

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ingeniería en Electricidad y Computación

Segmentación automática de la Sustancia Negra cerebral y Núcleo Subtalámico a partir de imágenes 3D de resonancia magnética

PROYECTO INTEGRADOR

Previo la obtención del Título de:

Ingeniero en Ciencias de la Computación

Presentado por:

Franklin Enmanuel Magallanes Pinargote

Juan Francisco Nebel Dunn

GUAYAQUIL - ECUADOR

Año: 2022

DEDICATORIA

Juan Fco. Nebel Dunn

El presente proyecto lo dedico a mi padre Francisco Antonio Nebel Bueno. Gracias por el infinito apoyo que me diste durante el poco tiempo que te tuve.

AGRADECIMIENTOS

*Franklin Enmanuel Magallanes
Pinargote*

Agradezco principalmente a mi familia, especialmente a mi mamá por todo el apoyo brindado para poder lograr este nuevo hito académico. También un enorme agradecimiento a todos los docentes, que estuvieron presentes durante mi carrera universitaria.

Juan Fco. Nebel Dunn

Agradezco a mi familia, quienes estuvieron siempre pendientes de mi desarrollo académico; a mis amigos, quienes durante años se mantuvieron a mi lado por más que yo no siempre pude disfrutar de muchas ocasiones junto a ellos y a mis profesores quienes fueron una guía importante para formarme como profesional.

DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, nos corresponden conforme al reglamento de propiedad intelectual de la institución; Franklin Enmanuel Magallanes Pinargote y Juan Francisco Nebel Dunn damos nuestro consentimiento para que la ESPOC realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"



**Franklin Enmanuel
Magallanes Pinargote**



**Juan Francisco
Nebel Dunn**

EVALUADORES

Boris Xavier Vintimilla Burgos Ph.D.

PROFESOR DE LA MATERIA

Colón Enrique Peláez Jarrín Ph.D.

PROFESOR TUTOR

RESUMEN

La Sustancia Negra y Núcleo Subtalámico son componentes vitales para el tratamiento de la Enfermedad de Parkinson (EP); no obstante, son difíciles de identificar en imágenes de resonancia magnética (IRMs). La Sustancia Negra cerebral permite llevar un control del daño que ha ocasionado la EP. El Núcleo Subtalámico es central en tratamientos como ablaciones, Estimulación Cerebral Profunda (DBS) y Ultrasonido Focal de Alta Intensidad (HIFU) los cuales requieren de una localización precisa de este componente. Presentamos una herramienta que permite segmentar de forma autónoma la Sustancia Negra y el Núcleo Subtalámico, a partir del análisis de imágenes de resonancia magnética y la aplicación de técnicas de aprendizaje profundo. El módulo de aprendizaje profundo está fundamentado en la arquitectura transformers SETR (*Segmentation Transformer*) y fue entrenado a partir de un conjunto de datos con 3000 ejemplos de imágenes potenciadas en T2 con cortes a 1mm de distancia. El modelo recibe una IRM 3D y devuelve otra con el Núcleo Subtalámico y Sustancia Negra segmentada y co-registrada a la imagen original. La propuesta basada en transformers (SETR) obtuvo la segmentación de los volúmenes de interés con un coeficiente DICE del 0,81 y AVD (*Average Volume Distance*) del 0,06. Se contrastó estos resultados con otras arquitecturas, tales como CLCI (*Cross-Level fusion and Context Inference Network*) y U-Net; sin embargo, estas arquitecturas no obtuvieron un desempeño a nivel de la arquitectura basada en SETR. El modelo basado en SETR brinda los mejores resultados y se acopla mejor a la tarea de segmentación del Núcleo Subtalámico y Sustancia Negra utilizando IRMs en 3D. La personalización de la arquitectura base influyó de manera significativa en las métricas de AVD, F1 Score y coeficiente DICE. Se obtuvo una buena calidad en las segmentaciones en 3D, a pesar de utilizar IRMs entre 1.5 y 3T.

Palabras Clave: Segmentación 3D, Núcleo Subtalámico, Sustancia Negra, Parkinson, Aprendizaje Profundo

ABSTRACT

The Substantia Nigra and Subthalamic Nucleus are vital components in the treatment of Parkinson's Disease (PD); however, they are difficult to identify on magnetic resonance imaging (MRI). The Substantia Nigra allows to keep track of the damage that has been caused by Parkinson's disease (PD). The Subthalamic Nucleus is central in treatments such as ablations, Deep Brain Stimulation (DBS) and High Intensity Focal Ultrasound (HIFU) which require precise localization of this component. We present a tool that allows autonomous segmentation of the Substantia Nigra and the Subthalamic Nucleus, based on the analysis of magnetic resonance images and the application of deep learning techniques. The deep learning module is based on the SETR (Segmentation Transformer) transformer architecture and was trained with a dataset with 3000 examples of T2-weighted images with 1-mm distance slices. The model receives a 3D MRI and returns another one with the Subthalamic Nucleus and Substantia Nigra segmented and co-registered to the original image. The transformer-based approach (SETR) obtained the segmentation of the volumes of interest with a DICE coefficient of 0.81 and AVD (Average Volume Distance) of 0.06. These results were contrasted with other architectures, such as CLCI (Cross-Level fusion and Context Inference Network) and U-Net; however, these architectures did not perform at the level of the SETR-based architecture. The SETR-based model provides the best results and is best suited to the task of segmenting the Subthalamic Nucleus and Substantia Nigra using 3D MRIs. Customization of the base architecture significantly influenced AVD metrics, F1 Score and DICE coefficient. Good quality segmentations were obtained despite using MRIs between 1.5 and 3T.

Keywords: 3D Segmentation, Subthalamic Nucleus, Substantia Nigra, Parkinson's Disease, Deep Learning

ÍNDICE GENERAL

RESUMEN	i
ABSTRACT	ii
ABREVIATURAS	v
ÍNDICE DE FIGURAS	v
ÍNDICE DE TABLAS.	vii
CAPÍTULO 1	1
1 INTRODUCCIÓN	1
1.1 Descripción del problema	2
1.2 Justificación del problema	2
1.3 Objetivos	2
1.3.1 Objetivo general	2
1.3.2 Objetivos específicos	3
1.4 Resultados esperados	3
1.5 Beneficiarios	3
1.6 Impacto comercial	4
1.7 Marco teórico	4
1.7.1 Enfermedad de Parkinson	5
1.7.2 Sustancia Negra	5
1.7.3 Núcleo Subtalámico	6
1.7.4 Resonancia magnética	8
1.7.5 Técnicas de preprocesamiento	8
1.7.5.1 Corrección del sesgo	9
1.7.5.2 Normalización de la intensidad	9
1.7.5.3 Registro de la imagen	10
1.8 Datos	12
1.8.1 Segmentación de la imagen	13

1.9	Aprendizaje profundo para segmentación	13
1.9.1	Modelos basados en convoluciones	13
1.9.1.1	Modelo U-Net	14
1.9.1.2	Modelo CLCI-Net	15
1.9.2	Modelo basado en <i>vision transformer</i>	16
1.9.2.1	Modelo SETR	17
CAPÍTULO 2	20
2	METODOLOGÍA	20
2.1	Etapas de la metodología	20
2.1.1	Preprocesamiento del conjunto de datos	22
2.1.2	<i>Data augmentation</i>	22
2.1.3	Adaptación de los modelos de base	23
2.1.4	Entrenamiento	23
2.1.5	Evaluación de métricas y calibración de parámetros	23
CAPÍTULO 3	26
3	RESULTADOS Y ANÁLISIS	26
3.1	Modelos de base	26
3.2	Modelos mejorados	29
CAPÍTULO 4	37
4	CONCLUSIONES Y TRABAJOS FUTUROS	37
4.1	Conclusiones	37
4.2	Recomendaciones	38
4.3	Trabajos Futuros	39
BIBLIOGRAFÍA	40

ABREVIATURAS

ESPOL	Escuela Superior Politécnica del Litoral
SNpc	Sustancia Negra Pars Compacta
SNpr	Sustancia Negra Pars Reticulata
GABA	Ácido Gamma-aminobutírico
IRM	Imágenes de resonancia magnética
FIMCP	Facultada de Ingeniería en Mecánica y Ciencias de la Producción
LSTM	Long Short Term Memory
ConvLSTM	Convolutional Long Short Term Memory

ÍNDICE DE FIGURAS

1.1	Sustancia Negra en relación al cerebro (plano coronal) [1]	6
1.2	Anatomía del cerebro, ganglios basales [2]	7
1.3	Ejemplos diversos de resonancia magnética [3]	8
1.4	Comparativa de IRM original vs IRM con corrección de sesgo [autoría propia]	9
1.5	Comparativa de IRM original vs IRM con normalización de intensidad [autoría propia]	10
1.6	Comparativa de IRM original vs IRM con transformación [autoría propia]	11
1.7	Transformaciones espaciales en imágenes [autoría propia]	11
1.8	Segmentación en resonancia del cerebro [4]	13
1.9	Proceso de convolución en una imagen [autoría propia]	14
1.10	Diagrama de la arquitectura U-Net [autoría propia]	15
1.11	Diagrama de la arquitectura CLCI-Net [5]	16
1.12	Diagrama de la arquitectura <i>vision transformer</i> [autoría propia]	17
1.13	Diagrama de la arquitectura SETR. (a) PUP (b) MLA [autoría propia]	19
2.1	Distribución de las coordenadas de las áreas de interés [autoría propia]	21
2.2	Flujo de datos del estudio [autoría propia]	22
3.1	Resultado de métricas del <i>dataset</i> de entrenamiento en los modelos base [autoría propia]	27
3.2	Resultado de métricas del <i>dataset</i> de validación en los modelos base [autoría propia]	28
3.3	Arquitectura mejorada para el modelo basado en transformer [autoría propia]	30
3.4	Resultado de métricas del <i>dataset</i> de entrenamiento de los modelos SETR [autoría propia]	32

3.5 Segmentación resultante de modelo basado en CLCI-Net [autoría propia]	35
3.6 Segmentación resultante de modelo basado en SETR 1 [autoría propia]	35
3.7 Segmentación resultante de modelo basado en SETR 2 [autoría propia]	36

ÍNDICE DE TABLAS

3.1	Resultados de las métricas para el <i>dataset</i> de pruebas	33
-----	--	----

CAPÍTULO 1

1. INTRODUCCIÓN

No es extraño que se encuentren contribuciones valiosas dentro de la interseccionalidad de las disciplinas. Dentro de la medicina existen desafíos complejos cuya falta de soluciones desembocan en consecuencias dolorosamente reales. Estos desafíos presentan problemas novedosos que otras disciplinas pueden ayudar a sobrellevar con éxito.

Uno de los órganos más sofisticados y elusivos es el cerebro humano. Este corresponde a un sistema incalculablemente complejo que permite, desde las funciones más básicas, hasta las expresiones más complejas; expresiones que incluso coquetean con dilemas filosóficos que ponderan lo que significa estar vivo y tener conciencia.

Desafortunadamente, las neuropatologías tienden a tener tratamientos complicados e intervenciones particularmente delicadas [6]. El mal de Parkinson cae dentro de esta categoría y éste afecta a componentes del cerebro que regulan el movimiento de un individuo, lo cual desemboca en el clásico movimiento involuntario que caracteriza a la enfermedad [7].

Las intervenciones que logran reducir los síntomas del mal de Parkinson requieren una precisa localización de estos componentes reguladores de movimiento [6]. Es aquí donde este proyecto busca utilizar técnicas y herramientas del estado del arte dentro de la Inteligencia Artificial y evaluar su rendimiento para asistir en el proceso de identificación de estos componentes del cerebro, a través de la segmentación de imágenes.

1.1 Descripción del problema

La enfermedad de Parkinson (EP) es un trastorno progresivo que afecta el movimiento, el control muscular y el equilibrio [7]. Parte del proceso patológico se desarrolla a medida que las células se van destruyendo progresivamente en ciertas partes del cerebro, particularmente la masa de células en media luna, conocida como Sustancia Negra [7], la cual es un núcleo de los ganglios basales que adquiere su nombre por sus neuronas dopaminérgicas pigmentadas debido a la neuromelanina que las recubre [1].

Uno de los principales problemas que enfrentan los neurocirujanos es justamente la identificación de la Sustancia Negra y el Núcleo Subtalámico a partir de imágenes médicas, esto por una multitud de razones, unas cuantas incluyen la poca diferencia de contraste respecto a las regiones aledañas, la alta cantidad de hierro en la zona (la cual afecta la imagen resultante) y una alta sensibilidad al ruido que presentan principalmente los resonadores por debajo de 1.5 Tesla [8, 9].

1.2 Justificación del problema

La delimitación eficaz de la Sustancia Negra y el Núcleo Subtalámico implica una mayor precisión quirúrgica. Esto conlleva un mejor tratamiento de enfermedades como el párkinson; especialmente cuando se habla de la mejoría de síntomas motores en un paciente, ya que, estos elementos participan en iniciar y controlar sus movimientos finos [10].

1.3 Objetivos

1.3.1 Objetivo general

Construir una herramienta que permita la identificación y delimitación de la Sustancia Negra y Núcleo Subtalámico a partir de la aplicación de técnicas de aprendizaje profundo en imágenes de resonancia magnética (IRM) para control de la progresión de la enfermedad de Parkinson y el tratamiento de sus síntomas.

1.3.2 Objetivos específicos

- Depurar imágenes de resonancia magnética (IRMs) a través de preprocesamiento y extracción de características.
- Diseñar modelos de aprendizaje autónomo para segmentación de la Sustancia Negra y el Núcleo Subtalámico.
- Implementar un prototipo de software.
- Evaluar el desempeño y comparar resultados entre varios modelos de aprendizaje autónomo basados en diferentes arquitecturas.

1.4 Resultados esperados

Una herramienta que permita ubicar a la Sustancia Negra y al Núcleo Subtalámico de forma autónoma, a partir del análisis de imágenes de resonancia magnética y la aplicación de técnicas de aprendizaje profundo.

Así mismo, esta herramienta debe sustentar su elección en base a la comparación de algoritmos de aprendizaje profundo supervisado y contar con acceso a través de una aplicación web.

1.5 Beneficiarios

Los beneficiarios resultantes de este proyecto son los siguientes:

- Doctores que tengan que realizar un procedimiento referente a males asociados a la Sustancia Negra y/o Núcleo Subtalámico.
- Organizaciones y emprendimientos cuyas operaciones dependen de una correcta segmentación de la Sustancia Negra y/o Núcleo Subtalámico.
- Investigadores que buscan explorar soluciones novedosas a problemas dentro de un dominio similar o con características compartidas.

1.6 Impacto comercial

Una delimitación eficaz de la Sustancia Negra y Núcleo Subtalámico facilita el cálculo de parámetros importantes como el volumen, perímetro, centro de masa, variación de intensidad y densidad. Estos son fundamentales para el control de la progresión de la enfermedad o para el éxito de intervenciones quirúrgicas que se consideren realizar al paciente [6].

Una herramienta como esta, disponible a través de una aplicación web, podría convertirse en una solución tecnológica de apoyo a radiólogos y neurocirujanos para la identificación de la Sustancia Negra y Núcleo Subtalámico.

Además, representaría un ahorro de recursos para la institución médica en aspectos de infraestructura y mitigación de errores emergentes durante la realización de resonancias magnéticas de secuencia T2, las cuales ofrecen una alta sensibilidad y contraste en las áreas del cerebro mencionadas.

1.7 Marco teórico

Este proyecto se encuentra en una intersección de disciplinas y comprende conocimiento epistémico de diferentes dominios. Tanto el origen del problema, como el contexto en donde una solución puede ser aplicada, se ven comprendidos por el dominio de la medicina.

Para este proyecto es relevante el área médica que genera conocimiento referente a las neuropatologías, en particular, el mal de Parkinson y los componentes del cerebro que este afecta; la Sustancia Negra y el Núcleo Subtalámico específicamente (mas no son los únicos componentes afectados por el mal de Parkinson).

Por otra parte, las técnicas y herramientas presentadas se encuentran fundamentadas en un trasfondo computacional, particularmente el área de Inteligencia Artificial, visión por computador y aprendizaje profundo.

1.7.1 Enfermedad de Parkinson

La enfermedad de Parkinson (EP) es una enfermedad compleja que no está comprendida en su totalidad [7]. Se conoce como una enfermedad neurodegenerativa progresiva, la cual afecta al sistema nervioso central, periférico y entérico. Entre los síntomas que experimentan pacientes que presentan dicho cuadro se pueden encontrar: espasmos musculares involuntarios de intensidad variante, rigidez muscular, problemas en mantener una postura estable, entre otros [11, 7].

La fisiopatología de la enfermedad vislumbra que esta afecta sustantivamente al conjunto de ganglios basales ubicados en el tallo central del cerebro. Los ganglios basales son responsables del correcto funcionamiento y ejecución de funciones como el movimiento muscular; de manera específica, la degeneración se concentra en subáreas de la Sustancia Negra donde, dependiendo de la progresión de la enfermedad, se puede apreciar daños extensivos [11, 7].

1.7.2 Sustancia Negra

La Sustancia Negra es parte de los ganglios basales. Está ubicada en el cerebro medio y está directamente conectada con el putamen. Juega un rol esencial en el control de movimientos motores, así como la regulación del sistema de recompensa del cerebro a través de la dopamina [1]. Su ubicación en relación al resto del cerebro puede ser observada en la Figura 1.1.

La Sustancia Negra deriva su nombre debido a la pigmentación oscura que esta posee en comparación a la Sustancia Blanca y la Sustancia Gris. Esta coloración es resultado de la neuromelanina presente en el recubrimiento de neuronas dopaminérgicas, las cuales presentan una alta densidad en esta área del cerebro [1].

La Sustancia Negra está dividida en 2 secciones principales: la Sustancia Negra Pars Compacta (SNpc) y la Sustancia Negra Pars Reticulata (SNpr). La SNpc es donde se encuentra la mayor cantidad de neuronas dopaminérgicas y por ende posee la coloración más oscura. Por otra parte, la SNpr es donde se encuentran la mayor cantidad de

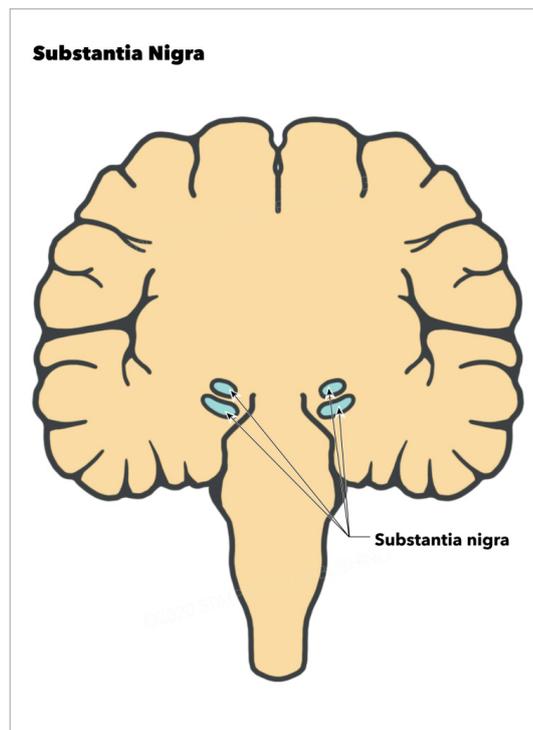


Figura 1.1: **Sustancia Negra en relación al cerebro (plano coronal) [1]**

neuronas GABAérgicas, las cuales obtienen su nombre por el neurotransmisor químico que utilizan, el ácido gamma-aminobutírico (GABA) [1].

La pérdida de Sustancia Negra está estrechamente relacionada con patologías cerebrales importantes. El enfoque de este proyecto se orienta a la enfermedad de Parkinson. Los pacientes de Parkinson pierden células en esta sección del cerebro, la cual, en consecuencia, pierde su coloración distintiva y empieza a difuminarse con sus alrededores, reduciendo su volumen [1].

1.7.3 Núcleo Subtalámico

El Núcleo Subtalámico es un componente pequeño con forma de lente ubicado en la unión entre el cerebro medio y el diencefalo. Naturalmente, se lo considera como parte del subtálamo y su función principal recae en regular el movimiento [12].

El Núcleo Subtalámico contiene una cantidad significativa de neuronas glutaminérgicas, las cuales cumplen un papel de neurotransmisores excitatorios para el sistema nervioso central. El Núcleo Subtalámico está conectado a la Sustancia

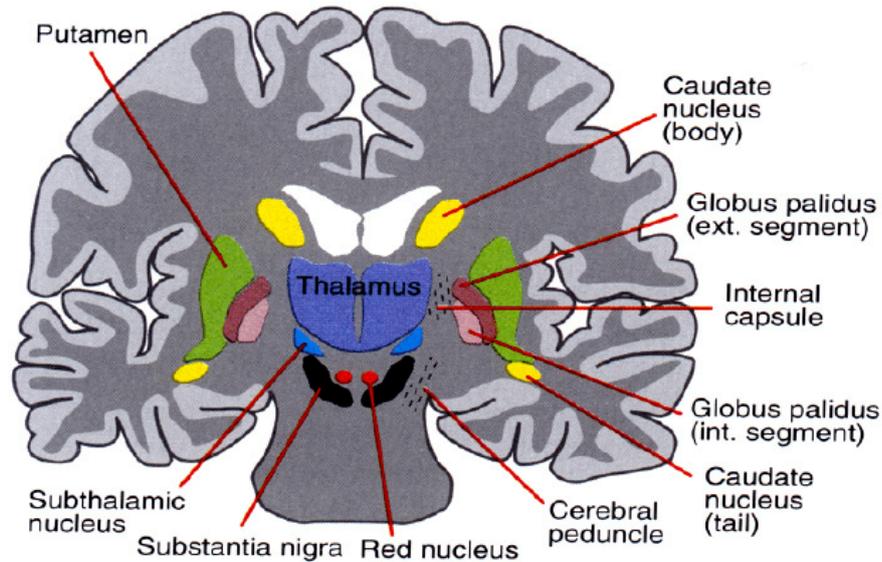


Figura 1.2: **Anatomía del cerebro, ganglios basales [2]**

Negra por medio de conexiones límbicas. Ambos componentes son parte de una cadena reguladora de movimiento en la cual sus eslabones excitan o inhiben señales usando neuronas glutamérgicas y GABAérgicas respectivamente [12]. La Figura 1.2 expone los diversos componentes de esta cadena reguladora del movimiento, tales como el putamen, el globo pálido, la Sustancia Negra, el Núcleo Subtalámico, etc.

Para inhibir señales no deseadas usando neuronas excitatorias, el Núcleo Subtalámico envía señales por medio de las neuronas glutamérgicas que posee a las neuronas GABAérgicas del globo pálido interno, activándolas [12]. Esto resulta en una inhibición de señales que ocurre antes de que estas lleguen al resto del tálamo y luego a la corteza cerebral, evitando que sean ejecutadas por el resto del cuerpo en forma de movimientos no deseados [12].

A grandes rasgos, la cadena de señales en la cual participa la Sustancia Negra es la siguiente: La corteza y la Sustancia Negra envían señales excitatorias usando neuronas glutamérgicas al cuerpo estriado (striatum). El cuerpo estriado envía señales inhibitorias usando neuronas GABAérgicas al globo pálido externo. El globo pálido externo envía señales inhibitorias al Núcleo Subtalámico. El Núcleo Subtalámico envía señales excitatorias al globo pálido interno y a la SNpr, Estas dos últimas estructuras envían señales inhibitorias al tálamo antes de que este envíe señales excitatorias de movimiento a la corteza cerebral y sean ejecutadas por el resto del cuerpo [12].

1.7.4 Resonancia magnética

La resonancia magnética es un proceso físico en el cual partículas como electrones, protones, núcleos atómicos con número impar de neutrones o protones son excitados mediante un campo magnético [13]. Esta excitación produce ondas de radiofrecuencia que pueden ser capturadas por una antena receptora, para posteriormente construir una imagen; de ahí el nombre, imagen de resonancia magnética (IRM) [13]. La Figura 1.3 muestra ejemplos de imágenes resultantes de este proceso desde un punto de vista axial y con diferentes secuencias (configuraciones); T2, FLAIR, DWI, etc.

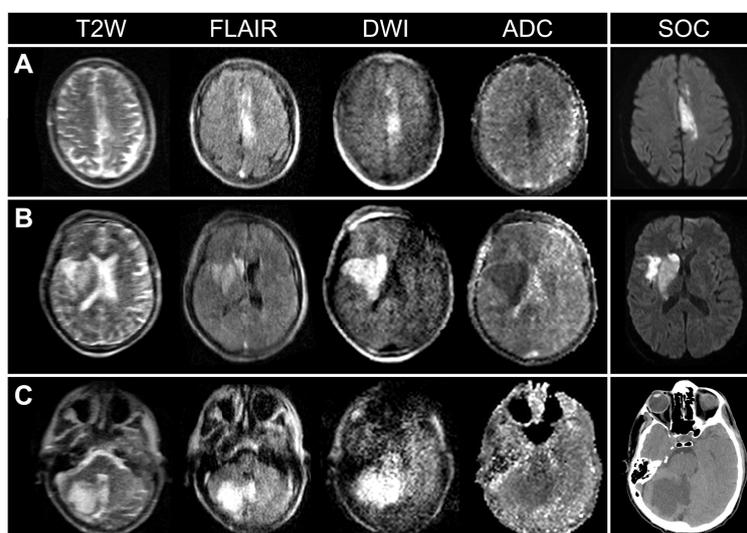


Figura 1.3: Ejemplos diversos de resonancia magnética [3]

Este método tomográfico presenta algunas ventajas sobre otros tipos de imágenes como son: la habilidad de registrar cortes en diferentes secciones del espacio a capturar, su alto nivel de contraste además de su capacidad para manipularlo luego de su captura y el poco riesgo o efectos nocivos al no hacer uso de radiaciones ionizantes [14].

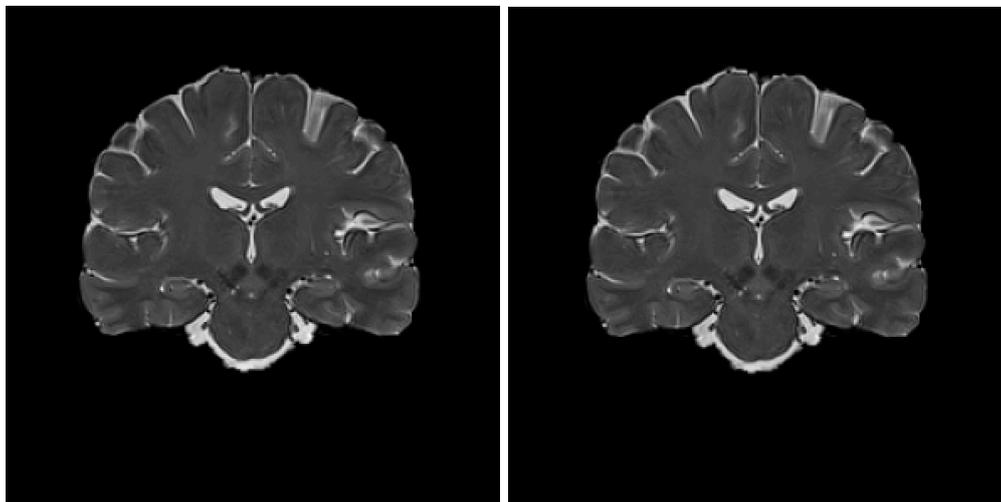
1.7.5 Técnicas de preprocesamiento

Como en todo proyecto de aprendizaje profundo que hace uso de datos, estos deben ser previamente procesados con el fin de poder hacerlos comparables entre sí y que todos los ejemplos compartan el mismo formato que espera el modelo. Esto se puede hacer a través de una serie de técnicas que se encargan de limpiar, corregir, filtrar y/o transformar las IRMs capturadas.

En el presente trabajo se realizaron las siguientes técnicas de preprocesamiento sobre las IRMs: corrección del sesgo, normalización de intensidad y registro de imagen.

1.7.5.1 Corrección del sesgo

Perturbaciones no deseadas en el campo magnético pueden producir que ciertas zonas de la IRM sean muy oscuras o muy brillantes o presenten un pobre contraste entre pequeños detalles, como se ve en la sección Figura 1.4a, esto puede causar una pérdida de información. Para solucionar este problema, se utilizó una técnica de corrección del sesgo sección Figura 1.4b, la cual calcula un campo de corrección para cada zona de la imagen con el fin de nivelar las diferencias entre las zonas [15].



(a) IRM sin corrección del sesgo (b) IRM con corrección del sesgo

Figura 1.4: Comparativa de IRM original vs IRM con corrección de sesgo [autoría propia]

1.7.5.2 Normalización de la intensidad

Las IRMs pueden variar en el valor de intensidad que le asignan a los vóxeles o píxeles, esto debido a que no todas las IRMs fueron capturadas con el mismo resonador. Es por esto que es necesario aplicar una técnica de normalización de intensidad, esto permite que los valores de los vóxeles de la imagen sean comparables entre todas las IRMs del conjunto de datos [16]. La Figuras 1.5a y la Figura 1.5b es una comparativa de una IRM original y una con normalización de intensidad, los efectos no son muy visibles a simple vista, debido a que el cambio principalmente consiste en una normalización del rango del

valor de los vóxeles conservando la naturaleza del contraste original.

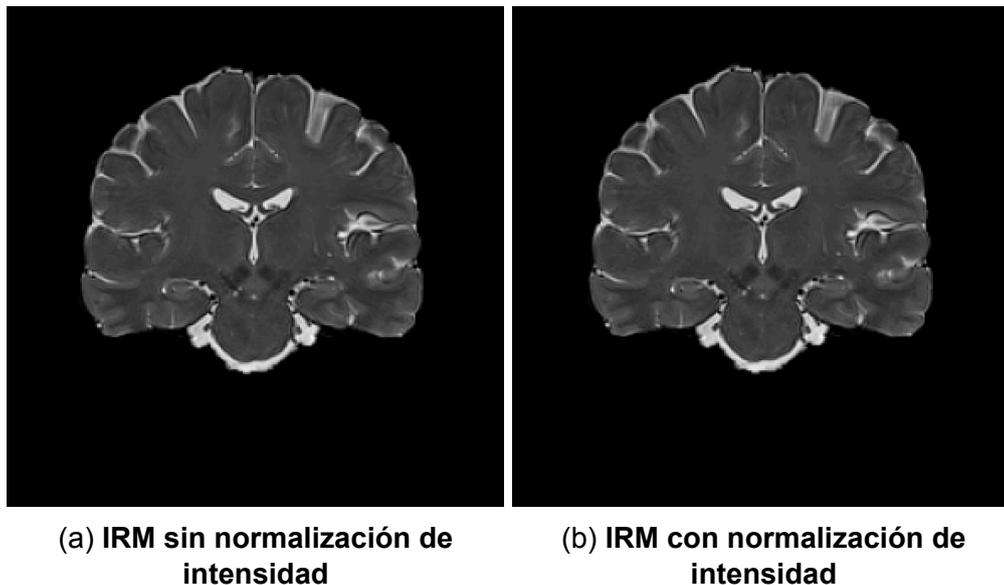


Figura 1.5: **Comparativa de IRM original vs IRM con normalización de intensidad [autoría propia]**

1.7.5.3 Registro de la imagen

Si bien durante la captura de la IRM se tratar de que el paciente se acomode en una posición determinada, esta posición no siempre es la misma. A esto sumemos el hecho de que la forma y volumen de cada cerebro nunca son iguales.

Estas variaciones, provocan que la inclinación u orientación del cerebro no sea la misma para todas las IRMs, lo que no las hace completamente comparables entre sí, más aún cuando lo que se desea en ellas es localizar ciertas zonas o elementos presentes.

La solución a este problema consiste en aplicar el proceso de registro, el cual realiza una transformación geométrica [17] a la imagen fuente hacia una imagen objetivo o plantilla; ejemplificada en la Figura 1.6a, donde se puede observar cierta diferencia entre la IRM objetivo (en tonos grises) y la IRM original (en tonos azules). Una vez aplicada la transformación a la IRM, esta se ajusta a la IRM objetivo usada, tal como se aprecia en la Figura 1.6b. Para este proceso existen dos tipos de transformaciones mostrados en la Figura 1.7, la lineal y la no lineal.

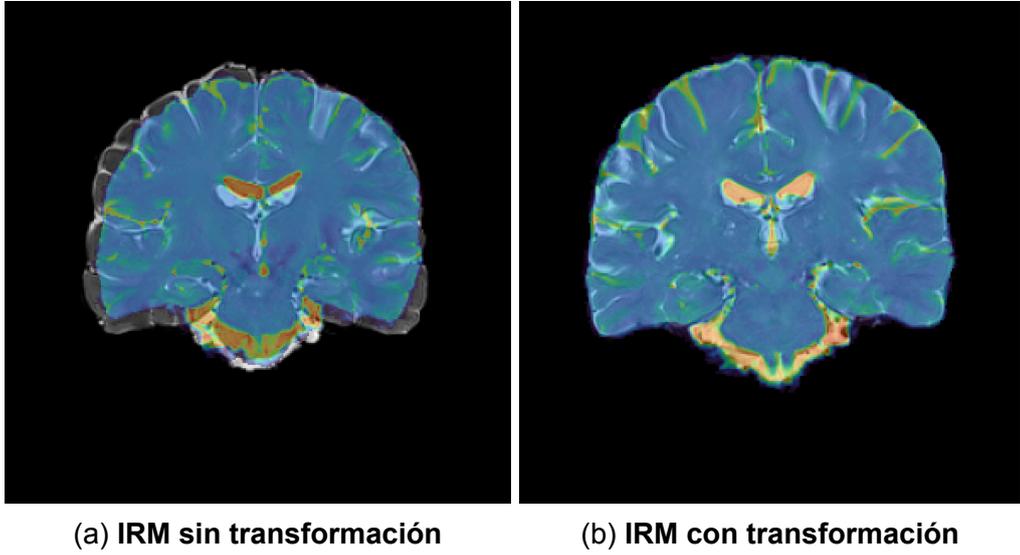


Figura 1.6: Comparativa de IRM original vs IRM con transformación [autoría propia]

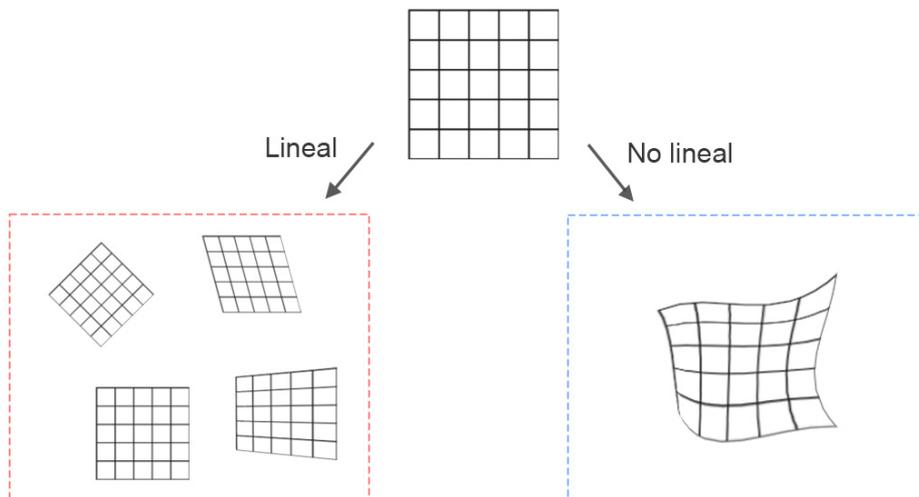


Figura 1.7: Transformaciones espaciales en imágenes [autoría propia]

- Transformación lineal: dispone de un total de 6 grados de libertad, con movimientos de rotación y translación.
- Transformación no lineal: dispone de miles grados de libertad aplicables a cada vóxel de la imagen.

1.8 Datos

Las IRMs objetivo para el presente trabajo tuvieron que cumplir con una serie de requisitos, los cuales eran:

- Cortes a no más de 2 milímetros de separación, esto debido a que el Núcleo Subtalámico tiende a tener no más de 5 mm de grosor. Por lo que, IRMs con cortes a mayor separación, se arriesgan a perder información importante.
- IRMs tomadas en secuencia T1 o T2. Esto debido a que son las secuencias en donde se puede empezar a distinguir las áreas de interés.
- Capturas de las áreas de la Sustancia Negra y Núcleo Subtalámico, debido a que son justamente las regiones que se desea segmentar.
- IRMs de cerebros sin lesiones o deformaciones anteriores, dado que los pacientes con la EP no presentan lesiones adicionales que pueden entorpecer el aprendizaje de los modelos de Inteligencia Artificial.

Las fuentes de donde se obtuvieron las imágenes fueron:

- Laboratorio de Neurociencias y Bioingeniería de la FIMCP: el cual brindó 60 IRMs 3D en secuencia T1 con cortes a 2 milímetros con interpolación a 1 milímetro.
- Colaboración con centro médico de España: el cual facilitó 20 IRMs 3D en secuencia T2 con cortes a 1 milímetro y T2 con parámetros optimizados para resaltar las zonas de interés.

1.8.1 Segmentación de la imagen

La segmentación es uno de los problemas presentes en el análisis de imágenes. Según [18], la segmentación consiste en la separación en varias regiones denominadas segmentos. Estos pueden ser diferenciados por una línea que delimita el área/volumen correspondiente al objeto o coloreando al objeto de un color de contraste. La Figura 1.8 muestra la segmentación de lesiones cerebrales como ejemplo; se puede apreciar claramente un contraste que delimita la región de interés del resto de la imagen.

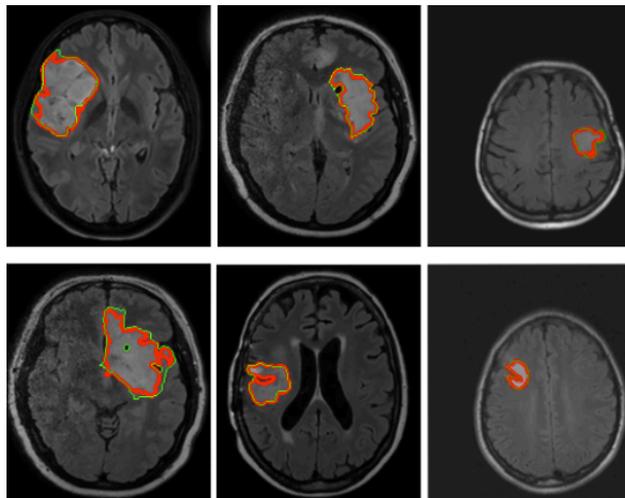


Figura 1.8: Segmentación en resonancia del cerebro [4]

1.9 Aprendizaje profundo para segmentación

En esta sección se describe al estado del arte referente a los modelos propuestos para la segmentación de imágenes. En el estado del arte existen dos principales enfoques:

- Modelos basados en convoluciones.
- Modelos basados en *vision transformers*.

1.9.1 Modelos basados en convoluciones

Se componen de una estructura *encoder-decoder*. En donde la primera sección se encarga de recibir la imagen y crear un mapa de características. Finalmente, la segunda sección se encarga de recibir el mapa de características y reconstruir una imagen segmentada.

Los modelos basados en convoluciones se utilizan principalmente en el procesamiento de imágenes. Estas capas de convolución aplican filtros entrenables que resaltan ciertas características de la imagen. Estos filtros, que generalmente son de un tamaño 3x3 o 5x5, funcionan como una ventana deslizante, la cual se mueve en la imagen con un desplazamiento en píxeles indicado por el parámetro de la capa llamado 'stride' [19]. La Figura 1.9 muestra este proceso, así como el mapa de características resultante; éste último contiene los patrones a aprender por el modelo.

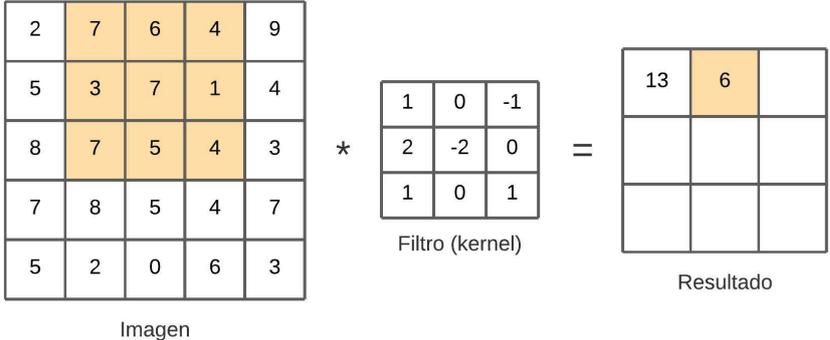


Figura 1.9: **Proceso de convolución en una imagen [autoría propia]**

1.9.1.1 Modelo U-Net

Este modelo también se compone de una estructura *encoder-decoder*. La primera sección se encarga de recibir la imagen y devolver un mapa de características, mientras que la sección del *decoder* reconstruye la imagen segmentada [20].

La Figura 1.10 muestra cómo esta arquitectura usa capas de convolución de la familia VGG en el *encoder*, específicamente la llamada VGG11. La arquitectura se compone de una serie de capas de convolución con la cual se obtiene un mapa de características.

Por otro lado, el *decoder*, utiliza capas de convolución transpuestas, las cuales duplican las dimensiones del mapa de características y reducen el número de canales a la mitad. Este proceso de *upsampling* se repite tres veces [20]. Además, para evitar el desvanecimiento del gradiente, cada proceso de *upsampling* se junta con una *skip connection*, para finalmente, obtener una imagen de las mismas dimensiones y canales que la imagen original, pero con los objetos de interés segmentados [20] [21].

Este modelo ha sido ampliamente utilizado para la segmentación en imágenes 2D e incluso en 3D. Como nos indica [22], el modelo ha demostrado ser capaz de llevar a cabo satisfactoriamente la segmentación del volumen de la próstata en IRMs de secuencia T2.

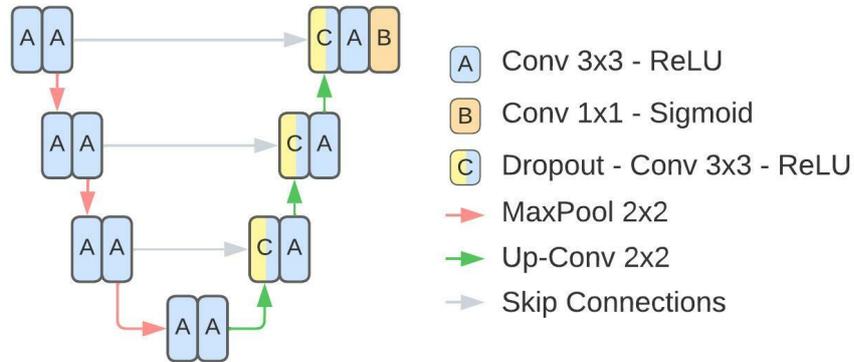


Figura 1.10: **Diagrama de la arquitectura U-Net [autoría propia]**

1.9.1.2 Modelo CLCI-Net

Otra arquitectura utilizada para la segmentación con IRMs, es el modelo CLCI-Net. Este modelo también se basa en una estructura *encoder-decoder* al igual que el modelo U-Net y presenta capas de convolución, principalmente con filtros de 3x3 y 1x1 [5].

Sin embargo, su principal diferencia radica en la manera en cómo se lleva a cabo las *skips connections*. Estas conexiones ya no son realizadas por una operación de concatenación, si no por una unidad de convolución LSTM (*Long Short Term Memory*), con el fin de capturar los patrones presentes mediante el análisis de la información recibida por el estado oculto, es decir, los mapas de características de la *skip connection* [5].

Si bien el estado del arte menciona que cuando se está trabajando con imágenes el mejor enfoque es usar redes basadas en convoluciones, esto no necesariamente aplica cuando lo que se desea es analizar una secuencia de imágenes, o procesar de manera más fina al estado oculto entre dos secuencias de imágenes. Para esto debemos utilizar un enfoque basado en convoluciones LSTM. Esta es una capa recursiva con la diferencia de que las multiplicaciones internas son realizadas por operaciones con convolución [23] [24].

Como se puede observar en la Figura 1.11, el modelo CLCI-Net presenta una estructura de *encoder-decoder*, donde 18 operaciones son realizadas con convoluciones de filtro 3x3 y 1 con filtro 1x1. Para evitar desvanecimientos del gradiente, en la sección del *encoder* se usan *skip connections* tradicionales en la capas 3ra, 6ta, 9na, 12da y 15ta. Por otro lado, en la sección del *decoder* las *skip connections* se realizan con una capa *ConvLSTM* luego de la 1ra, 4ta, 7ma y 9na capa.

A pesar de que el modelo explicado ha sido evaluado en un área de análisis diferente, el cual es detectar la presencia de tumores en resonancias magnéticas, este sí presenta una evaluación en un contexto, el cual consiste en el uso de imágenes en 3D (IRMs) como dato de entrada para el modelo. Convirtiéndolo en un candidato apropiado y afín para el área investigación de este proyecto.

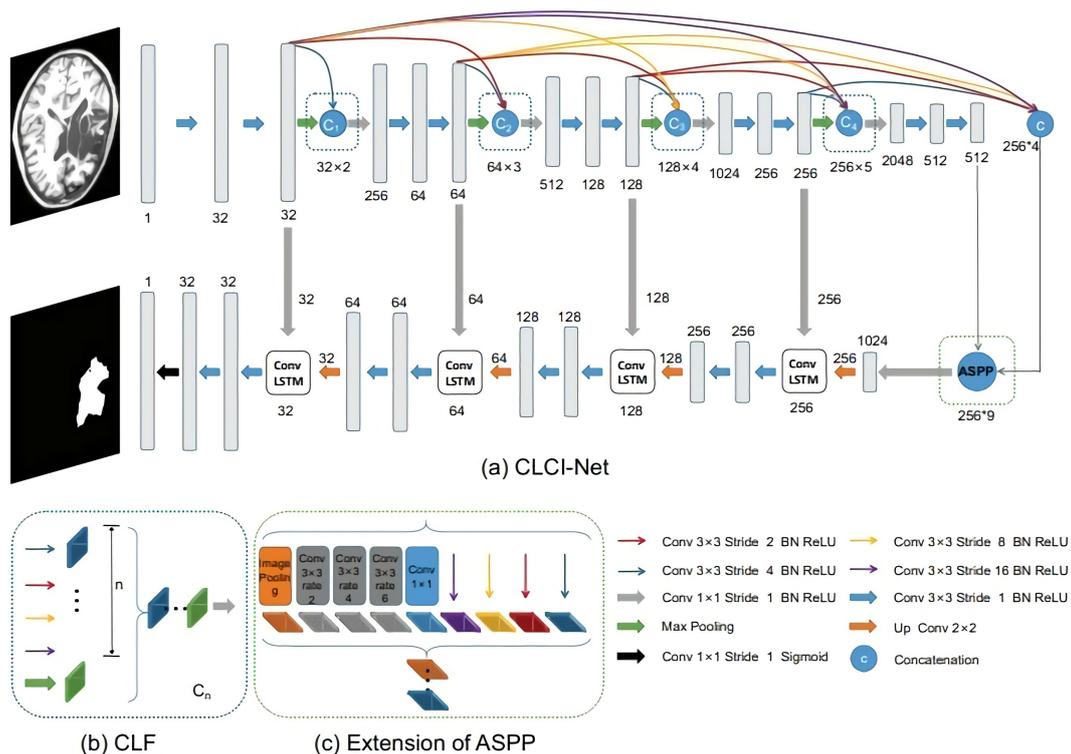


Figura 1.11: Diagrama de la arquitectura CLCI-Net [5]

1.9.2 Modelo basado en *vision transformer*

Estos modelos, al igual que los basados en capas de convolución, se basan en secciones de *encoder-decoder*. Sin embargo, su principal diferencia radica en la composición del

encoder, el cual está conformado por capas de multi atención [25].

Por otro lado, existen modelos que usan como *decoder* capas de multi atención y otros que usan capas de convolución. Entre los modelos basados en *transformers* tenemos a SETR (*Segmentation Transformer*) [25].

La arquitectura de *transformer* normalmente se usa para procesamiento natural de texto, en la que recibe una secuencia unidimensional de *tokens* que representan el texto de entrada [26].

Ahora bien, una imagen no es una secuencia unidimensional como lo es el texto. Primero se debe convertir la imagen en una secuencia de sub-conjuntos 2D llamados *patches* como ejemplifica la Figura 1.12; estos son análogos a los *tokens* que representan al texto. Una vez realizada esta conversión, se puede pasar a un modelo de *transformers* para que procese la secuencia de *patches* al igual que en el procesamiento de texto [27].

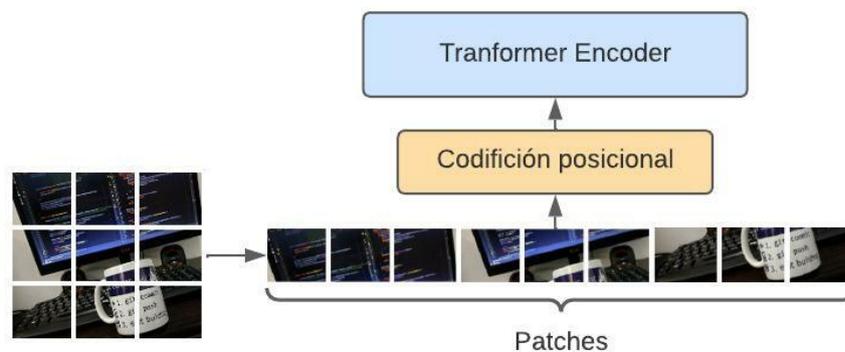


Figura 1.12: Diagrama de la arquitectura *vision transformer* [autoría propia]

1.9.2.1 Modelo SETR

Su comportamiento es análogo a los modelos basados en capas de convolución, con la diferencia de que, para la sección del *encoder*, este modelo está basado totalmente en capas de multi atención [25]. Una vista general de la arquitectura se encuentra en la Figura 1.13. En esta figura vemos como una serie de *patches* pasan por un proceso de codificación posicional antes de pasar a las capas de *transformers* y como la salida del *encoder* entra directamente a una cadena de *upsampling*.

Una visión general del funcionamiento es que la imagen de entrada es codificada por un *transformer* y decodificada por capas de convolución que reconstruyen una imagen con las mismas dimensiones y canales que la imagen original [25].

La sección de *encoder* se compone de un *transformer* con la siguiente estructura [25]:

- Codificación posicional.
- Codificación de *patches*.
- Bloque de transformación:
 - Capa de normalización.
 - Capa de multi atención.
 - Capa de normalización.
 - Multi capa de perceptrones (MLP).

Finalmente, para la sección del *decoder*, el modelo SETR presenta dos variantes: *upsampling* progresivo (PUP) y agregación de características en multi nivel (MLA) [25].

La variante de PUP consiste en tomar como entrada del *decoder* a la última salida del *encoder* y aplicar cuatro bloques de convolución con *upsampling* para reconstruir la imagen con los objetos de interés segmentados [25].

Por otro lado, la variante MLA consiste en tomar como input del *decoder* salidas intermedias de los bloques de *transformer* en el *encoder*. Estas salidas intermedias se combinan en una operación de agregación para luego, al igual que en la variante PUP, aplicar cuatro bloques de convolución con *upsampling* para reconstruir la imagen con los objetos de interés segmentados [25].

Si bien los trabajos relacionados a segmentación a través de *vision transformer* arrojan rendimientos aceptables, estos mismos modelos representan un alto costo computacional y requieren de un conjunto de datos bastante grande para que el rendimiento sea el mejor posible. Según [28], conjuntos de datos con una cardinalidad por debajo de los ocho mil ejemplares provocan un rendimiento bajo en los modelos entrenados.

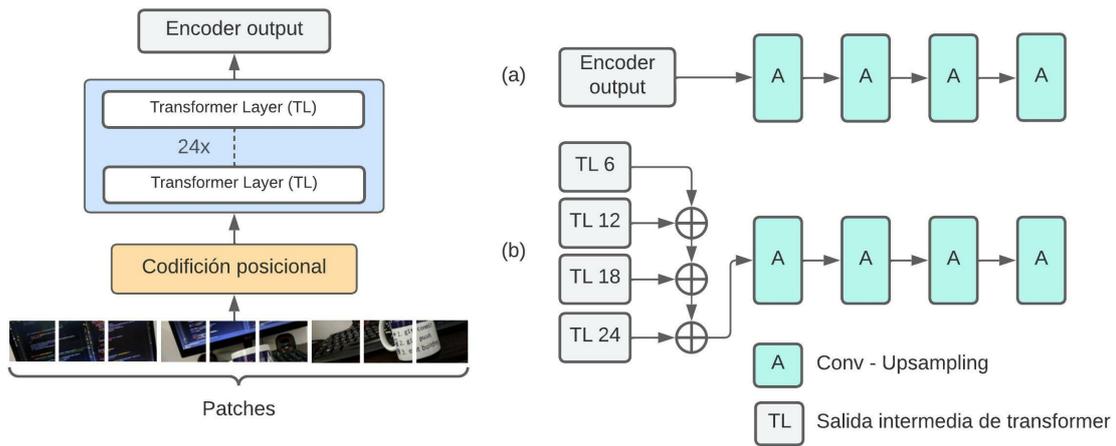


Figura 1.13: **Diagrama de la arquitectura SETR. (a) PUP (b) MLA [autoría propia]**

Dado lo expuesto en el presente capítulo, los modelos CLCI-Net [5] y U-Net [22], son los que han sido evaluados para trabajar con un tipo de imágenes similar al de este proyecto, las cuales son IRM de secuencia T1 y T2.

Por otro lado, las boyantes arquitecturas basadas en *transformers* [25], están demostrando dar muy buenos resultados en diferentes áreas, incluyendo la segmentación de imágenes. Con la consideración según [28], de que estas arquitecturas demandan un hardware robusto y conjuntos de datos bastante extensos.

Es por esto que, en el presente trabajo se hizo un análisis del rendimiento de los modelos CLCI-Net, U-Net y SETR. En el siguiente capítulo se hablará del proceso metodológico a seguir, así como se explicarán las métricas de evaluación.

CAPÍTULO 2

2. METODOLOGÍA

En esta sección se desglosa el procedimiento al cual se sujeta el estudio y los detalles de evaluación a considerar cuando se presenten resultados en el siguiente capítulo. Esta metodología se centra en un análisis del rendimiento de los modelos CLCI-Net, U-Net y SETR para la segmentación de la Sustancia Negra y Núcleo Subtalámico.

2.1 Etapas de la metodología

Los pasos necesarios para llevar a cabo este análisis, ilustrados por la Figura 2.2, se desglosan a continuación:

- **Procesamiento del conjunto de datos:** Tuvo el fin de extraer sólo el volumen objetivo, es decir, un volumen reducido del cerebro que contiene tanto al Núcleo Subtalámico como a la Sustancia Negra. El conjunto de datos contenía imágenes en un espacio estándar. No fue necesario hacer un registro de espacio nativo a estándar en este caso.
- **Data augmentation:** Se aplicaron técnicas de aumento de datos de manera aleatoria tales como la corrección del sesgo, normalización de intensidad, rotación (rotaciones de 90 grados), *flip* horizontal y vertical.
- **Adaptación del modelo base:** Se necesitó modificar los modelos a usar para acomodarlos a un entrenamiento usando imágenes de resonancia magnética en 3D.
- **Entrenamiento:** Se procedió a entrenar al modelo adaptado con el conjunto de datos aumentado.

- **Evaluación de métricas y tuneo de parámetros:** Se analizó el rendimiento del modelo usando las métricas *Precision*, *Recall*, *F1 Score*, coeficiente Sorensen-Dice (DICE), *Hausdorff Distance (HD)* y *Average Volume Distance (AVD)*. Para, en base a los resultados del modelo, poder ajustar los parámetros con el objetivo de mejorar su rendimiento previo.

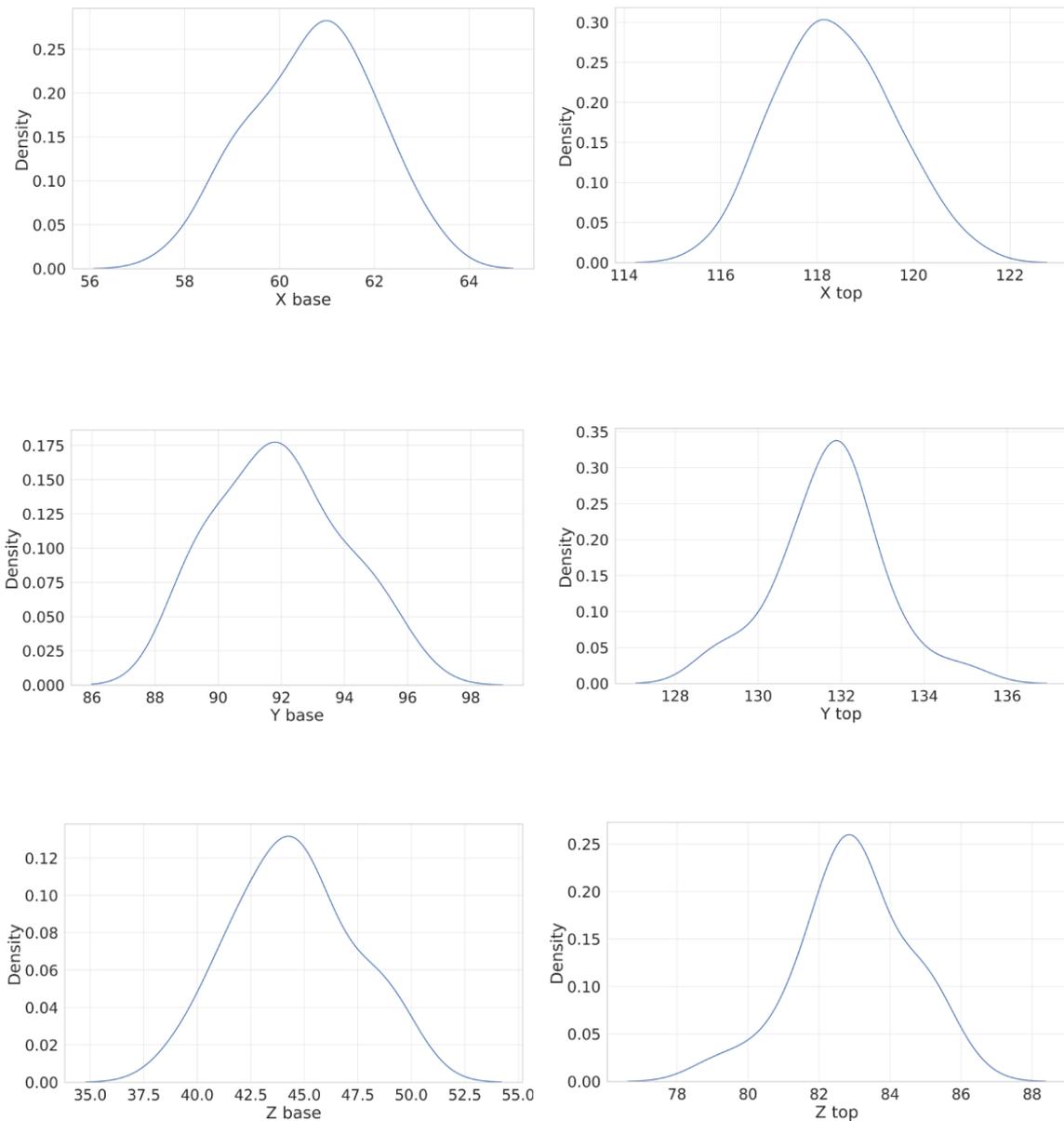


Figura 2.1: Distribución de las coordenadas de las áreas de interés [autoría propia]

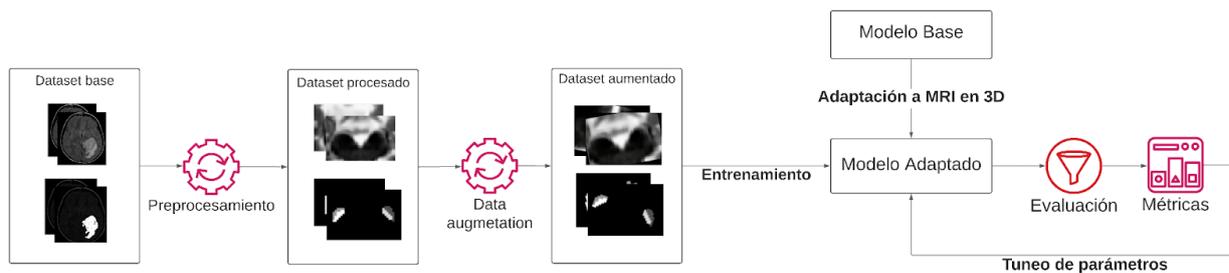


Figura 2.2: Flujo de datos del estudio [autoría propia]

2.1.1 Preprocesamiento del conjunto de datos

En este paso se tomó y recortó las IRMs descritas previamente en la sección 1.8, dejando solo un corte donde se encuentran los *targets* del estudio; la Sustancia Negra y Núcleo Subtalámico.

Las coordenadas de los cortes de estas regiones fueron determinadas mediante un análisis sobre las IRMs ya segmentadas. Se registraron las coordenadas que encierran a los puntos de interés para luego construir un gráfico de distribución de los valores. La Figura 2.1 contiene la distribución de coordenadas en cada uno de los ejes (x, y, z) en cuanto a dónde se encontraban los *targets* dentro del conjunto de datos. Esto indicó, de manera general, los valores promedio de las coordenadas para las regiones de interés.

2.1.2 Data augmentation

Luego de realizar la extracción del volumen que contiene a los *targets*, se procedió definir las siguientes transformaciones:

- **Flip horizontal:** Con 2 opciones, aplicar o no aplicar.
- **Flip vertical:** Con 2 opciones, aplicar o no aplicar.
- **Corrección del bias:** Con 2 opciones, aplicar o no aplicar.
- **Normalización de intensidad:** Con 2 opciones, aplicar o no aplicar.
- **Rotación:** Con 4 opciones de ángulos rectos (90 grados).

Al aplicar un producto cartesiano entre estas transformaciones se obtiene un valor de 64 combinaciones, es decir, que se tuvo un conjunto de datos con un total de 3840 IRMs partiendo de 60.

2.1.3 Adaptación de los modelos de base

Para los modelos escogidos fue necesario aplicar cambios en ciertas capas, con el fin de que el modelo sea capaz de recibir una matriz correspondiente a una IRM y devolver otra de iguales dimensiones con las secciones de interés ya segmentadas. A continuación, se describen los cambios realizados para los respectivos modelos:

- **CLCI-Net (Figura 1.11):** Se cambió la última capa de convolución. Antes del cambio solo devolvía una matriz plana de $80 \times 80 \times 1$. Después del cambio devuelve la misma cantidad de *slides* de la IRM recibida como input, esto es $80 \times 80 \times 56$.
- **U-Net (Figura 1.10):** Se cambió la última capa de convolución. Antes del cambio solo devolvía una matriz con su última dimensión de 3 canales (RGB), esto es: $80 \times 80 \times 3$. Después del cambio devuelve la misma cantidad de *slides* de la IRM recibida como input, esto es $80 \times 80 \times 56$.
- **SETR (Figura 1.13):** Este modelo no necesitó de una adaptación como tal, pero sí requería una codificación de la arquitectura en TensorFlow 2, ya que, al momento de este estudio, solo existen implementaciones en PyTorch.

2.1.4 Entrenamiento

Una vez que los modelos estuvieron listos para recibir las IRMs se procedió con el entrenamiento de estos. Durante su entrenamiento se registró los valores de la función de pérdida (DICE) y métricas adicionales como *Precision*, *Recall* y *F1 Score*, esto de manera preliminar. Los valores definitivos de estas métricas se definieron al momento de evaluar los modelos.

2.1.5 Evaluación de métricas y calibración de parámetros

Una vez que el aprendizaje del modelo se estabilizó, se analizó el comportamiento para la toma de decisiones sobre qué cambios se debían hacer en los parámetros del modelo.

Esto con el objetivo de mejorar puntos débiles y en consecuencia mejorar su rendimiento. Las métricas consideradas con su respectiva intuición relevante al problema fueron las siguientes:

- **Precision:** Indica cuál es la capacidad del modelo para garantizar que aquellos vóxeles considerados como parte del área de interés estén correctamente clasificados [29]. Tal como se aprecia en la Ecuación 2.1.

$$Precision = \frac{TP}{TP + FP} \quad (2.1)$$

- **Recall:** Indica la capacidad del modelo para encontrar todos los vóxeles que pertenecen al área de interés [29]. Definida por la Ecuación 2.2.

$$Recall = \frac{TP}{TP + FN} \quad (2.2)$$

- **F1 Score:** Es el promedio armónico entre el *Precision* y *Recall* y permite evaluar de manera holística el modelo cuando se busca comprender el *trade off* que generalmente ocurre entre estas 2 métricas [29]. Se expone la definición en la Ecuación 2.3.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (2.3)$$

- **Coeficiente Sorensen-Dice (DICE):** Indica el traslape entre la segmentación manual y la del modelo tomando en cuenta desaciertos [30]. En la Ecuación 2.4 se puede observar su definición.

$$DICE = \frac{2 \times TP}{(TP + FP) + (TP + FN)} \quad (2.4)$$

- **Hausdorff Distance (HD):** Evalúa la distancia entre dos subconjuntos de un mismo espacio métrico [30].
- **Average Volume Distance (AVD):** Muestra porcentualmente la diferencia volumétrica entre la segmentación manual y la segmentación del modelo en relación a la segmentación manual [30]. Dada por la Ecuación 2.5.

$$AVD = \frac{|VG - VP|}{VG} \quad (2.5)$$

Cabe recalcar que, en esta problemática en específico, cuando hay que elegir entre aumentar *Precision* o *Recall*, se debe dar más importancia a *Precision*. La motivación nace de que el trabajo está orientado a intervenciones quirúrgicas existentes en pacientes con Parkinson. En un escenario como este, es más importante tener certeza que la región segmentada sea la región de interés a encontrar exhaustivamente cada vóxel que pertenece a la región de interés.

Para otros problemas, inclusive dentro de la misma área médica, *Recall* puede ser de más importancia. Esto queda a criterio de los investigadores.

Esta sección expuso el marco de trabajo bajo el cual se desenvuelve este proyecto, metodología sobre la cual se destilan los resultados presentados en el siguiente capítulo. A continuación, se desglosan los resultados obtenidos.

CAPÍTULO 3

3. RESULTADOS Y ANÁLISIS

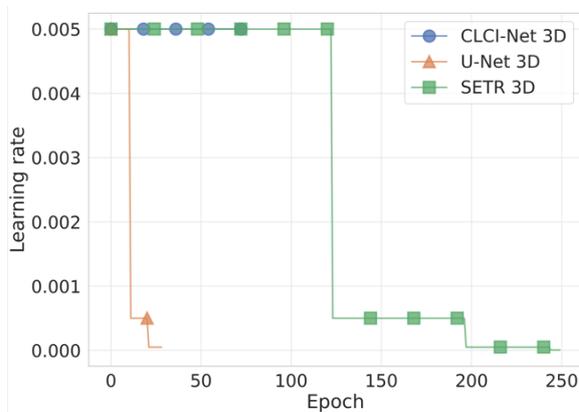
En este capítulo se exponen los resultados de los experimentos realizados, medidos a través de las métricas seleccionadas para evaluar los modelos previamente descritos. Este capítulo toma como referencia los resultados de los modelos base y luego se muestran los resultados con el modelo más prometedor ya afinado.

3.1 Modelos de base

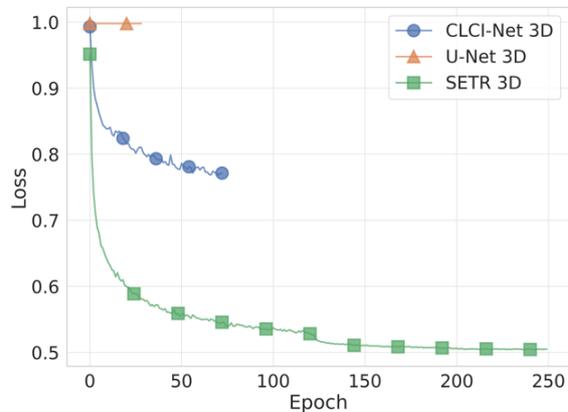
Como se puede observar en las Figuras 3.1 y 3.2, solo el modelo SETR pudo aprender a segmentar los volúmenes de interés. Como evidencia, se obtuvo un valor superior de coeficiente DICE en la Figura 3.1c y el *F1 Score* en la Figura 3.1e. En el caso de los otros modelos correspondientes a CLCI-Net y U-Net se puede observar que su desempeño fue inferior al obtenido por el modelo SETR; en particular el modelo U-Net, que desde el principio se estancó en valores bajos.

La Figura 3.1b nos muestra cómo el valor de la función pérdida fue disminuyendo conforme pasaban las épocas para SETR y CLCI-Net. *Precision* y *Recall* se desglosan a través de las épocas en las Figuras 3.1d y 3.1f respectivamente. *Precision* fue incrementando para ambos modelos mientras que *Recall* presentó un decremento para CLCI-Net.

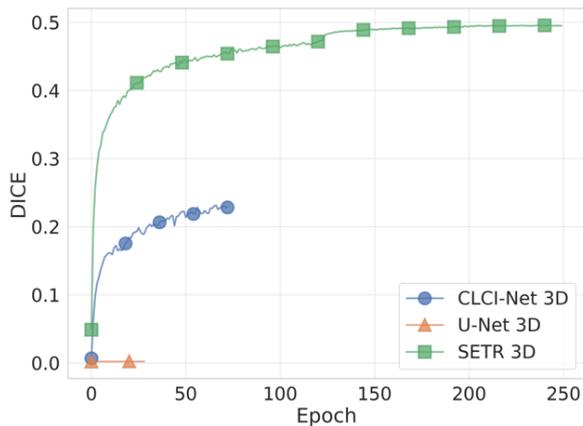
Adicionalmente, se realizaron pruebas con la porción de validación del conjunto de datos con el fin de confirmar que el modelo de hecho aprendía a generalizar y no caía en *overfitting*. Finalmente, los resultados del conjunto de validación en las métricas de coeficiente DICE (Figura 3.2b) y *F1 Score* (Figura 3.2d) son muy similares a los obtenidos



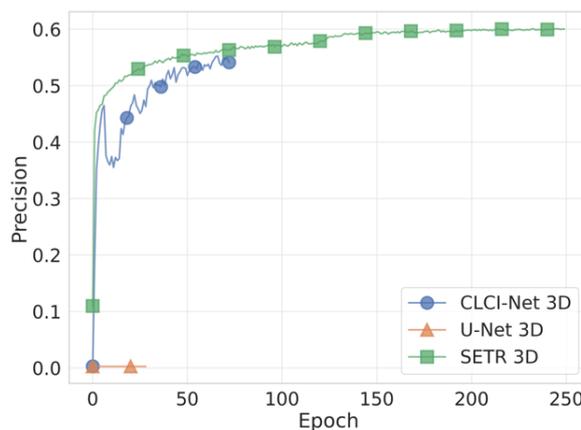
(a) Aprendizaje vs. épocas



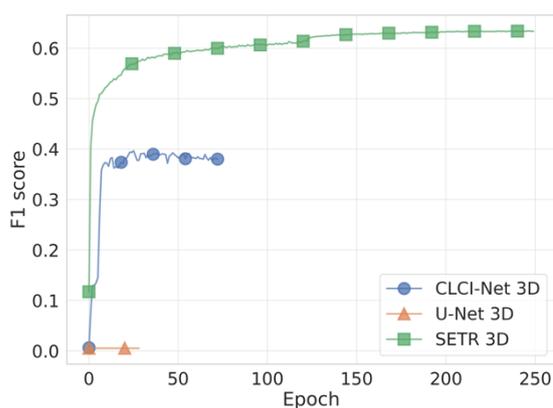
(b) Función de pérdida vs. épocas



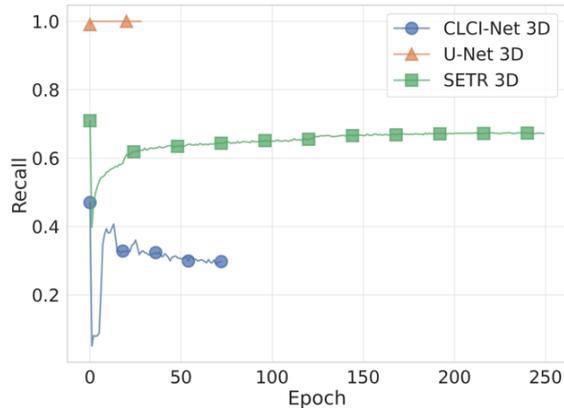
(c) Coeficiente DICE vs. épocas



(d) *Precision* vs. épocas



(e) *F1 Score* vs. épocas



(f) *Recall* vs. épocas

Figura 3.1: Resultado de métricas del *dataset* de entrenamiento en los modelos base [autoría propia]

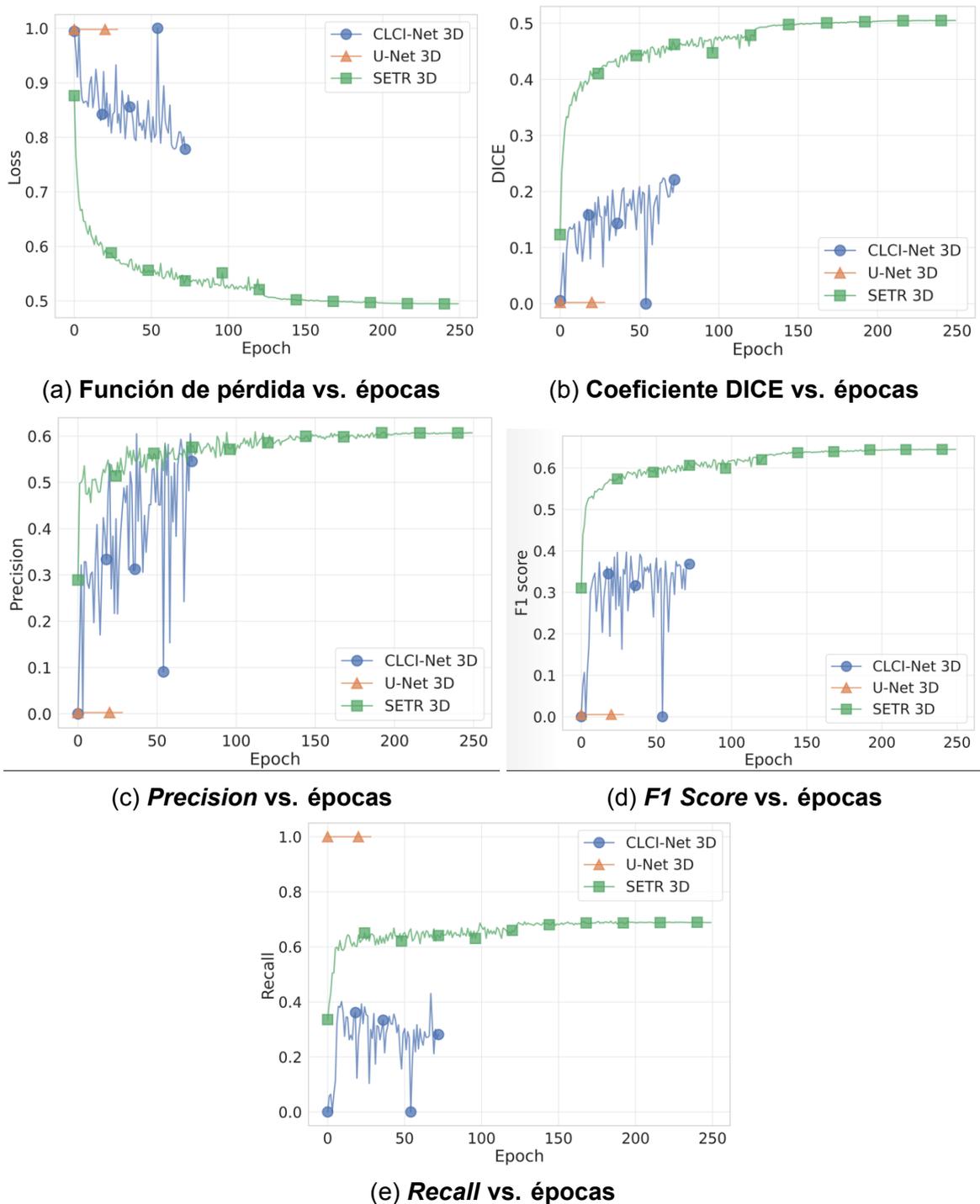


Figura 3.2: Resultado de métricas del *dataset* de validación en los modelos base [autoría propia]

en las métricas de la Figura 3.1, lo que confirma que el modelo si aprendió a generalizar.

Cabe recalcar que los valores presentados por el modelo basado en CLCI-Net fueron erráticos cuando se los calculaba a partir de la porción de validación. Esto está reflejado en todas las métricas pero particularmente en *Precision* (Figura 3.2c) y en *Recall* (Figura 3.2e), donde se puede ver que la curva que representa a CLCI-Net subía y bajaba de manera abrupta conforme pasaban las épocas.

Aparte, era de esperar que métricas promedio basadas (directa o indirectamente) en *Precision* y *Recall*, como lo son *F1 Score* y *DICE*, compartieran esta conducta previamente descrita. Esto está visualmente descrito en las Figuras 3.2d y 3.2b respectivamente.

Si bien el modelo SETR fue el que obtuvo los mejores resultados en comparación a los demás, estos valores no son lo suficientemente satisfactorios como para asegurar la calidad de la segmentación. Por este motivo se realizaron cambios al modelo SETR base con el fin de mejorar los resultados obtenidos.

3.2 Modelos mejorados

Los cambios realizados al modelo basado en *transformers* (Figura 1.13) que alcanzaron los mejores resultados son observables en la Figura 3.3, en la cual se incluyen las variaciones más importantes descritas a continuación:

- Se reemplazó el método para obtener los *patches*; se pasó de realizar simples cortes a los *items* del conjunto de datos a usar una capa de convolución 3D para obtener los *patches*.
- Se hizo pasar la salida obtenida por la sección *transformer*, por una capa completamente conectada (*fully connected*). Esta capa se encarga de procesar todo el estado oculto de las capas intermedias de la sección del *transformer*. Dando como resultado un nuevo estado oculto con el cual se reconstruye los tensores de la forma $10 \times 10 \times 7 \times 25$, para luego con la ayuda de capas *Upsampling* y Convoluciones 3D terminar con la forma final de $80 \times 80 \times 56 \times 5$.

Con respecto a los demás hiperparámetros del modelo se tuvo:

- Capas de *transformer*: 12
- *Patch size*: 8
- Vector de proyección: 64
- Número de cabezas del *transformer*: 8
- MLP de *transformer*: Dos capas ocultas de 128 y 64 neuronas cada una.

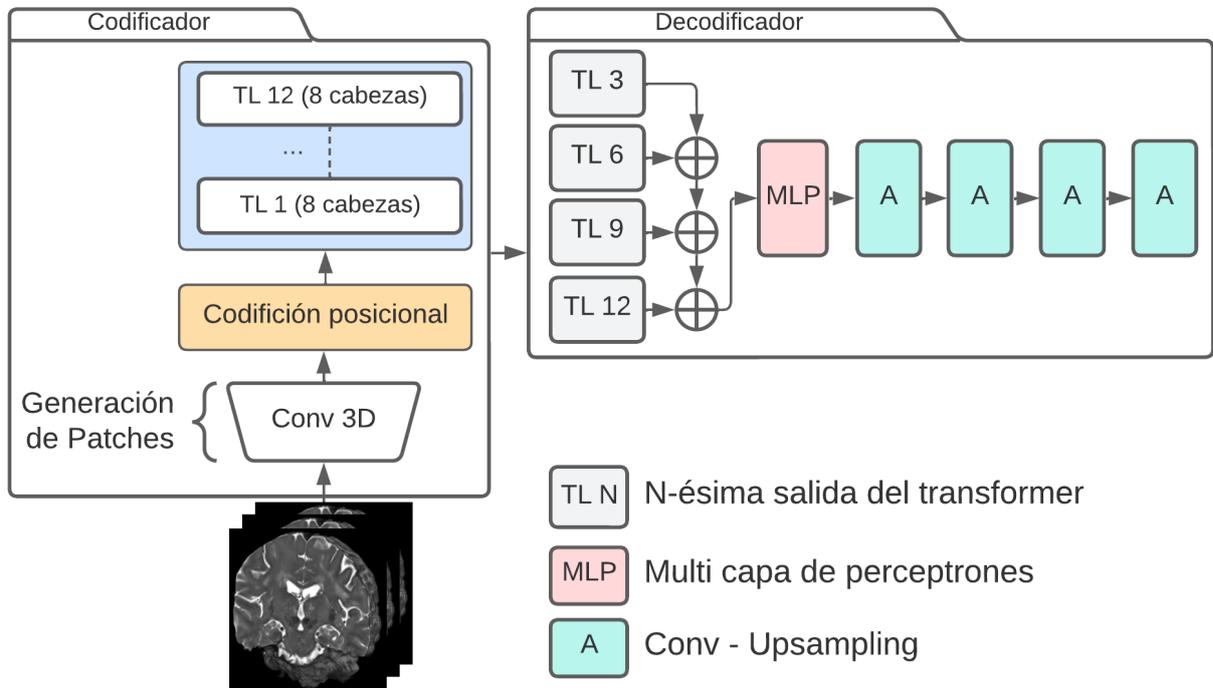


Figura 3.3: **Arquitectura mejorada para el modelo basado en transformer [autoría propia]**

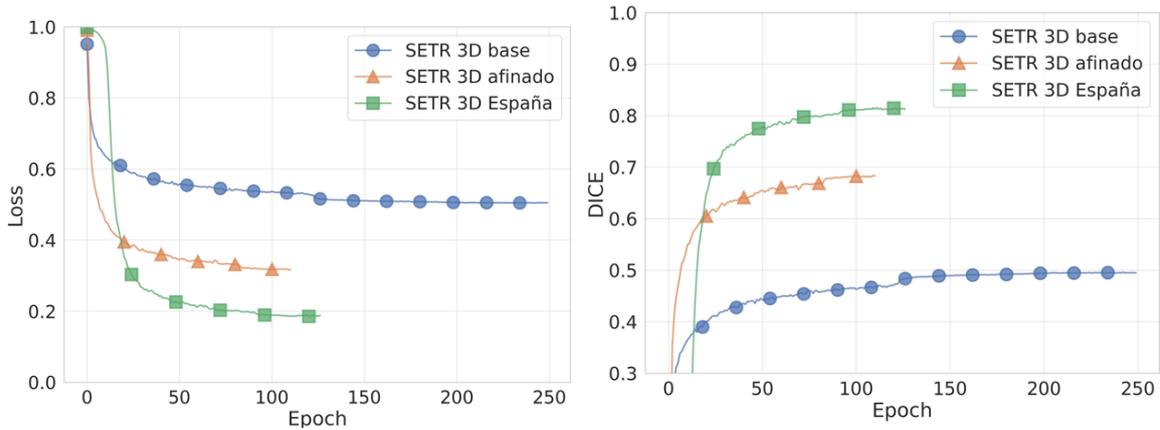
Una vez que se realizó la implementación de estos cambios en el modelo SETR, se procedió con el entrenamiento del modelo mejorado. Como se puede observar en la Figura 3.4, la cual contrasta las métricas del modelo base con el afinado y con el modelo usando el conjunto de datos de España, se obtuvieron resultados significativamente mejores que los obtenidos con el modelo base, pasando de un coeficiente DICE de 0,6 a 0,7 (Figura 3.4b). Además, la convergencia al mismo fue mucho más rápida que en el modelo base.

Con estos resultados mejorados del modelo SETR, se procedió a realizar un entrenamiento final con el conjunto de datos provisto por el laboratorio de España. Como se puede ver en la Figura 3.4, se obtuvieron resultados prometedores, obteniendo un coeficiente DICE del 0,8 (Figura 3.4b) y un *F1 Score* (Figura 3.4d) y un ritmo de convergencia similar al del modelo afinado.

La Figura 3.4a muestra que se obtuvo una caída rápida deseable en la función de pérdida para los tres modelos antes de las 50 épocas. La Figura 3.4c muestra como el modelo mejorado logró superar de manera temprana al modelo en cuanto a *Precision*.

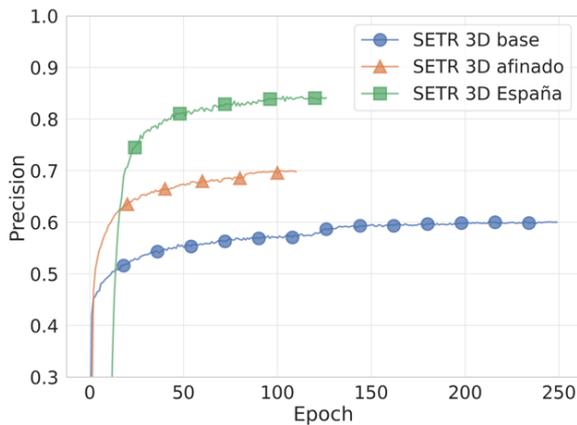
La Figura 3.4e muestra que, si bien el modelo afinado con el conjunto de datos español comenzó con un *Recall* en caída, logró recuperarse al punto de sobrepasar tanto al modelo afinado como al modelo base; esto con valores de 0.87, 0.70 y 0.74 respectivamente.

Finalmente, se procedió a realizar un cálculo para las métricas complementarias; estas son: *Hausdorff Distance*, *Average Volumen Distance* y *Cross-entropy* disponibles en la Tabla 3.1 expone los resultados obtenidos, la misma se explica a continuación.

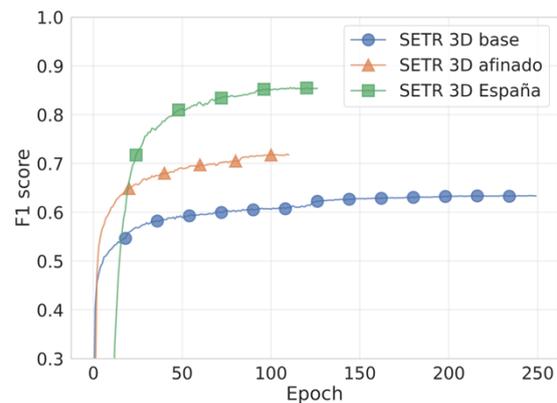


(a) Función de pérdida vs. épocas

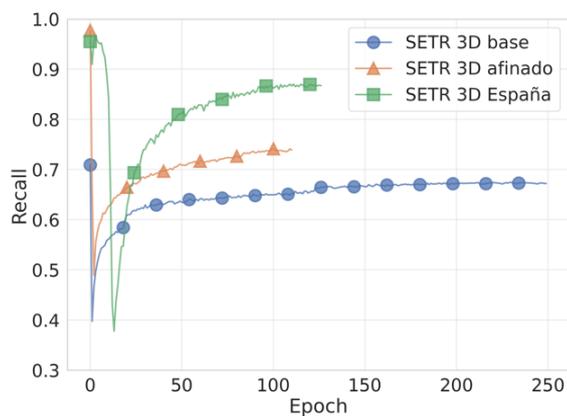
(b) Coeficiente DICE vs. épocas



(c) Precision vs. épocas



(d) F1 Score vs. épocas



(e) Recall vs. épocas

Figura 3.4: Resultado de métricas del *dataset* de entrenamiento de los modelos SETR [autoría propia]

Tabla 3.1: **Resultados de las métricas para el *dataset* de pruebas**

	CLCI-Net 3D	U-Net 3D	SETR 3D base	SETR 3D afinado	SETR 3D España
DICE	0.15	0.01	0.48	0.66	0.81
Recall	0.26	1.00	0.70	0.74	0.87
Precision	0.25	0.01	0.58	0.67	0.83
F1 Score	0.25	0.01	0.63	0.70	0.85
Loss	0.84	0.99	0.51	0.33	0.18
AVD	0.74	429.99	0.23	0.16	0.06
Hausdorff Distance	45.36	51.95	5.56	3.46	2.40
Cross entropy	0.03	16.10	0.02	0.01	0.001

En la Tabla 3.1 se puede observar que los modelos SETR son aquellos que también obtuvieron mejores resultados para las métricas: *Hausdorff Distance*, la cual indica la medida en la cual el volumen predicho se debe expandir o contraer para encapsular al volumen objetivo; *Average Volumen Distance*, la cual indica qué tan similar es el volumen entre lo predicho por el modelo y el volumen objetivo usando porcentajes; y, *Cross-entropy* que indica la capacidad del modelo para etiquetar correctamente la clase de cada vóxel.

Del análisis de cada una de las métricas obtenidas, expuestos en la Tabla 3.1, se obtienen las siguientes observaciones. Esto recordando que, en todo momento, el modelo afinado de SETR usando el conjunto de datos español fue superior a las otras alternativas.

El coeficiente DICE toma en cuenta los verdaderos positivos, falsos positivos y falsos negativos y sirve para entender el traslape entre el espacio predicho por el modelo y el espacio que debió ser predicho por el modelo. El traslape mejoró notoriamente en esta tarea al afinar la arquitectura SETR de base; 0.48 (base) a 0.66 (afinado), terminando con un (0.81) dado el conjunto de datos español. El traslape con CLCI fue bastante bajo (0.15) pero esto no se debió a que el modelo fue incapaz de encontrar espacialmente los objetivos, sino que no los clasificó de manera correcta. En otras

palabras, el modelo reconoció que un vóxel era parte de uno de los objetivos (Núcleo Subtalámico izq./der. Sustancia Negra izq./der.) pero cometió un número considerable de errores para determinar a cuál de estas categorías pertenecía.

SETR presenta un valor coherente y satisfactorio entre *Recall* y *Precision* (0.87 y 0.83 respectivamente). Los valores de *Recall* y *Precision* de U-Net vislumbran cuál fue el problema del modelo para conseguir buenas métricas promedio como lo es el *F1 Score* o métricas de similitud como el coeficiente DICE. U-Net presentó la tendencia a categorizar una enorme cantidad de vóxeles como parte las clases en cuestión encontrando todos los vóxeles correctos (*Recall* de 1.00) pero terminó equivocándose en casi todas sus clasificaciones (*Precision* de 0.01).

En cuanto al promedio armónico entre *Recall* y *Precision*, el *F1 Score*, se obtuvo que SETR tuvo un valor prometedor de (0.85). CLCI-Net tuvo problemas similares para obtener un buen valor (0.25) a lo que se explicó al momento de analizar el coeficiente DICE de esta misma arquitectura.

Visualmente se contrastaron resultados del modelo basado en SETR afinado con el conjunto de datos de España junto con resultados de un modelo basado en CLCI-Net utilizando el mismo conjunto de datos. La Figura 3.5, correspondiente a CLCI-Net, muestra una clasificación de vóxeles difusa. Si bien el modelo lograba identificar vóxeles como perteneciente a alguna de las clases de interés, fallaba en definir a qué clase de interés pertenecía con consistencia. Por otra parte, las Figuras correspondientes a los modelos basados en SETR (3.6 3.7) logran delimitar claramente los componentes de interés y agruparlos correctamente como sólidos definidos (no mezclados); NS izquierdo siendo rojo, NS derecho amarillo, SN izquierda siendo verde y SN derecha azul.

En este capítulo se expusieron los resultados obtenidos por los modelos al ser evaluados con las métricas expuestas en el capítulo anterior. En el siguiente capítulo se procede a exponer las conclusiones derivadas de los resultados presentados.

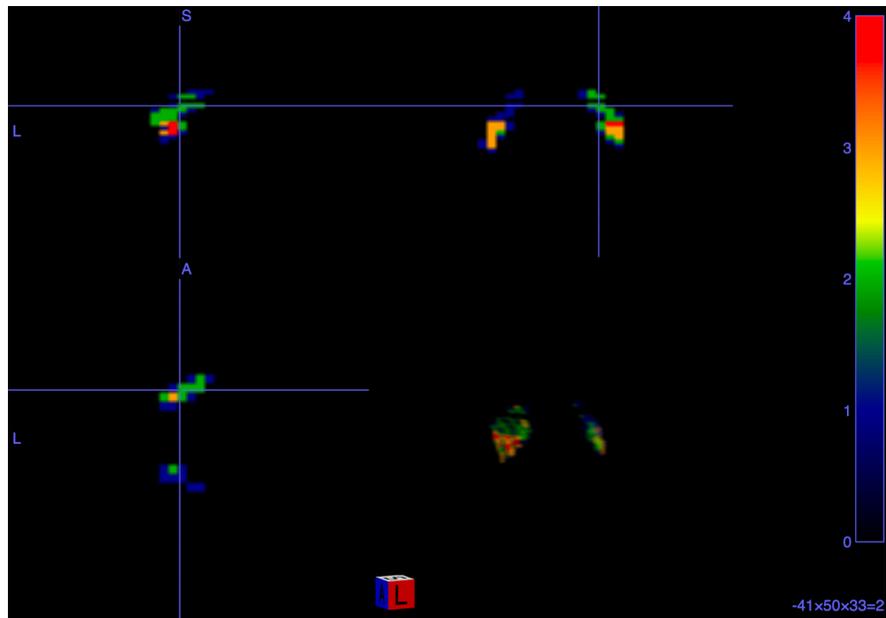


Figura 3.5: Segmentación resultante de modelo basado en CLCI-Net [autoría propia]

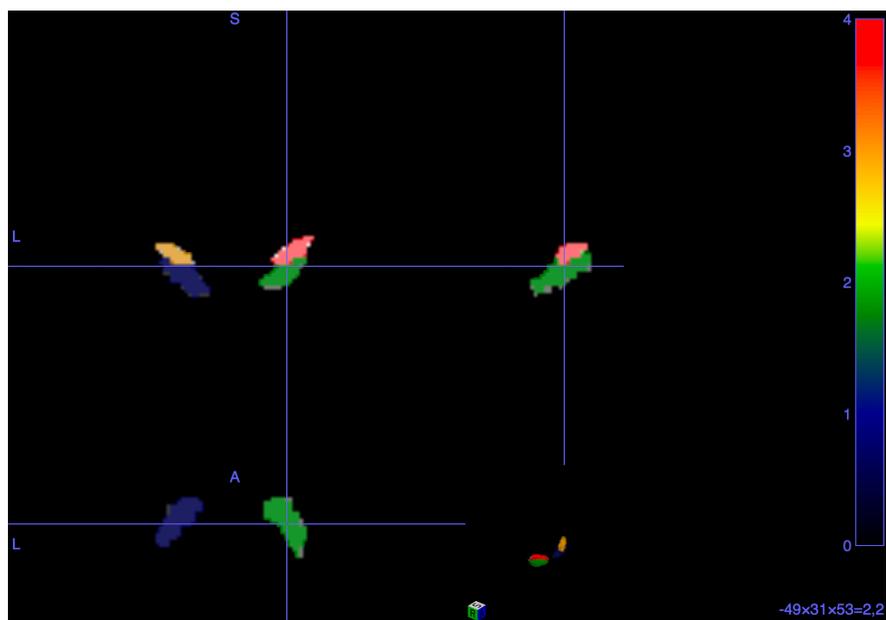


Figura 3.6: Segmentación resultante de modelo basado en SETR 1 [autoría propia]

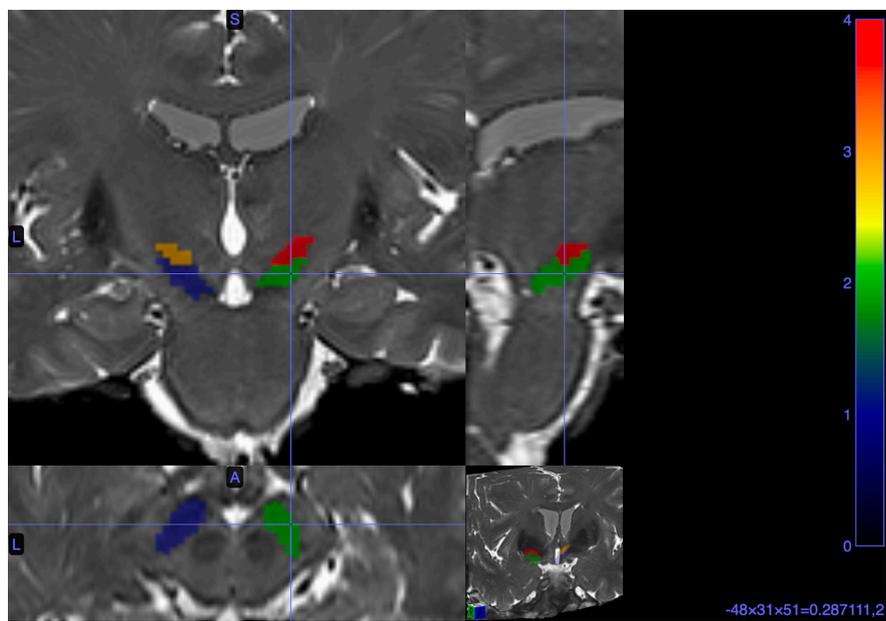


Figura 3.7: Segmentación resultante de modelo basado en SETR 2 [autoría propia]

CAPÍTULO 4

4. CONCLUSIONES Y TRABAJOS FUTUROS

En este capítulo se presentan las conclusiones a las que se llegó luego del proceso realizado y los resultados obtenidos.

4.1 Conclusiones

Tomando en consideración únicamente las arquitecturas base. Para la tarea de segmentación de imágenes 3D, la arquitectura SETR brinda los mejores resultados y se acopla mejor a la tarea de segmentación del Núcleo Subtalámico y Sustancia Negra utilizando imágenes en 3D (IRMs) cuando se la compara con las arquitecturas U-Net y CLCI-Net. Esto queda resumido en la métrica *F1 Score* de cada modelo; 0.63, 0.01 y 0.25 respectivamente.

Los cambios que se realizaron a SETR influyeron de manera significativa en el desempeño de la arquitectura base, la cual pasó de tener un *F1 Score* de 0.63 a 0.70 con un conjunto de datos de prueba y un 0.85 cuando se evalúa al modelo usando imágenes de mejor calidad provistas por el laboratorio español. Estos cambios corresponden a utilizar una capa de convolución 3D y alimentar la salida de la sección *Transformer* a una capa *fully connected*.

SETR resultó ser una mejor opción en comparación al modelo CLCI-Net porque este último sufría de un entrenamiento inaceptablemente lento debido a su sección LSTM, la cual no permitía una paralelización propia. Paralelización con la cual SETR sí contaba.

Finalmente, se prefiere SETR a U-Net para esta tarea debido a que U-Net usaba demasiada RAM perteneciente a la GPU. Esto se debe a la cantidad considerable de capas de convolución 3D que conforman U-Net, notablemente mayor al número de convoluciones que la versión propuesta de SETR posee.

4.2 Recomendaciones

El conjunto de datos es un elemento importante para replicar o mejorar este esfuerzo. En este trabajo se usaron imágenes de resonancia magnética de secuencia T2 principalmente, con cortes de 1 mm de separación en el espacio nativo. La separación de los cortes es importante porque el Núcleo Subtalámico en promedio tiene 5 mm de diámetro. Una separación mayor a 1 mm brinda muy pocos cortes con uno de los objetivos a segmentar y, en lo posible, es mejor evitar más interpolación de la necesaria.

Es muy probable que sea difícil conseguir una cardinalidad grande de resonancias magnéticas con estas características, entonces las técnicas de *data augmentation* expuestas en este trabajo pueden ser de utilidad. Se recomienda llegar al menos a una cardinalidad de 3000 ejemplos. Cualquier esfuerzo por generar un conjunto de datos mayor y compartirlo para la posteridad tomando en cuenta el consentimiento y anonimidad respectivo a la naturaleza delicada de este tipo de datos es bienvenido.

Algo importante a mencionar es que, en el paso de *data augmentation*, la técnica de rotación puede ocasionar problemas si no se procura que las funciones que realicen este procedimiento **no tengan interpolación**. De lo contrario, el valor que dicta a qué clase pertenece un vóxel pasa de ser un valor discreto a ser uno continuo. Esto podría derivar en problemas como clasificación difusa, en la cual no se ven *clusters* claramente definidos de vóxeles de una sola clase, sino que vóxeles de diferente clase están mezclados entre sí. Como sugerencia, se puede utilizar funciones de Numpy que sólo roten por ángulos rectos.

Otro punto importante por recalcar, es que las IRMs deben pasar del espacio nativo al espacio estándar para evitar problemas de inconsistencias entre el ángulo en que la

imagen fue capturada, la forma del cráneo, la forma del cerebro e incluso la posición del paciente. Esto se logra con funciones disponibles en la librería de Python ANTsPy y usando un *template* para hacer el registro usando las librerías mencionadas.

4.3 Trabajos Futuros

El dominio referente al ámbito médico tiene muchos desafíos que afrontar, soluciones a estos desafíos se pueden beneficiar claramente de incorporar conocimiento dentro de áreas de la computación.

Dentro de un marco práctico, un avance construido sobre lo encontrado durante este trabajo constituye la inclusión de una versión refinada de la solución en software utilizado en neuroimagen y neuronavegación. Mejor aún si es incorporado en sistemas disponibles para usuarios como radiólogos que puedan beneficiarse de herramientas de esta naturaleza una vez que logre pasar estándares debidos.

Dentro de un marco teórico/epistémico, las técnicas exploradas y personalizadas para la solución del problema en específico pueden extrapolarse a problemas médicos de segmentación similares; tales como la segmentación de lesiones vasculares, tumores, derrames, etc.

BIBLIOGRAFÍA

- [1] J. Sonne, V. Reddy, and M. R. Beato, *Neuroanatomy, Substantia Nigra*. 2020.
- [2] G. Leisman and R. Melillo, "The basal ganglia: Motor and cognitive relationships in a clinical neurobehavioral context," *Reviews in the Neurosciences*, vol. 24, 2013.
- [3] M. H. Mazurek, B. A. Cahn, M. M. Yuen, A. M. Prabhat, I. R. Chavva, J. T. Shah, A. L. Crawford, E. B. Welch, J. Rothberg, L. Sacolick, M. Poole, C. Wira, C. C. Matouk, A. Ward, N. Timario, A. Leasure, R. Beekman, T. J. Peng, J. Witsch, J. P. Antonios, G. J. Falcone, K. T. Gobeske, N. Petersen, J. Schindler, L. Sansing, E. J. Gilmore, D. Y. Hwang, J. A. Kim, A. Malhotra, G. Sze, M. S. Rosen, W. T. Kimberly, and K. N. Sheth, "Portable, bedside, low-field magnetic resonance imaging for evaluation of intracerebral hemorrhage," *Nature Communications*, vol. 12, 2021.
- [4] W. Benešová, "Segmentation of brain tumors from mri using adaptive thresholding and graph cut algorithm – vision & graphics group."
- [5] H. Yang, W. Huang, K. Qi, C. Li, X. Liu, M. Wang, H. Zheng, and S. Wang, "Clci-net: Cross-level fusion and context inference networks for lesion segmentation of chronic stroke," 2019.
- [6] A. M. Lozano, N. Lipsman, H. Bergman, P. Brown, S. Chabardes, J. W. Chang, K. Matthews, C. C. McIntyre, T. E. Schlaepfer, M. Schulder, Y. Temel, J. Volkmann, and J. K. Krauss, "Deep brain stimulation: current challenges and future directions," 2019.
- [7] K. G. Prakash, B. M. Bannur, M. D. Chavan, K. Saniya, S. S. Kumar, and A. Rajagopalan, "Neuroanatomical changes in parkinson's disease in relation to cognition: An update," *Journal of Advanced Pharmaceutical Technology and Research*, vol. 7, 2016.

- [8] M. Hutchinson and U. Raff, "Structural changes of the substantia nigra in parkinson's disease as revealed by mr imaging," *American Journal of Neuroradiology*, vol. 21, 2000.
- [9] M. Adachi, T. Hosoya, T. Haku, K. Yamaguchi, and T. Kawanami, "Evaluation of the substantia nigra in patients with parkinsonian syndrome accomplished using multishot diffusion-weighted mr imaging," *American Journal of Neuroradiology*, vol. 20, 1999.
- [10] S. Nunta-Aree, D. Tuntongtip, B. Sitthinamsuwan, P. Srirabheebhat, A. Nitising, and T. Witthiwej, "Accuracy of various subthalamic nucleus targeting methods and an appropriated formula for thai patients," *Journal of the Medical Association of Thailand*, vol. 93, 2010.
- [11] H. Braak, U. Rüb, and E. Braak, "Neuroanatomie des morbus parkinson," *Der Nervenarzt*, vol. 71, 2000.
- [12] H. Basinger and J. Joseph, *Neuroanatomy, Subthalamic Nucleus*. 2022.
- [13] J. Gili and J. Alonso, "Introducción biofísica a la resonancia magnética en neuroimagen con la colaboración del,"
- [14] W. V. Vargas, *Contribution to public health data processing architectures applied to resilient smart cities affected by natural hazards = Contribución a las arquitecturas de procesamiento de datos de salud pública aplicadas a ciudades inteligentes resilientes afectadas por peligros naturales*. PhD thesis, Escuela Técnica Superior de Ingenieros en Telecomunicación, 2019.
- [15] J. Juntu, J. Sijbers, D. V. Dyck, and J. Gielen, "Bias field correction for mri images," pp. 543–551, 10 2005.
- [16] L. G. Nyú and J. K. Udupa, "On standardizing the mr image intensity scale," *Magn Reson Med*, vol. 42, pp. 1072–1081, 1999.
- [17] J. V. Hajnal, D. L. Hill, and D. J. Hawkes, "Medical image registration," *Europhysics News*, vol. 31, pp. 5–8, 7 2000.

- [18] Y. J. Zhang, "A survey on evaluation methods for image segmentation," *Pattern Recognition*, vol. 29, pp. 1335–1346, 8 1996.
- [19] K. O'shea and R. Nash, "An introduction to convolutional neural networks,"
- [20] V. Iglovikov and A. Shvets, "Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation," 1 2018.
- [21] A. Orhan, "Skip connections as effective symmetry-breaking," 01 2017.
- [22] L. Zhu, C. Han, X. Liu, G. Gao, W. Liu, X. Wang, J. Zhang, and X. Zhang, "U-net deep learning network for automatic segmentation and localization of prostate cancer on mri apparent diffusion coefficient map," *Chinese Journal of Radiology (China)*, vol. 54, pp. 974–979, 10 2020.
- [23] L. Costantini, C. Nwafor, S. Lorenzi, A. Marrano, P. Ruffa, P. Moreno-Sanz, S. Raimondi, A. Schneider, I. Gribaudo, and M. Grando, "Caratterizzazione di varianti apirene di cultivar di vite," *7° Convegno Nazionale di Viticoltura, Piacenza, 9-11 luglio 2018*, p. p.43, 2018.
- [24] M. Phi, "Illustrated guide to lstm's and gru's: A step by step explanation."
- [25] S. Zheng, J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang, Y. Fu, J. Feng, T. Xiang, P. H. S. Torr, and L. Zhang, "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers,"
- [26] A. Gillioz, J. Casas, E. Mugellini, and O. Abou Khaled, "Overview of the transformer-based models for nlp tasks," pp. 179–183, 09 2020.
- [27] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale,"
- [28] R. Strudel, R. Garcia, I. L. Inria, and C. S. Inria, "Segmenter: Transformer for semantic segmentation,"
- [29] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *European conference on information retrieval*, pp. 345–359, Springer, 2005.

- [30] H. Li, G. Jiang, J. Zhang, R. Wang, Z. Wang, W.-S. Zheng, and B. Menze, “Fully convolutional network ensembles for white matter hyperintensities segmentation in mr images,” *NeuroImage*, vol. 183, pp. 650–665, 2018.