaciones no locales de las observaciones, sirviendo para forzar la segmentación de la lesión entre diferentes escalas, pero también permitiendo guardar un mapa sintético ($[\mathbf{M_s}(\mathbf{x})]$). El mapa sintético es la respuesta entre el nivel anterior del decodificador y la entrada del codificador.
Al entrenarse el esquema propuesto se obtienen segmentaciones a baja escala, con limitada descripción local, pero capturando información intermedia que puede ser utilizada como entrada para forzar un nuevo entrenamiento de la arquitectura. Esta arquitectura es representada en la figura 13. Cabe resaltar que esta red puede ser apropiada por sí sola para recuperar lesiones que tienen mejor definición textural. Sin embargo, su principal objetivo es codificar respuestas de atención que puedan complementar un re-entrenamiento en una segunda fase de ajuste.

### 5.4. Segunda fase de entrenamiento: refinamiento de la lesión ACV

En la segunda etapa, una segunda arquitectura, idéntica en composición a la entrenada en la primera fase, es ajustada a partir de una nueva representación de entrada. Esta representación incluye mapas sintéticos de atención, calculados a partir del módulo de atención aditiva cruzada de la última capa de la arquitectura entrenada en la primera etapa. En otras palabras, la entrada presente en la segunda etapa de la red se compone por una unión entre los mapas paramétricos presentes en la primera fase con los nuevos mapas sintéticos generados en esta ($[\mathbf{I}(\mathbf{x}); \mathbf{M_p}(\mathbf{x}); \mathbf{M_s}(\mathbf{x})]$).
El mapa sintético de saliencia resalta patrones de la lesión isquémica, logrando una mejor

segmentación de las lesiones isquémicas sobre secuencias multiparamétricas de CT. Esta nueva imagen es la que será añadida en la entrada de la red durante esta segunda fase. Al representar de manera clara las zonas de importancia, al igual que una posible indicación de la forma de la lesión, se plantea que estos pueden guiar a la red a la hora de segmentar si son introducidos desde el inicio de esta. Un ejemplo de estos mapas puede verse en la figura 15

**Figura 15.** En orden vertical, mapa extraído de la primera fase, lesión delineada por un profesional y comparación entre las imágenes



Ya que se posee una arquitectura similar a la de la primera fase, se plantea el uso de los mismos mecanismos de atención de la primera fase en la segunda. Ahora, para poder rea-

lizar este proceso se necesita determinar la configuración inicial de las fases, para ello se tiene como entrada la imagen CT más un subconjunto de mapas paramétricos, que son acompañados además por los nuevos mapas de atención resultantes de la primera fase de entrenamiento. Este mapa extraído tendrá una serie de filtros de profundidad al provenir de la mitad del mecanismo de atención, por lo que se obtiene la media de los píxeles de cada capa de esta imagen para obtener una sola imagen que represente el contenido de todas ellas. Esta imagen es la que es añadida en la entrada de la red de la segunda fase.
Como resultado de esta segunda fase, se esperan tener mapas de segmentación ($\mathbf{s_2}(\mathbf{x})$) refinados, los cuales capturan detalles de la geometría de la lesión y puede ser clave para el diagnóstico y pronóstico del paciente.

## 5.5. Configuración experimental

**5.5.1. Conjuntos de datos de CT** La metodología propuesta se validó sobre el conjunto de datos *Ischemic Stroke Lesion Segmentation* de 2018 (*ISLES2018*). Este conjunto de datos contiene 156 estudios de pacientes diagnosticados con accidente cerebrovascular isquémico agudo. El conjunto de datos viene dividido en 94 estudios para entrenamiento y 62 estudios para prueba. Cada estudio contiene una imagen de CT sin contraste, cuatro mapas de perfusión de CT (CBF, CBV, MTT, Tmax) y el estudio de perfusión (CTP). Las delineaciones de las lesiones isquémicas fueron llevadas a cabo manualmente por radiólogos expertos y sólo estaban disponibles para los estudios de entrenamiento. Por último, se excluyeron aleatoriamente 19 estudios de la partición de entrenamiento para la validación.
Previo al entrenamiento de las arquitecturas, se llevó a cabo un preprocesamiento que consistió en normalizar las intensidades de los voxeles entre 0 y 1. Posteriormente, se extrajeron cortes axiales de las secuencias CT, los cuales fueron redimensionados a 224x224 píxeles. De esto se crean varios conjuntos de datos con los que experimentar:

- CT original

- Mapas paramétricos individuales
- Grupos de mapas paramétricos. En este caso, la experimentación consiste en tomar subconjuntos de parejas de los tres mapas con mejores resultados de segmentación. A esto se le añaden pruebas combinando los tres mapas y también entrenamientos usando todos los mapas.

**5.5.2. Especificaciones de la red** La arquitectura implementa un backbone convolucional que une bloques con una capa convolucional, batch normalization y una activación ReLU, dos veces por bloque. La red consta de seis capas, en las cuales al final se incluye un bloque convolucional seguido de un max pooling con factor de 2 en el codificador. En el decodificador se incluyeron en su lugar capas upsampling con factor de 2 seguidos de una interpolación bilineal. En la primera capa del codificador cada capa tiene 32 filtros de profundidad, duplicándose en tamaño hasta tener 1024 en la última capa. Los tamaños son los mismos en el decodificador, empezando desde 1024 capas hasta bajar a 32 y finalmente una salida con una sola capa.

Para las redes que entrenan con los mapas paramétricos se utilizó un batch size de 16 y se modificó el tamaño de las imágenes a 224x224 para todos los casos. El entrenamiento se realizó por 600 epochs con learning rate de 3e-2 y función de pérdida focal (con pesos de 0.7 para la lesión y 0.3 para el fondo) para enfrentar el problema del desbalance de datos. Como función de pérdida se utiliza binary crossentropy y como función de optimización se usa AdamW.

Los mecanismos de atención utilizados reciben de entrada imágenes con tamaño 224x224x64 que fueron integradas utilizadas con una función aditiva.

## 6. EVALUACION Y RESULTADOS

En este trabajo se realizó un estudio exhaustivo de los componentes del enfoque propuesto y las entradas CT multiparamétricas para medir su aporte en el desempeño de los modelos de segmentación. En primer lugar, para realizar este estudio de componentes, se tuvieron en cuenta las siguientes variaciones del enfoque propuesto: 1) U-Net estándar (U-Net), 2) U-Net con entrenamiento de refinamiento multinivel (U-Net + RM) y U-Net con entrenamiento de refinamiento multinivel y mecanismos de atención aditiva cruzada (U-Net + RM + Att).

En un primer experimento se midió el desempeño de modelos unimodales para explotar la información presente en cada una de las modalidades en el conjunto de datos. En la Tabla 1 se muestran los Dice Score para todas las configuraciones definidas con respecto a cada una de las modalidades presentes en el conjunto de datos. La mejor configuración resultó de utilizar el mapa paramétrico CBF y el modelo U-Net con entrenamiento de refinamiento multinivel (0.58). Es de resaltar que para cada modalidad en el conjunto de datos, se obtiene un mejor desempeño con los componentes del enfoque propuesto. Más específicamente, se observa una mejora en el dice score de 0.026 y 0.022 para las configuraciones *RM* y *RM + ATT* con respecto a la U-Net estándar. Cabe destacar que la diferencia entre los esquemas RM y RM + ATT no tienen una diferencia significativa. También se observa una gran diferencia entre el desempeño mostrado por los modelos que utilizan NCCT y los que utilizan los mapas paramétricos derivados de la perfusión, siendo los NCCT los de menor capacidad. Este hecho puede estar asociado a la baja sensibilidad del NCCT para identificar lesiones de ACV isquémico con pocas horas de evolución. Por otro lado, las variables temporales y volumétricas, modeladas por los mapas paramétricos resultan ser útiles para la caracterización de las lesiones de ACV. Este hecho resulta clave para nuestra investigación, en la cual se propende por definir mapas alternativos que permitan soportar las observaciones de stroke y complementen fuertemente los estudios basados únicamente en NCCT.

En la figura 16 se ilustran algunos ejemplos de anotaciones obtenidas, de acuerdo a las con-

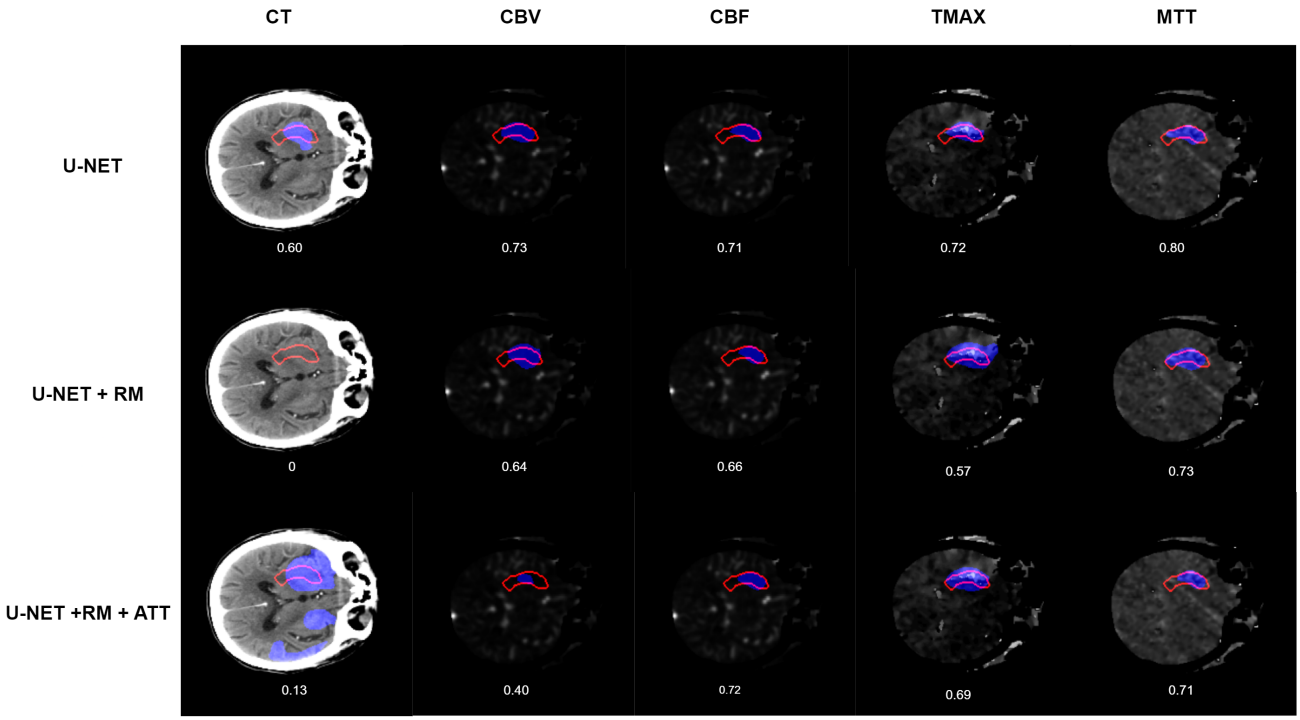
**Cuadro 1.** Métrica Dice score para distintas configuraciones del método propuesto con cada una de las modalidades presentes en el conjunto de datos ISLES2018

| Configuración modelo | | Modalidades | | | | |
|---|---|---|---|---|---|---|
| RM | Att | NCCT | CBV | CBF | Tmax | MTT |
| ✗ | ✗ | 0.20 ± 0.26 | 0.51± 0.30 | 0.57 ± 0.26 | 0.50 ± 0.25 | 0.56 ± 0.30 |
| ✓ | ✗ | **0.26 ± 0.24** | 0.51 ± 0.27 | **0.58 ± 0.22** | **0.55 ± 0.26** | **0.57 ± 0.26** |
| ✓ | ✓ | 0.24 ± 0.19 | **0.54 ± 0.28** | 0.57 ± 0.23 | **0.55 ± 0.28** | 0.55 ± 0.21 |

figuraciones validadas en la tabla 1. En las columnas se exponen los resultados de acuerdo a cada modalidad considerada, mientras que en las filas se tiene en cuenta los diferentes componentes involucrados en la red, por ejemplo la Unet, Unet+Rm. Unet +Rm+att. Como se puede observar, los modelos logran un mejor sobrelapamiento con respecto a la anotación brindada por un experto radiólogo, cuando la entrada son los mapas paramétricos. Además existe un claro aporte cuando se utilizan mecanismos RM y de atención.

En un segundo experimento se midió la capacidad de explotar información multicontexto. En este caso, se realizó una validación exhaustiva con los componentes propuestos y múltiples combinaciones de modalidades CT multiparametricas. La combinación de las modalidades consiste en la concatenación de los modalidades como canales de una misma imagen. En la tabla 2 se muestran los resultados obtenidos para cada una de las combinaciones de entrada y configuración del modelo. Como se esperaba, la capacidad de segmentar lesiones isquémicas de los modelos que incluyen información multicontexto, es superior a la de los modelos unimodales. El mejor desempeño (0.63) es alcanzado por dos configuraciones, la que utiliza todos los mapas paramétricos y la que utiliza la combinación de Tmax, CBV y CBF. Pero al analizar los datos, se puede observar que la media obtenida por la configuración que incluye todos los mapas paramétricos es mayor a la combinación que utiliza solo tres de estos. Adicionalmente a esto, se observa una diferencia de 0.08 entre el mejor modelo unimodal y el modelo que incluye todas las modalidades. Además, al igual que en el experimento con una sola modalidad en la entrada, la capacidad de segmentar lesiones isquémicas aumenta cuando se incluyen los componentes propuestos. Más específicamente, se logra un aumento de 0.024 con las configuraciones *RM* y *RM + ATT* con respecto a la

**Figura 16.** Ejemplos de las segmentaciones realizadas para cada modalidad de la primera tabla. La línea roja representa el contorno de la lesión indicada por el profesional y la zona azul lo predicho por la red




U-Net estándar.

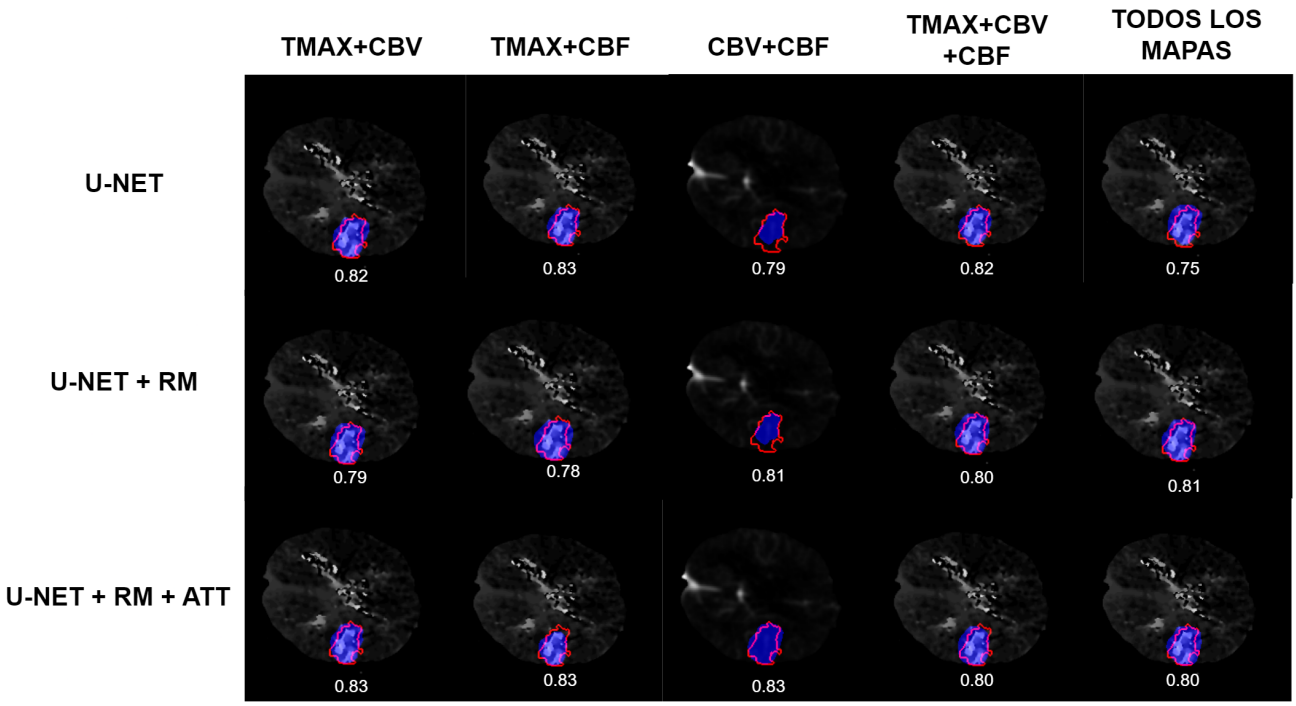
**Cuadro 2.** Métrica Dice score para distintas configuraciones del método propuesto con distintas combinaciones de las modalidades presentes en el conjunto de datos ISLES2018

| Configuración modelo | | Modalidades | | | | |
|---|---|---|---|---|---|---|
| RM | Att | TMax + CBV | TMax + CBF | CBV + CBF | TMax + CBV + CBF | Todas |
| ✗ | ✗ | 0.58 ± 0.25 | 0.58 ± 0.24 | 0.53 ± 0.26 | 0.60 ± 0.24 | 0.62 ± 0.23 |
| ✓ | ✗ | **0.61 ± 0.23** | **0.62 ± 0.20** | 0.54 ± 0.27 | **0.63 ± 0.21** | 0.63 ± 0.22 |
| ✓ | ✓ | 0.60 ± 0.23 | **0.62 ± 0.21** | **0.55 ± 0.25** | 0.61 ± 0.25 | **0.63 ± 0.16** |

Para complementar los resultados analizados en la tabla 2, en la figura 17 se ilustran algunos ejemplos de segmentaciones obtenidas con diferentes entradas multimodales. Como se ilustra, los mecanismos de RM y atención aportan a un mayor grado de delineación local, conservando estructuras particulares de la lesión y brindando un mayor sobrelapamiento

entre las máscaras. En este caso, la información multimodal brinda mejores aproximaciones para modelar el problema.

**Figura 17.** Ejemplos de las segmentaciones realizadas para cada modalidad de la segunda tabla. La línea roja representa el contorno de la lesión indicada por el profesional y la zona azul lo predicho por la red




Ya determinados los elementos que dan mejores resultados se entrenó una configuración del modelo que incluye la fase de refinamiento explicada en el enfoque propuesto. La segunda red posee las mismas características que la primera: mecanismos de atención aditiva y refinamiento multinivel.

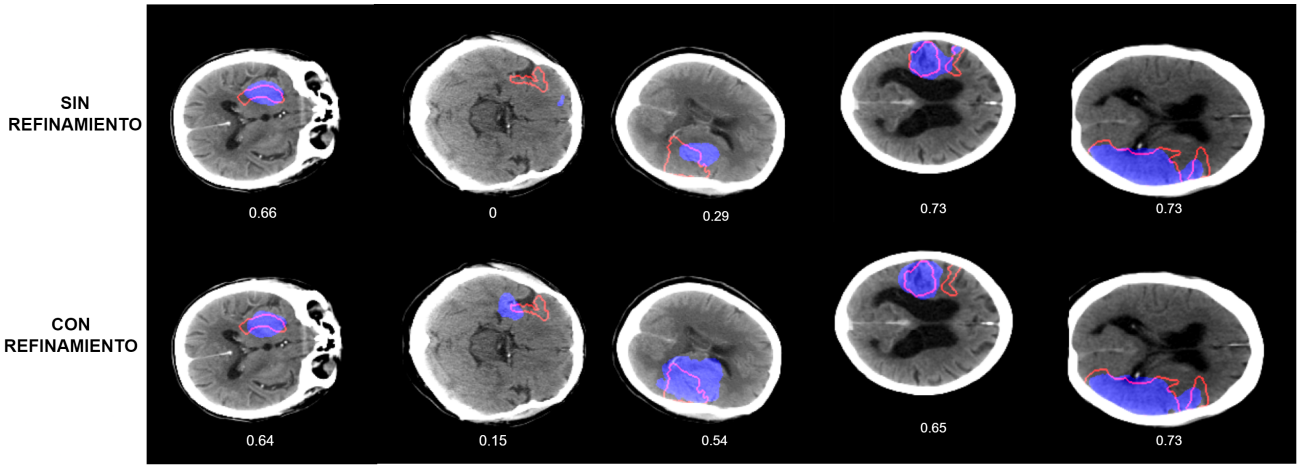
Los resultados obtenidos muestran una pérdida de sensibilidad al aplicar la fase de refinamiento al pasar de (0.73 ± 0.19) a (0.68 ± 0.12), pero también se obtuvo una significativa mejora en la precisión y el dice score obtenido. Sumado a esto se puede observar que la desviación estándar es menor en todas las métricas, lo cual significa que al aplicar el refinamiento se obtienen resultados más consistentes.

**Cuadro 3.** Métricas Dice score, precisión y sensibilidad para el mejor modelo con y sin fase de refinamiento en el conjunto de datos ISLES2018

| Fase de refinamiento | Dice Score | Precisión | Sensibilidad |
|---|---|---|---|
| ✗ | 0.63 ± 0.16 | 0.63 ± 0.21 | 0.73 ± 0.19 |
| ✓ | 0.66 ± 0.14 | 0.67 ± 0.19 | 0.68 ± 0.12 |

Para complementar los resultados cuantitativos de la tabla 3, en la figura 18 se ilustran algunos estudios con los resultados de incluir la red de refinamiento en la segunda fase de ajuste. Por ejemplo, para la segunda columna, se puede apreciar que una red sin refinamiento no logra en absoluto una aproximación de la lesión. Esto, por supuesto, tiene que ver con lo desafiante de la lesión (su estadio temprano), además del bajo contraste del estudio. Sin embargo, la fase de refinamiento, logra localizar la lesión, lo cual podría ser suficientemente interesante como soporte al análisis observacional realizado por el radiólogo. También se observa, que para otras lesiones, la arquitectura permite una aproximación equivalente en la etapa de refinamiento.

**Figura 18.** Ejemplos de las segmentaciones realizadas en ambas redes de la tercera tabla. Las de la izquierda corresponden a la red con solo la primera fase, las de la derecha corresponden a la red con fase de refinamiento. La linea roja representa el contorno de la lesión indicada por el profesional y la zona azul lo predicho por la red



# 7. CONCLUSIONES Y PERSPECTIVAS DEL TRABAJO

En este trabajo se planteó una metodología para segmentar lesiones de ACV desde secuencias NCCT. Para ello, se implementó una arquitectura de tipo codificador-decodificador (U-Net) que además integraba módulos de atención para preservar la estructura geométrica y resaltar los patrones de la lesión, durante la decodificación. Sumado a ello, durante el entrenamiento se siguió un esquema supervisado en múltiples niveles, que permitía monitorear la reconstrucción de la lesión en diferentes escalas de representación.

Como se evidenció en el trabajo, uno de los principales componentes para obtener una segmentación coherente son los mapas de perfusión. Así, emulando estos mapas, en este trabajo se reporta una metodología en dos etapas de entrenamiento, siendo la primera etapa dedicada a aprender mapas sintéticos de la lesión, a partir de la salida de los mecanismos de atención. Estos mapas son entonces usados en una segunda fase de entrenamiento como entrada y guía para la representación de la lesión, desde las secuencias de entrada. El método fue validado en un conjunto de datos públicos mostrando resultados competitivos en el contexto del problema de segmentación de lesiones de ACV.

Trabajos futuros incluyen la adaptación de esta arquitectura en estudios locales, recopilados en el grupo de investigación y que pueden ser más desafiantes teniendo en cuenta las características de las máquinas de adquisición de secuencias de imágenes, así como de la ausencia de los mapas de perfusión. Así, se podrían plantear aproximaciones donde los mapas sintéticos puedan ser usados como una aproximación para acompañar estos estudios que no cuentan con perfusión. También se espera seguir estudiando nuevos mecanismos de atención y entrenamiento supervisado a diferentes escalas, que pueda impactar de mejor manera la caracterización de la lesión.

## BIBLIOGRAFÍA


Albers, Gregory W et al. “Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging”. En: *New England Journal of Medicine* 378.8 (2018), págs. 708-718 (vid. pág. 6).

Balakrishnan, Guha et al. “Voxelmorph: a learning framework for deformable medical image registration”. En: *IEEE transactions on medical imaging* 38.8 (2019), págs. 1788-1800 (vid. pág. 22).

Calamante, Fernando et al. “Measuring cerebral blood flow using magnetic resonance imaging techniques”. En: *Journal of cerebral blood flow & metabolism* 19.7 (1999), págs. 701-735 (vid. pág. 9).

Calamante, Fernando and Christensen, Søren and Desmond, Patricia M and Østergaard, Leif and Davis, Stephen M and Connelly, Alan. “The physiological significance of the time-to-maximum (Tmax) parameter in perfusion MRI”. En: *Stroke* 41.6 (2010), págs. 1169-1174 (vid. pág. 12).

Carroll, Timothy J and Horowitz, Sandra and Shin, Wanyong and Mouannes, Jessy and Sawlani, Rahul and Ali, Saad and Raizer, Jeffrey and Futterer, Stephen. “Quantification of cerebral perfusion using the “bookend technique”: an evaluation in CNS tumors”. En: *Magnetic resonance imaging* 26.10 (2008), págs. 1352-1359 (vid. págs. 10, 11).

Clèrigues, Albert et al. “Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks”. En: *Computers in biology and medicine* 115 (2019), pág. 103487 (vid. pág. 19).

Dolz, Jose, Ismail Ben Ayed y Christian Desrosiers. “Dense multi-path U-Net for ischemic stroke lesion segmentation in multiple image modalities”. En: *International MICCAI Brainlesion Workshop*. Springer. 2018, págs. 271-282 (vid. pág. 18).

French, Brandi R, Raja S Boddepalli y Raghav Govindarajan. “Acute ischemic stroke: current status and future directions”. En: *Missouri medicine* 113.6 (2016), pág. 480 (vid. pág. 6).

Gómez, Santiago and Florez, Sebastian and Mantilla, Daniel and Camacho, Paul and Tarazona, Nick and Martínez, Fabio. “An attentional unet with an auxiliary class learning to support acute ischemic stroke segmentation on CT”. En: *Medical Imaging 2023: Image Processing*. Ed. por Olivier Colliot e Ivana Išgum. Vol. 12464. International Society for Optics y Photonics. SPIE, 2023, 124640S. DOI: 10.1117/12.2654269 (vid. pág. 25).

Hoeffner, Ellen G et al. “Cerebral perfusion CT: technique and clinical applications”. En: *Radiology* 231.3 (2004), págs. 632-644 (vid. pág. 11).

Kuang, Hulin et al. “EIS-Net: Segmenting early infarct and scoring ASPECTS simultaneously on non-contrast CT of patients with acute ischemic stroke”. En: *Medical Image Analysis* 70 (2021), pág. 101984 (vid. pág. 20).

Li, Renjie et al. “A Comprehensive Review on Deep Supervision: Theories and Applications”. En: *arXiv preprint arXiv:2207.02376* (2022) (vid. pág. 28).

Liu, Liangliang et al. “Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease”. En: *Medical Image Analysis* 65 (2020). DOI: 10.1016/j.media.2020.101791 (vid. pág. 4).

Liu, Pengbo. “Stroke lesion segmentation with 2D novel CNN pipeline and novel loss function”. En: *International MICCAI Brainlesion Workshop*. Springer. 2018, págs. 253-262 (vid. pág. 20).

Liu, Pengbo. “Stroke lesion segmentation with 2D novel CNN pipeline and novel loss function”. En: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11383 LNCS (2019), págs. 253-262. DOI: 10.1007/978-3-030-11723-8_25 (vid. pág. 4).

Mair, Grant et al. “Sensitivity and specificity of the hyperdense artery sign for arterial obstruction in acute ischemic stroke”. En: *Stroke* 46.1 (2015), págs. 102-107 (vid. pág. 3).

Martel, Anne L. et al. “Measurement of infarct volume in stroke patients using adaptive segmentation of diffusion weighted MR images”. En: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 1999. DOI: 10.1007/10704282_3 (vid. pág. 3).

Neumann, Anders B. et al. “Interrater agreement for final infarct mri lesion delineation”. En: *Stroke* 40.12 (2009), págs. 3768-3771. DOI: 10.1161/STROKEAHA.108.545368 (vid. pág. 3).

Rekik, Islem et al. “Medical image analysis methods in MR/CT-imaged acute-subacute ischemic stroke lesion: Segmentation, prediction and insights into dynamic evolution simulation models. A critical appraisal”. En: *NeuroImage: Clinical* 1.1 (2012), págs. 164-178 (vid. pág. 3).

Ronneberger, Olaf, Philipp Fischer y Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. En: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, págs. 234-241 (vid. pág. 26).

Roth, Gregory A et al. “Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study”. En: *Journal of the American College of Cardiology* 76.25 (2020), págs. 2982-3021 (vid. pág. 3).

Tureckova, Alzbeta y Antonio J Rodríguez-Sánchez. “ISLES challenge: U-shaped convolution neural network with dilated convolution for 3D stroke lesion segmentation”. En: *International MICCAI Brainlesion Workshop*. Springer. 2018, págs. 319-327 (vid. pág. 18).

Vaswani, Ashish et al. “Attention is all you need”. En: *Advances in neural information processing systems* 30 (2017) (vid. págs. 15, 16).

Wang, Guotai et al. “Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks”. En: *Medical Image Analysis* 65 (2020), pág. 101787. DOI: 10.1016/j.media.2020.101787. arXiv: 2007.03294 (vid. pág. 4).

— “Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks”. En: *Medical Image Analysis* 65 (2020), pág. 101787 (vid. pág. 21).