

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ingeniería en Electricidad y Computación

Segmentación autónoma de lesiones producidas por ultrasonido focalizado como tratamiento a enfermedades con desórdenes de movimiento, utilizando técnicas de inteligencia artificial

PROYECTO INTEGRADOR

Previo la obtención del Título de:

Ingeniero en Ciencias de la Computación

Presentado por:

Bryan Paúl Alava Calderón

GUAYAQUIL - ECUADOR

AÑO: 2023

Dedicatoria

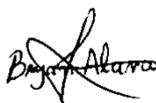
El presente proyecto se lo dedico a mis padres, por confiar en mí en todo momento y darme ánimos en momentos difíciles.

Agradecimientos

Agradezco principalmente a mi familia que nunca me dejó solo desde que comencé este camino; a mis amigos de primer semestre que aún siguen conmigo desde el primer día y a los cuales los considero casi hermanos; a mis amigos de la carrera con los cuales compartí conocimientos y habilidades. Y, por último, a los docentes que fueron la guía para formarme como profesional.

Declaración expresa

“Los derechos de titularidad y explotación, me corresponde conforme al reglamento de propiedad intelectual de la institución; Bryan Paúl Alava Calderón doy mi consentimiento para que la ESPOL 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”

A handwritten signature in black ink, appearing to read "Bryan Alava", with a stylized flourish at the end.

Bryan Alava

Evaluadores

Miguel Andrés Realpe Robalino Ph.D.

Profesor de la materia

Colón Enrique Peláez Jarrín Ph.D.

Profesor tutor

Resumen

El tratamiento con ultrasonido focalizado es una intervención quirúrgica no invasiva, que mejora los efectos motores de enfermedades como el Parkinson, Disquinesias y Temblor Esencial. Este tratamiento produce una lesión en el Núcleo Subtalámico, que requiere de una identificación precisa en el menor tiempo posible. Se presenta una herramienta web que permite segmentar de forma autónoma las lesiones presentes en el Núcleo Subtalámico, a partir del análisis y procesamiento de imágenes de resonancia magnética (MRI) en conjunto con la aplicación de técnicas de aprendizaje profundo. El módulo de aprendizaje profundo emplea como backbone la arquitectura Swin Transformers y se basa en una arquitectura Unet, el cual fue entrenado a partir de un conjunto de datos 4000 ejemplos de imágenes potenciadas en T2 con cortes a 1mm de distancia. La propuesta basada en transformers obtuvo la segmentación de los volúmenes de interés con un coeficiente DICE del 0,80 y AVD (Average Volume Distance) del 0,0019. Se contrastó estos resultados con otras arquitecturas, tales como CCT (Compact Convolutional Transformer) y MobileViT las cuales no obtuvieron un desempeño a nivel de la arquitectura Swin Transformer. El módulo que corresponde al prototipo acopla el modelo de segmentación con una interfaz web que permite subir una MRI, procesarla y devolver una máscara de segmentación superpuesta, además de proveer un mecanismo de retroalimentación para la evaluación por parte de personal administrativo.

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

Abstract

Focused ultrasound treatment is a non-invasive surgical intervention that improves motor effects in diseases such as Parkinson's, Dyskinesias, and Essential Tremor. This treatment induces a lesion in the Subthalamic Nucleus, requiring precise identification in the shortest possible time. A web-based tool is presented, allowing for autonomous segmentation of lesions within the Subthalamic Nucleus, based on the analysis and processing of magnetic resonance imaging (MRI) data in conjunction with deep learning techniques. The deep learning module uses the Swin Transformers architecture as its backbone and is based on a Unet architecture, which was trained on a dataset of 4000 examples of T2-weighted enhanced images with 1mm slice spacing. The transformer-based proposal achieved segmentation of volumes of interest with a DICE coefficient of 0.80 and an Average Volume Distance (AVD) of 0.0019. These results were compared to other architectures, such as Compact Convolutional Transformer (CCT) and MobileViT, which did not perform at the level of the Swin Transformer architecture. The prototype module integrates the segmentation model with a web interface that allows users to upload an MRI, process it, and return an overlaid segmentation mask, in addition to providing a feedback mechanism for evaluation by administrative personnel.

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

Índice general

Resumen	6
Abstract.....	7
Índice general	8
Abreviaturas.....	10
Índice de figuras.....	11
Índice de tablas.....	11
Capítulo 1	12
1.1 Introducción.....	13
1.2 Descripción del problema.....	15
1.3 Justificación del problema	16
1.4 Descripción del problema	16
1.4.1 Objetivo general	16
1.4.2 Objetivos específicos	16
1.5 Marco teórico	17
Capítulo 2	27
2.1 Metodología	28
2.2 Análisis de datos	28
2.3 Pre-procesamiento del conjunto de datos	30
2.3.1 Corrección del sesgo	30
2.3.2 Normalización de intensidad.....	30
2.3.3 Realineamiento de la imagen	30
2.3.4 Selección de cortes	31
2.4 Data augmentation	32
2.4.1 ZCA.....	32
2.4.2 Rotaciones	33
2.5 Entrenamiento del modelo.....	34
2.5.1 Adaptación de modelos	34

2.5.2 Entrenamiento	35
2.6 Métricas de evaluación.....	36
2.6.1 Precision.....	36
2.6.2 Recall	36
2.6.3 Dice	37
2.6.4 F1-score	37
2.6.5 Distancia de Hausdorff (HD)	37
2.6.6 Distancia promedio de volumen (AVD)	37
2.7 Implementación de prototipo web.....	38
2.7.1 Requerimientos funcionales.....	38
2.7.2 Requerimientos no funcionales.....	38
2.7.3 Diseño del prototipo.....	38
2.7.4 Elección de tecnología.....	39
Capítulo 3	41
3.1 Resultados y análisis	42
3.1.1 Modelos base.....	42
3.1.2 Dataset mejorado.....	45
3.1.3 Modelo mejorado	45
3.1.4 Prototipo.....	52
Capítulo 4	55
4.1 Conclusiones y recomendaciones.....	56
4.1.1 Conclusiones	56
4.1.2 Recomendaciones	56
Bibliografía.....	58

Abreviaturas

ESPOL Escuela Superior Politécnica del Litoral

MRI Imágenes de resonancia magnética

Índice de figuras

Figura 1. 1 - Imagen de resonancia magnética de la lesión causada por HIFU.	14
Figura 1. 2 - Núcleo Subtalámico (Línea Roja) y Ganglios basales (Líneas Azules) [24].	19
Figura 1. 3 - (a) Imagen de MR en secuencia T1 vs (b) Imagen de MR en secuencia T2.	20
Figura 1. 4 - Arquitectura Multi escala de CNN's que se empleó para la clasificación de nódulos malignos y benignos en los pulmones [13].	23
Figura 1. 5 - Arquitectura del modelo Vision Transformer (ViT) [35].	24
Figura 1. 6 - Arquitectura MobileViT [41]	25
Figura 1. 7 - Arquitectura CCT [42].	25
Figura 1. 8 - Arquitectura Swin Transformer [43]	26
Figura 2.1 - Flujo de datos de este estudio [autoría propia]28	
Figura 2. 2 - Distribución de áreas y volúmenes de las lesiones y edemas [autoría propia].....	29
Figura 2. 3 - Distribución de área y volúmenes de lesiones y edemas [autoría propia].....	29
Figura 2. 4 - Ejemplo de la región seleccionada [autoría propia].	32
Figura 2. 8 - Prototipo inicial.....	39
Figura 3. 1 - Resultado de métricas del dataset de entrenamiento en los modelos base [autoría propia].....	43
Figura 3. 2 - Resultado de métricas del dataset de validación en los modelos base [autoría propia].....	44
Figura 3. 3 - Arquitectura mejorada para el modelo base de Swin Transformer [autoría propia].....	46
Figura 3. 4 - Resultado de métricas del dataset de entrenamiento del modelo Swin Transformer base vs Swin Transformer modificado [autoría propia].	47
Figura 3. 5 - Resultado de métricas del dataset de validación del modelo Swin Transformer base vs Swin Transformer modificado [autoría propia].....	48

Índice de tablas

Tabla 3. 1 - Resultados de las métricas para el dataset de pruebas [autoría propia].	49
Tabla 3. 2 - Matriz de confusión del modelo Swin Transformer [autoría propia].	50

Capítulo 1

1.1 Introducción

Las enfermedades que provocan desórdenes de movimiento como la Enfermedad de Parkinson (EP), el Temblor esencial y las Disquinesias afectan notablemente la movilidad de las personas evitando que puedan realizar gran parte de sus actividades diarias [1]. Estas enfermedades dependen muchas veces de factores genéticos o hereditarios [1], o se producen por el uso de medicamentos para tratarlas [2,3]. Los síntomas más comunes de estas enfermedades suelen manifestarse a través de movimientos involuntarios, rigidez o lentitud, producidos por las alteraciones en el cerebro y el control motor [1].

Las imágenes de resonancia magnética (MRI, por sus siglas en inglés) son una valiosa técnica que se emplea en el área de la medicina que ayuda a la visualización de estructuras anatómicas del cuerpo humano [4]. Existen diferentes secuencias en los tipos de resonancia magnética, tales como: T1, T2 y FLAIR (FLuid-Attenuated Inversion Recovery por sus siglas en inglés) [4,5]. La secuencia T1 se usa para resaltar tejidos grasos, como el cerebro o tejidos blandos, como músculos. La secuencia T2, por otro lado, se utiliza para resaltar los líquidos, como el líquido cefalorraquídeo en el cerebro. O, la secuencia FLAIR que permite detectar lesiones o anomalías en el cerebro al suprimir la señal del líquido cefalorraquídeo y resaltar las alteraciones en los tejidos circundantes [4,6,7].

Existen múltiples estructuras ubicadas en el cerebro relacionadas con el control motor, una de ellas es el Núcleo Subtalámico (NST). El cual juega un papel esencial en la coordinación y el control de los movimientos, así como en el equilibrio y la postura [8,9]. Razón por la que se ha convertido en el foco actual de muchos tratamientos aplicados a pacientes en las últimas décadas. Como el uso de fármacos (L-DOPA en pacientes con Parkinson), hasta intervenciones quirúrgicas invasivas y no invasivas que causan lesiones a fin de reducir los síntomas de las enfermedades que padecen [3,10,11].

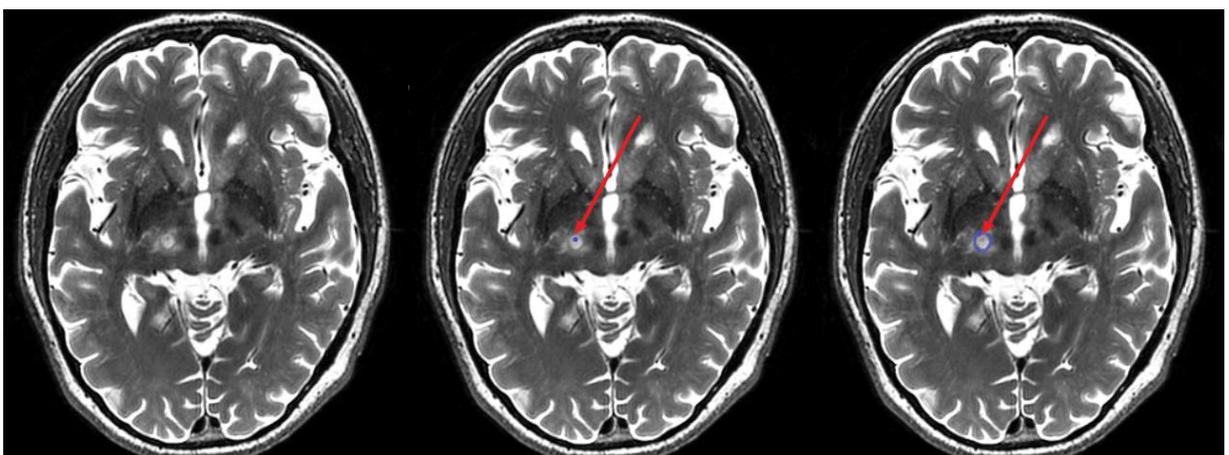
Entre los tipos de lesiones causadas por los tratamientos quirúrgicos están las de tipo Termolíticas y por HIFU (High Intensity Frequency Ultrasound). Las lesiones Termolíticas son

métodos invasivos y conllevan una gran cantidad de riesgos, la técnica más común utilizada es la termoablación por radiofrecuencia donde se emplea una sonda que genera calor focalizado en la región del NST causando una lesión y edema (tejido necrótico alrededor de la lesión) [3]. Por otro lado, la lesión por HIFU es un método mínimamente invasivo guiado por Resonancia Magnética (MR), que consiste en focalizar alrededor de 1000 haces de ultrasonido en la diana localizada generalmente en el NST, aumentando su temperatura y produciendo una lesión mínima junto a un edema. Siendo este último, considerado actualmente como uno de los tratamientos más eficaces en contra de las enfermedades que provocan desórdenes en el movimiento [11,12].

En la Figura 1.1 se puede observar el resultado del tratamiento con HIFU en un paciente, en la MRI de la izquierda a simple vista no es perceptible la lesión ni el edema, por otro lado, en la imagen central se observar un pequeño punto azul que representa a la lesión y por último en la imagen de la derecha se aprecia una región azul delimitada que corresponde al tamaño del edema.

Figura 1.1 – Imagen de resonancia magnética de la lesión causada por HIFU.

[autoría propia]



El aprendizaje profundo empleado en el campo de la medicina ha sido de gran ayuda, especialmente en tareas de análisis de imágenes médicas. Esta tecnología se ha utilizado para

la clasificación de imágenes y su aplicación en tratamientos, para la predicción del avance de enfermedades, para el diagnóstico de enfermedades a través de datos clínicos, entre otros [8,13,14].

1.2 Descripción del problema

La lesión y el edema provocados en el NST en pacientes de Parkinson intervenidos por HIFU, produce la reducción de los temblores involuntarios causados por la enfermedad. Sin embargo, identificar de forma precisa la tanto la lesión como el edema del NST y las condiciones en las cuales fueron realizados no es una tarea sencilla. Adicionalmente existen aspectos asociados a la lesión y el edema que dificultan aún más la tarea de segmentación, entre los que destacan la forma, el tamaño, volumen, ubicación, ya que a partir de estos se realizan análisis más preciso del tratamiento que debe recibir un paciente.

Resulta desafiante realizar la identificación y ubicación de las lesiones en esta área del cerebro de forma precisa, conocido como segmentación de la lesión [15,16] , debido a las variaciones anatómicas y estructurales de las lesiones. Aunque actualmente existen métodos basados en inteligencia artificial (IA) que se utilizan para evaluar y segmentar lesiones en el cerebro a través del análisis de imágenes de Resonancia Magnética [16, 17], no hay herramientas que realicen la segmentación de las lesiones en el NST de forma autónoma. Lo que implica que personal médico experto en el tema deba identificar, evaluar y obtener la segmentación de la lesión el edema de forma manual y lo más precisa posible, lo que a su vez consume tiempo y recursos durante todo el procedimiento. Cabe destacar que este proceso ocurre no solo al inicio de la operación sino a lo largo de la evolución de los pacientes con enfermedades de desórdenes de movimiento, esto con el fin de poder realizar ajustes en el tratamiento postquirúrgico.

1.3 Justificación del problema

Contar con una herramienta que permita segmentar las lesiones en el NST de forma autónoma a través del análisis de imágenes MR, se podría mejorar la precisión en la identificación y ubicación de las lesiones en pacientes con enfermedades que provocan desórdenes en el movimiento sometidos a HIFU, lo que a su vez aportaría en la disminución del tiempo y recursos de identificación y segmentación realizado por parte del personal médico.

Además, la segmentación autónoma de las lesiones facilitaría el monitoreo de la efectividad del tratamiento realizado a los pacientes, logrando contribuir a la mejora de la planificación del tratamiento postquirúrgico y a la optimización de la respuesta farmacológica de forma eficaz y oportuna en relación con los efectos que el tratamiento por HIFU produce.

1.4 Descripción del problema

1.4.1 Objetivo general

Implementar una herramienta para la segmentación autónoma de la lesión en el Núcleo Subtalámico a partir del análisis de imágenes de resonancia magnética utilizando técnicas de aprendizaje profundo.

1.4.2 Objetivos específicos

- Aplicar técnicas de pre-procesamiento de imágenes de resonancia magnética de pacientes con enfermedades de desórdenes de movimiento para la extracción de características de interés.
- Diseñar la arquitectura de un modelo de segmentación, basado en aprendizaje autónomo profundo para la identificación eficaz de la región del edema (lesión) en el Núcleo Subtalámico.
- Evaluar el desempeño del modelo de segmentación de la lesión en el Núcleo Subtalámico, para determinar su eficacia y precisión.

- Desarrollar e implementar un prototipo de la herramienta basada en una página web, con el modelo de segmentación, que facilite la carga de imágenes de resonancia magnética, para segmentar el edema en el Núcleo Subtalámico.

1.5 Marco teórico

1.5.1 Desórdenes de movimiento

Los trastornos del movimiento son afecciones que afectan la velocidad, el flujo, la calidad y la coordinación de los movimientos corporales. Estos trastornos pueden manifestarse como movimientos involuntarios, anormales o excesivos, así como también como dificultades para iniciar o controlar los movimientos voluntarios [1].

Estos trastornos del movimiento pueden tener un impacto significativo en la calidad de vida de las personas afectadas. Entre estos se destacan la enfermedad de Parkinson, el Temblor Esencial y las Disquinesias.

1.5.1.1 Parkinson

La EP es uno de los trastornos neurológicos más comunes en la actualidad afectando a cerca del 1% de la población de adultos mayores a 60 años [18]. Y se caracteriza por producir cuadros de enlentecimiento (bradicinesia), temblores, rigidez, entre otros síntomas [1,19]. La EP no tiene cura, pero existen tratamientos disponibles para ayudar a mitigar los síntomas. Estos pueden incluir medicamentos para aumentar los niveles de dopamina en el cerebro, terapia física y ocupacional para mejorar la movilidad y la calidad de vida, y en algunos casos, cirugía cerebral [1].

1.5.1.2 Temblor esencial

El Temblor Esencial es un trastorno neurológico común que se caracteriza por temblores involuntarios y rítmicos en diferentes partes del cuerpo, como las manos, los brazos, la cabeza

o la voz. Es una enfermedad que no se puede curar, sin embargo, hay medidas que se pueden tomar con el fin para reducir la gravedad de los temblores y mejorar la calidad de vida [12,20].

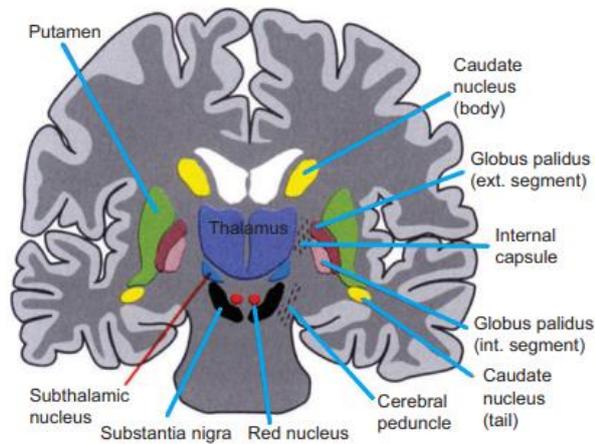
1.5.1.3 Disquinesias

Las disquinesias, también conocidas como discinesias son trastornos del movimiento caracterizados por movimientos involuntarios y anormales [21]. De forma general se manifiesta como movimientos rápidos y repetitivos (discinesias hiperquinéticas) [22]. Este tipo de trastorno es causado por diferentes factores, como efectos secundarios de algunos medicamentos, lesiones en el sistema nervioso, y una de las más comunes y difíciles de tratar causa del tratamiento con Levodopa que reciben los pacientes con EP, donde afecta aproximadamente al 40% de los pacientes [21].

1.5.2 Núcleo Subtalámico – NST

El NST es una estructura de crucial importancia en el control motor junto al resto de órganos que conforman el circuito de ganglios basales encargados de la coordinación del movimiento y otras funciones cognitivas [23,24]. Por lo que actualmente es considerada como la diana quirúrgica en el estudio y tratamiento de enfermedades como el Parkinson, el temblor esencial y discinesias [23,25,26]. Estudios recientes muestran que, al realizarse intervenciones quirúrgicas en esta zona, se evidencia mejoría en el control de los temblores musculares involuntarios en pacientes a lo largo del tiempo y disminución de los síntomas de la enfermedad que padecen [3,10,11].

Figura 1. 2 - Núcleo Subtalámico (Línea Roja) y Ganglios basales (Líneas Azules) [24].

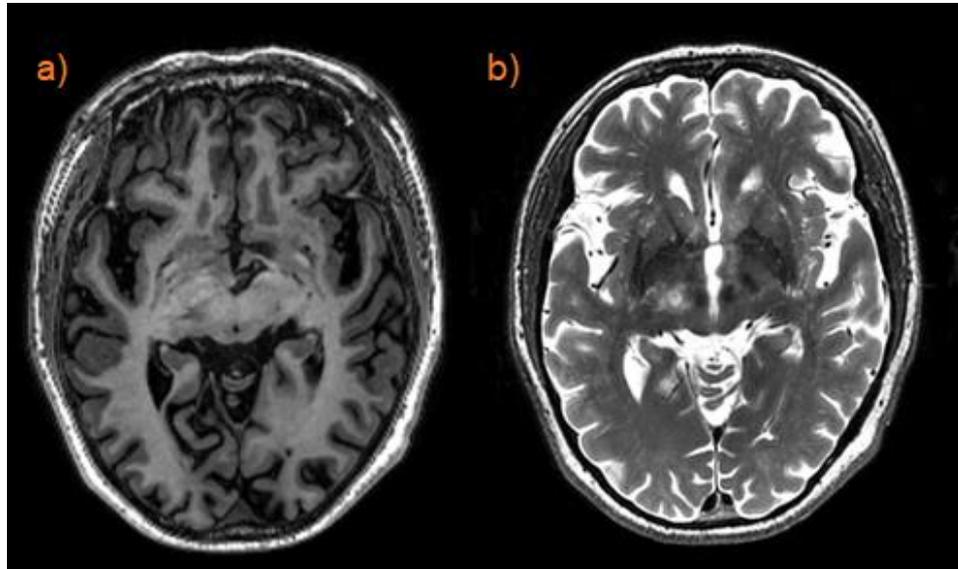


1.5.3 MRI

Las MRI son una de las herramientas más útiles dentro del campo de la medicina, siendo una forma no invasiva en la que el médico examina los órganos, los tejidos y el sistema esquelético, lo que a su vez le permite visualizar y evaluar posibles patologías en estas estructuras internas. Durante el proceso de adquisición de las MRI, los campos magnéticos interactúan con los protones presentes en el cuerpo, generando señales de radio que son detectadas y procesadas para producir imágenes de alta resolución [4,7], tal como se muestra en la Figura 1.3.

Figura 1. 3 – (a) Imagen de MR en secuencia T1 vs (b) Imagen de MR en secuencia T2

[autoría propia].



1.5.4 Técnicas de Pre-procesamiento

Para desarrollar la tarea de segmentación es necesario emplear imágenes que estén estandarizadas de acuerdo con lo que el modelo necesita y espera como entrada y así poder generalizar a partir de los ejemplos; sin embargo, se debe contar con una cantidad suficiente de imágenes para lograr este propósito.

Esto se logra aplicando técnicas de limpieza, corrección de datos, como corrección del sesgo (Bias field correction) y registro de imagen. Y para aumentar la cantidad de datos de imágenes médicas se aplican rotaciones, generalmente en ángulos rectos, volteos de imagen verticales y horizontales, entre otros [8] que serán tratados en el siguiente capítulo.

1.5.4.1 Corrección del sesgo

Durante la captura de imágenes de resonancia magnética pueden producirse 2 tipos de perturbaciones. La primera es inherente a la máquina de MR, la cual es causada por las limitaciones del equipo con el que se realizan las MRI [27]; y la segunda está asociada a las interacciones del paciente, como la respiración o ligeros movimientos de la cabeza, durante el proceso de generación de la imagen, o incluso implantes metálicos [27].

Aunque no siempre es visible para un observador humano, este sesgo puede causar clasificaciones erróneas cuando se utilizan técnicas de segmentación basadas en la intensidad [24,28]. Por lo que, con el fin de obtener mejores resultados en la segmentación esto se lo realiza a través de técnicas estimación de los valores de intensidad en las MRIs [28], permitiendo corregir la no uniformidad de estos valores en la imagen, lo que garantiza análisis más precisos de las MRIs [24].

1.5.4.2 Normalización de intensidad

Las MRIs cuentan con variaciones en las intensidades de los vóxeles generalmente, y esto se debe a que no todas las imágenes son capturadas de la misma forma, ya que existen factores asociados a su adquisición que pueden cambiar, como el tipo de resonador o los parámetros de ajuste de este [29,30]. De esta forma se produce un sesgo adicional, que impide analizar las imágenes de forma general si se cuenta con diferencias significativas en el nivel de detalle de los tejidos y otras regiones del cerebro.

Por lo cual es necesario emplear técnicas de normalización que permiten ajustar o escalar los valores de intensidad de los vóxeles presentes en una imagen, para que estos sigan una distribución de intensidad deseada o se mantengan dentro de un rango específico [29,30].

1.5.4.3 Realineamiento de imagen

Otro problema que se enfrenta cuando se trabaja con MRI es que no todas las personas poseen un cerebro con iguales dimensiones, sumado a que durante el proceso de captura de las imágenes la posición de la cabeza de la persona influye en la calidad de las imágenes generando que no sean comparables entre sí [31,32,33].

Esto se soluciona mediante la aplicación de técnicas de realineamiento de la imagen de forma automática o manual. El realineamiento automático consiste en aplicar una transformación geométrica a partir de una plantilla predefinida generando valores de rotación, traslación e

interpolación entre los cortes de la MRI [31,32]. El realineamiento manual consiste en tener de la misma forma una plantilla predefinida, y obtener los datos de origen y dirección de esta imagen para poder colocárselos a la MRI [33]. Con ambos métodos se logra conseguir una nueva imagen resultante que permite compararla con el resto de las imágenes de un dataset.

1.5.5 Análisis de imágenes médicas mediante aprendizaje profundo

Los algoritmos de aprendizaje profundo se fundamentan en redes neuronales artificiales que cuentan con la capacidad de aprender relaciones no lineales [14,34]. Estos, han demostrado grandes avances en cuanto al análisis de imágenes de resonancia magnética [13,14,37-39].

Para la realización de este trabajo se analizaron dos tipos de arquitecturas que se utilizaron para el análisis de imágenes médicas [13,35], y que a su vez fueron tema de investigación en trabajos pasados para la segmentación del NST [8]. Los tipos de arquitecturas empleadas fueron las Redes Neuronales Convolucionales (CNN) y los Transformers [8].

1.5.5.1 CNN

Es una de las arquitecturas más comunes en aprendizaje profundo diseñadas para visión artificial, y están compuestas de capas convolucionales, capas de reducción (Pooling) y capas completamente conectadas (Fully Connected Layers) [36,40]. Estas redes están diseñadas para recibir como entrada imágenes, asignando importancia (pesos y sesgos aprendibles) [40] a varios aspectos/objetos pertenecientes a cada imagen de un conjunto de datos, donde puede usarse algún tipo de etiqueta que delimite, ya sea una región de interés o algún patrón específico que se desee que aprenda la red [5,36,37,38]. Este proceso lo realizan a través de bloques encoder y decoder, donde el primero se encarga de crear mapas de características a través de las imágenes que recibe como entrada y el segundo se encarga de reconstruir las imágenes a partir de esos mapas, creando la segmentación final [38].

Debido a su capacidad de análisis de patrones complejos en imágenes, estas arquitecturas han sido empleadas para realizar clasificaciones y segmentaciones a partir de imágenes de resonancia magnética, y otros tipos de imágenes médicas [13, 38, 39], obteniendo grandes resultados de precisión y efectividad en el desarrollo de dichas tareas [13].

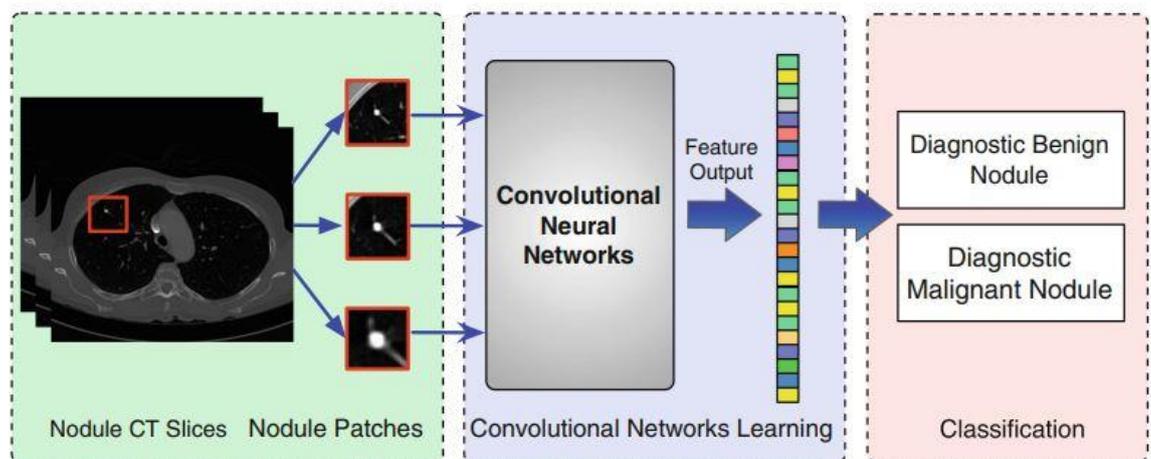


Figura 1. 4 - Arquitectura Multi escala de CNN's que se empleó para la clasificación de nódulos malignos y benignos en los pulmones [13].

1.5.5.2 Transformers

Los Transformers son arquitecturas de redes neuronales que se basan en una configuración de encoder y decoder, o en redes neuronales recurrentes [35]. Estas arquitecturas están diseñadas con un enfoque basado en mecanismos de atención para el procesamiento de secuencias [35]. La ventaja de utilizar mecanismos de atención es que eliminan la necesidad de capas recurrentes en modelos de procesamiento de lenguaje natural [35]. Aunque estas arquitecturas fueron creadas para procesamiento de texto [35], están siendo utilizadas en el desarrollo de tareas de clasificación y segmentación de imágenes. Por ejemplo, la arquitectura llamada "Vision Transformer" (ViT) utiliza una versión modificada de los Transformers y los aplica directamente a imágenes, este modelo utiliza una estrategia de descomposición de la imagen en

lotes y luego los transforma en secuencias para ser procesados, similar al texto en el modelo Transformer [35].

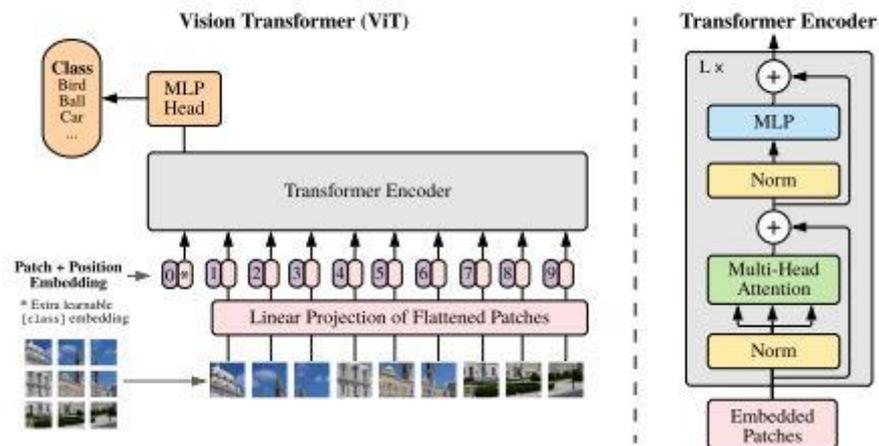
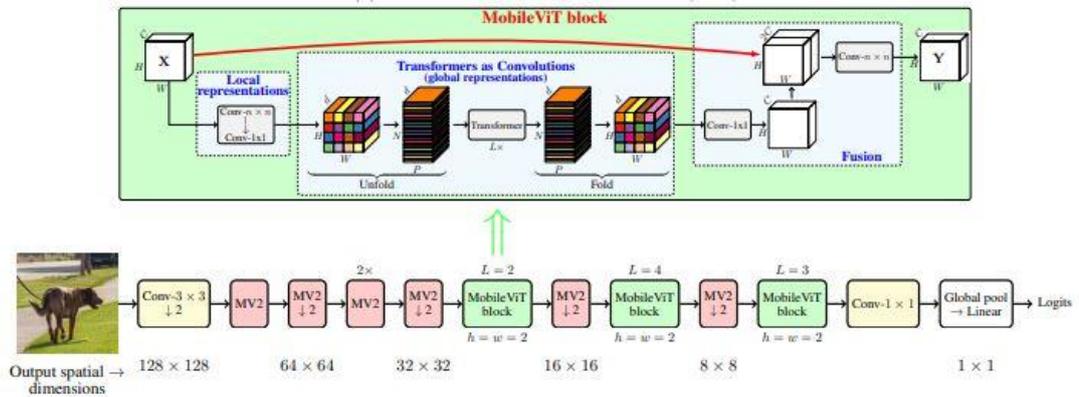


Figura 1. 5 - Arquitectura del modelo Vision Transformer (ViT) [35].

1.5.5.3 MobileViT

MobileViT es una variante del ViT que destaca por reducir el coste computacional del modelo inicial siendo considerada una versión más ligera al reducir el tamaño de las capas y de los bloques de atención, y tal como su nombre lo indica es una arquitectura pensada principalmente para dispositivos móviles o con pocos recursos computacionales [41]. La creación de este modelo parte de modelos CNN ligeros tal como Mobilenetv1, Mobilenetv2 o MixNet. Sin embargo, la comparativa de estos con MobileViT muestra que MobileViT brinda mejores resultados de generalización que los modelos basados en CNN [41]. Este modelo consta de bloques CNN y bloques MobileViT que reemplazan el procesamiento local en convoluciones con procesamiento global mediante Transformers tal como se muestra en la Figura 1.6 obteniendo características de los modelos CNN y Transformers [41].

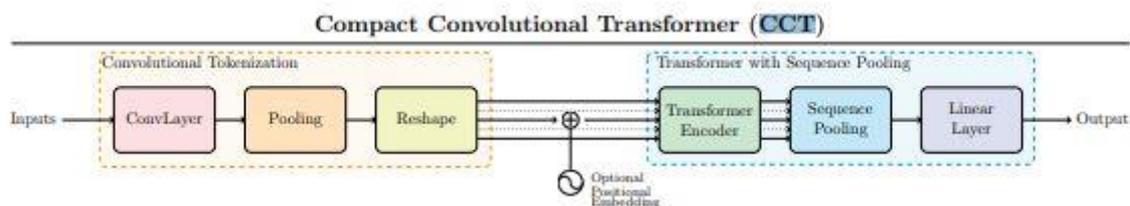
Figura 1. 6 – Arquitectura MobileViT [41]



1.5.5.4 Compact Convolutional Transformers (CCT)

CCT es otra arquitectura que combina los Transformers con CNN para lograr mayor eficiencia y rendimiento en tareas de generalización a partir de los datos, permitiendo que a través de los bloques convolucionales se capture de mejor forma las características locales de las imágenes [42]. La arquitectura se muestra a continuación en la Figura 1.7:

Figura 1. 7 – Arquitectura CCT [42]



A diferencia del modelo ViT esta arquitectura muestra una forma diferente de preprocesar los datos de entrada mediante un bloque convolucional en vez de usar bloques de atención [42].

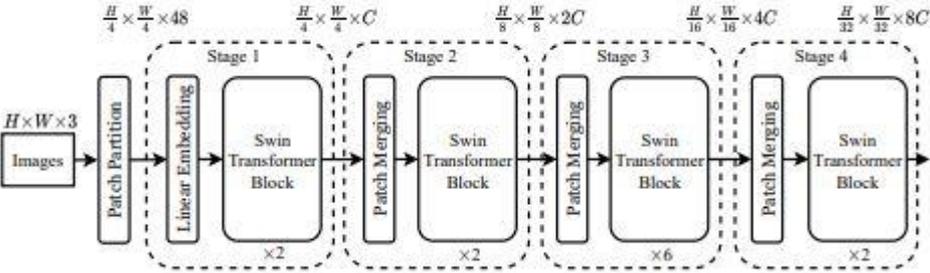
1.5.5.5 Swin Transformers

Swin Transformers es una arquitectura que se basa en ViT la cual consiste en modificar el bloque de atención que posee el ViT por una autoatención basada en ventana deslizante, lo que quiere decir que se reduce la complejidad y mejora la eficiencia al no aplicar el sistema de

atención a la imagen completa, sino que la divide la imagen en bloques más pequeños y lo agrupa en diferentes niveles de jerarquía [43].

La arquitectura del Swin Transformer se muestra a continuación en la Figura 1.8, la cual tiene una estructura jerárquica, en la que cada uno los niveles jerárquicos poseen bloques de atención y transformaciones lineales para procesar las imágenes de entrada [43]. Los bloques que se encuentran más abajo en la jerarquía (izquierda de la Figura 1.8) procesan parches de imágenes de mayor tamaño en comparación a los bloques superiores que procesan parches de imágenes mucho más pequeños (derecha de la Figura 1.8) [43]:

Figura 1. 8 – Arquitectura Swin Transformer [43]

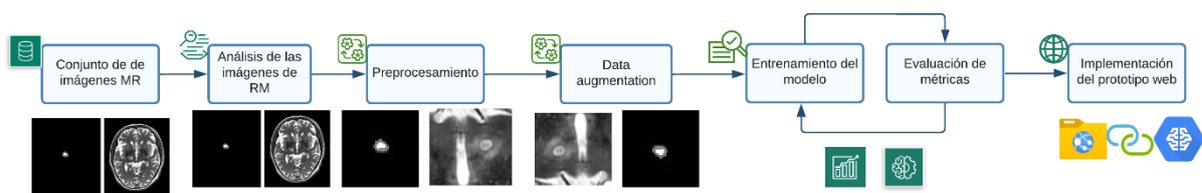


Capítulo 2

2.1 Metodología

En la Figura 2.1 se observa la división realizada para la metodología de este estudio, constando de 6 etapas que son: análisis de datos, preprocesamiento, data augmentation, diseño de un modelo de aprendizaje profundo, evaluación de métricas y parámetros, desarrollo de prototipo funcional.

Figura 2.1- Flujo de datos de este estudio [autoría propia]



2.2 Análisis de datos

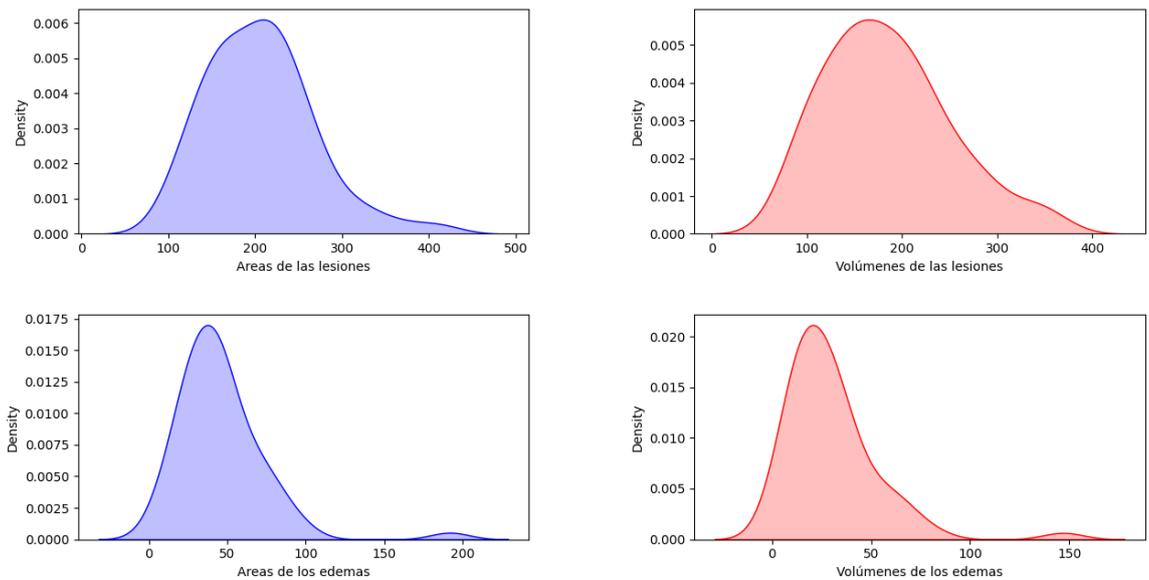
La primera etapa de la metodología realizada en este trabajo consistió en analizar el conjunto de imágenes de MR brindadas por parte del cliente, Francis Loayza, PhD. Dentro de este conjunto se tenía un total de 69 MRI's tomadas a pacientes en secuencia T1 y T2, así como las segmentaciones realizadas manualmente (Ground Truth) realizadas por expertos.

Conociendo las diferencias entre las secuencias de las MRI's se decidió emplear MRI's en secuencia T2 debido a la naturaleza de la lesión causada, donde la intensidad y el contraste que se tiene en estas imágenes permiten visualizar tejidos blandos con alto contraste mostrando con mayor claridad la región del NST en donde se producen las lesiones.

Adicionalmente se aplicó dos filtros a este conjunto de datos que consistieron en examinar que las MRI's en secuencia T2 tuvieran un Ground Truth de las lesiones para luego verificar que no muestren un área y volumen anormales, obteniendo un total de 60 MRI's. En la Figura 2.2 y

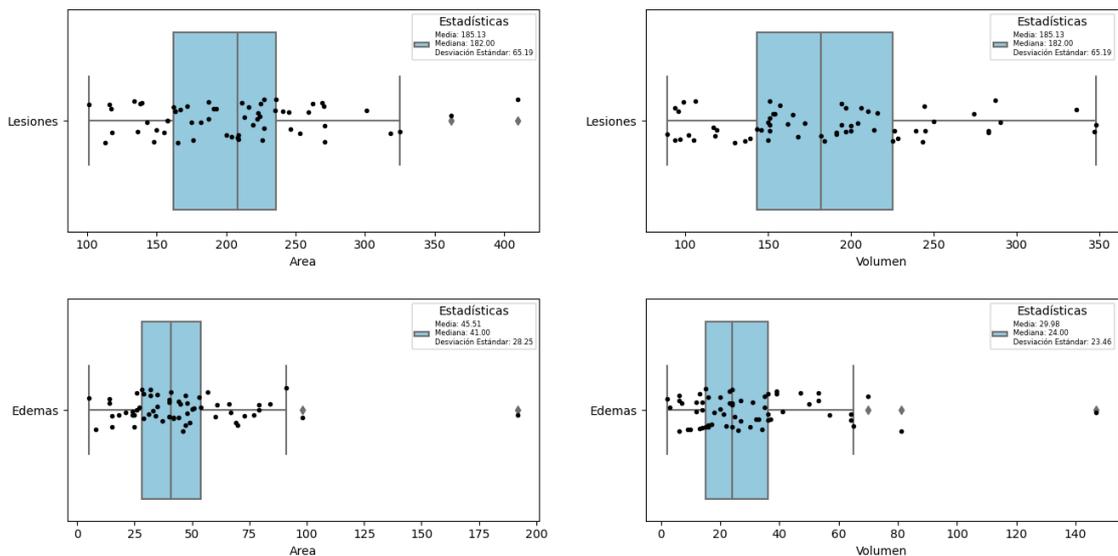
Figura 2.3 se observa la distribución obtenida del total de MRI's que se consiguieron luego de aplicar estos dos filtros.

Figura 2. 2 – Distribución de áreas y volúmenes de las lesiones y edemas [autoría propia]



Las imágenes muestran una distribución simétrica negativa para todas las gráficas, para el área y volumen de las lesiones se muestra una menor variabilidad de los datos obtenidos. Por otro lado en el área y volumen de los edemas existe una mayor variabilidad en los datos obtenidos.

Figura 2. 3 - Distribución de área y volúmenes de lesiones y edemas [autoría propia]



La Figura 2.3 muestra 4 diagramas de cajas del área y volumen de las lesiones y edemas, para las lesiones el área se muestra dentro del rango de 100 a 400 con una media y mediana

que indican que los datos se encuentran alineados hacia la derecha y para el volumen los datos se encuentran en el rango de 100 a 350 donde su media y mediana indican que la distribución de datos está hacia la izquierda. Por otro lado, para los edemas el área se muestra datos dentro del rango de valores de 5 hasta 100 con una media y mediana que indican , y para el volumen se muestran valores dentro del rango de 5 hasta 65 con más valores atípicos.

2.3 Pre-procesamiento del conjunto de datos

Para el pre-procesamiento de las MRI mencionadas en la sección 1.5, se realizaron los pasos que se muestran a continuación, tomando en cuenta la relación espacial de vóxeles (spacing) y la orientación de las imágenes:

2.3.1 Corrección del sesgo

Se empleó la corrección de sesgo por medio del algoritmo N4 de la biblioteca ANTsPy [44] que permitió resaltar detalles oscuros de las MRIs.

2.3.2 Normalización de intensidad

Para este paso se utilizó la librería de Intensity normalization [45] de Python para corregir los valores de intensidad de los vóxeles de las MRI.

2.3.3 Realineamiento de la imagen

Para este paso se usó un template para el análisis de las MRI's ya que cuenta con la característica de estar orientada con respecto a la comisura anterior y posterior o ACPC, lo que permite alinear las MRI's gracias a que ambas comisuras son puntos de referencia anatómicos que conectan diferentes áreas del cerebro entre los hemisferios derecho e izquierdo [46]. De forma manual se tomó la dirección y el origen para poder colocárselos a cada sujeto del conjunto de datos, alineando cada imagen en un mismo marco de referencia.

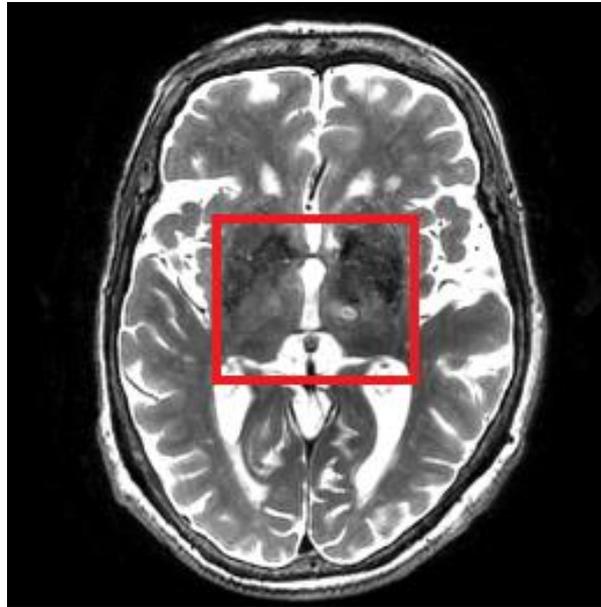
2.3.4 Selección de cortes

Para este paso se necesitan dos consideraciones importantes, la primera relacionada con el área de análisis, esto es un cubo con dimensiones similares para el eje "x" y "y" necesario para el data augmentation, la segunda consideración, verificar que se pudiese apreciar correctamente la lesión y el edema. La selección de los cortes fue realizada a través de la visualización 2D de las lesiones presentes en las MRIs, determinando que para todos los ejes se utilizarían valores máximos y medios de sus respectivas dimensiones.

Primero se procedió a verificar que cortes eran necesarios para el análisis, para esto se realizó una inspección manual de todas las MRIs debido a las variaciones anatómicas de los sujetos y el lugar en donde se realizaron las lesiones. Finalmente, se realizaron un total de 14 cortes a lo largo del eje z, dividiéndolo en dos mitades. Se llevaron a cabo 9 cortes debajo de dicho punto y 5 cortes por encima.

Así mismo, fue necesario evaluar de forma manual el área de análisis de las imágenes, para esto se tomó como referencia un corte medio en el eje z de una MRI seleccionada al azar con la lesión en el NST presente tal como se muestra en el Figura 2.4. Donde, para el eje x se determinó un valor total de 120 píxeles aproximadamente, es decir partiendo desde mitad del eje se tomó 60 píxeles hacia la izquierda y 60 hacia la derecha. Y por último para el eje y se tomaron 120 píxeles partiendo del centro, esto es 25 hacia arriba y 95 hacia abajo, dando suficiente cobertura para la ubicación del NST.

Figura 2. 4 – Ejemplo de la región seleccionada [autoría propia].



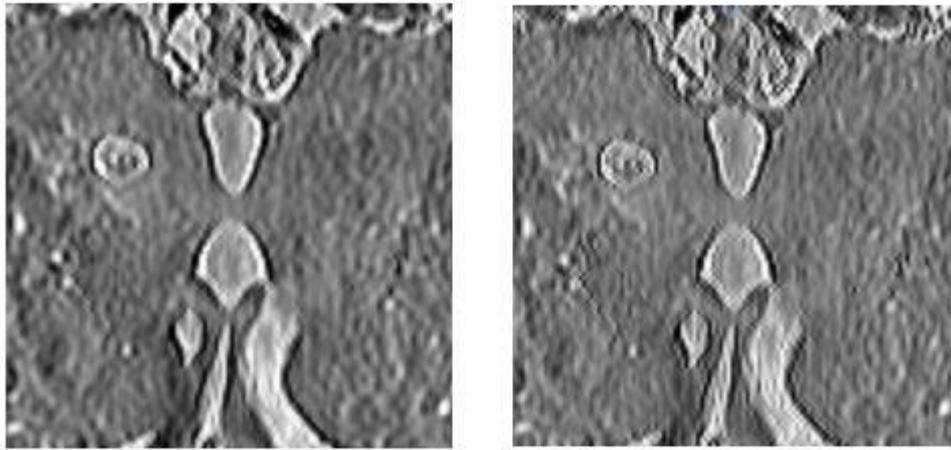
2.4 Data augmentation

Para este paso de la metodología se aplicaron técnicas de data augmentation sobre el conjunto de imágenes; que consistió en aplicar la técnica de Zero Component Analysis (ZCA) para descorrelacionar píxeles y resaltar bordes [47], y aplicar rotaciones a las imágenes cada 5 grados. Lo que finalmente permitió crear un total de 4440 imágenes, sin realizar combinaciones entre las técnicas mencionadas; partiendo del conjunto inicial de 60 MRIs. A continuación, se muestran los diferentes cambios realizados:

2.4.1 ZCA

En la Figura 2.6 se muestran dos variaciones del algoritmo ZCA, del lado izquierdo está definido un valor de ϵ de 0.1 mientras que el lado derecho se aplica un valor de 0.00001 [47].

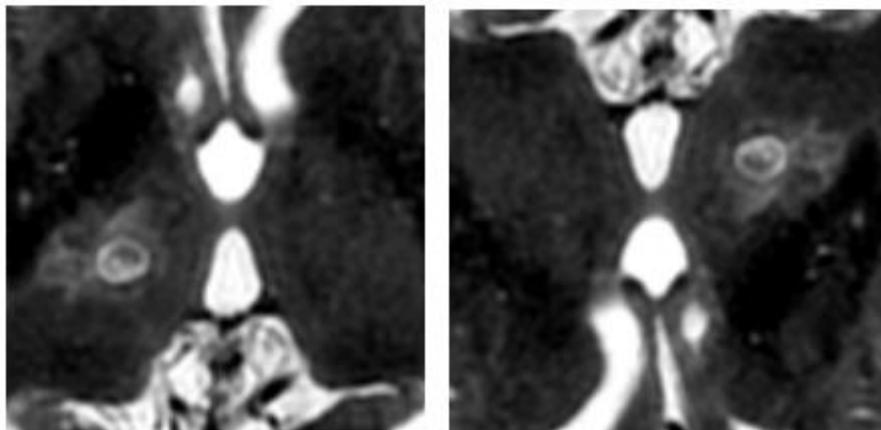
Figura 2. 5 - Variaciones del algoritmo ZCA [Autoría propia]



2.4.2 Rotaciones

En la Figura 2.7 se muestra un ejemplo de una rotación de 180 grados tomando como eje de rotación al plano XY similar al que hace un flip pero con la diferencia que la lesión no se ve reflejada como si fuera un espejo [8].

Figura 2. 6 - Comparación de Imagen original vs Rotación 180° [Autoría propia]



2.5 Entrenamiento del modelo

2.5.1 Adaptación de modelos

El análisis de modelos exitosos en el desarrollo de este tipo de segmentación fue clave para determinar la mejor arquitectura del modelo a ser utilizado en este trabajo, para ello, se analizó los resultados obtenidos de las métricas de los modelos basados en transformers, comparados con los modelos tipo convoluciones [8]. Los cambios fueron principalmente aplicados a las capas de los modelos, con la finalidad de que reciban como entrada una imagen 3D tipo MRI y devuelva como salida una matriz con el segmento correspondiente a la lesión. A continuación, se detallan las arquitecturas analizadas y los cambios propuestos para este trabajo:

2.5.1.1 MobileViT

Esta arquitectura cuenta con capas convolucionales para procesar y extraer las características iniciales de una imagen 2D con 3 canales, para luego pasar a un codificador Transformer que captura las relaciones entre los lotes 2D de imagen inicial y generar características de alto nivel y finalmente mostrar un resultado de clasificación de imágenes mediante el uso de una capa de activación que muestra las probabilidades de que una imagen pertenezca a cierta clase.

Las modificaciones se realizaron en varios bloques del modelo para permitir que acepte los valores de la matriz que se ha definido en el paso de preprocesamiento y selección de cortes con dimensiones 120x120x14. En la sección de las capas convolucionales fue necesario modificar la entrada para permitir el ingreso de una imagen 3D, el codificador también fue modificado para que sea compatible con la salida de los bloques de convolución; y, finalmente la capa de salida fue modificada para mostrar como resultado la máscara de segmentación con dimensiones similares a la entrada.

2.5.1.2 Compact Convolutional Transformer (CCT)

La arquitectura de este modelo cuenta con un tokenizer basado en convoluciones el cual se encarga de disminuir el foco de los mecanismos de atención, haciendo que se enfoquen en imágenes más pequeñas. Este componente es el encargado de dividir en tokens las imágenes que serán procesados por el resto del modelo, durante el preprocesamiento de las imágenes. El resto de la arquitectura sigue la estructura normal de los Transformers, con los mecanismos de atención captando información necesaria de los tokens generados por el tokenizer.

En este modelo se ha modificado la capa de entrada para que recibir como entrada una imagen MRI 3D junto al tokenizer, el resto del modelo fue modificado en la generación de sus capas para procesar los tokens y dé como resultado una matriz de imagen con la máscara predicha. Por último, se modificó la capa de salida para mostrar el resultado de la imagen de la máscara predicha.

2.5.1.3 Swin Transformer

Esta arquitectura cuenta con una capa de convolución inicial que se encarga de recibir imágenes 2D, después cuenta con bloques Transformer que aplican transformaciones lineales y bloques de atención basada en ventana desplazada. Y cuenta con jerarquías de bloques Transformer, donde cada una de estas jerarquías se encarga de analizar porciones más pequeñas de la imagen de entrada.

En este último modelo se ha realizado el mismo tipo de re-diseño, esto es en las capas de entrada y salida, así como para la extracción de parches y las secciones de las capas de atención y transformaciones lineales.

2.5.2 Entrenamiento

Una vez listas las MRIs procesadas y los modelos, se procedió a realizar el entrenamiento respectivo. El entrenamiento de estos modelos permitió observar los valores de las métricas que

fueron evaluadas, como el Precision, Recall, Dice, F1 score, Distancia de Hausdorff (HD) y Distancia promedio de volumen (AVD). El resumen de cada una de las métricas obtenidas se detalla en el siguiente paso de la metodología.

2.6 Métricas de evaluación

Después de la calibración de los modelos y su posterior entrenamiento, se analizaron las características de segmentación de los modelos resultantes para determinar si se debían realizar cambios en los parámetros de diseño o entrenamiento del modelo. Esto con el objetivo de mejorar el rendimiento de los modelos. Las métricas seleccionadas para la evaluación fueron las siguientes:

2.6.1 Precision

Es una métrica que indica la proporción de verdaderos positivos o píxeles correctamente clasificados del área o región de interés [48]. En la Ecuación 2.1 se puede apreciar la fórmula de esta métrica donde las siglas TP indican los verdaderos positivos y FP son los falsos positivos.

$$Recall = \frac{TP}{TP + FN} \quad (Ec. 2.1)$$

2.6.2 Recall

Esta métrica indica la proporción de verdaderos positivos sobre el total de píxeles que fueron clasificados y encontrados por el modelo [48]. La Ecuación 2.2 puede ser vista a continuación, donde las siglas TP indican los verdaderos positivos y FN son los falsos negativos

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

2.6.3 Dice

Es una métrica que combina las dos métricas anteriores Precision y recall, y mide la intersección entre la máscara de segmentación predicha y la máscara de segmentación de referencia [49]. En la Ecuación 2.3 se puede observar cómo está definida.

$$DICE = \frac{2 \times TP}{(TP + FP) + (TP + FN)} \quad (Ec. 2.3)$$

2.6.4 F1-score

Esta métrica se define como otra combinación de las métricas Precision y Recall en una sola medida. Y refleja la proporción de instancias clasificadas correctamente como positivas sobre el total de instancias clasificadas como positivas [49]. Está definida por la Ecuación 2.4.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (Ec. 2.4)$$

2.6.5 Distancia de Hausdorff (HD)

Es una métrica que permite realizar la evaluación la similitud o disimilitud entre dos conjuntos de puntos o segmentaciones [49]. En la Ecuación 2.5 se puede ver la definición de esta métrica:

$$HD(A, B) = \max(h(A, B), h(B, A)) \quad (Ec. 2.5)$$

2.6.6 Distancia promedio de volumen (AVD)

Es una métrica que permite realizar la evaluación de la precisión de las segmentaciones tridimensionales de forma que se mide la diferencia que existe entre la segmentación que se envía al modelo como entrada y la segmentación predicha [49]. En la Ecuación 2.6 se puede ver cómo está definida esta métrica:

$$AVD = \frac{|VG - VP|}{VG} \quad (Ec. 2.6)$$

2.7 Implementación de prototipo web

2.7.1 Requerimientos funcionales

Los requerimientos del prototipo son:

- Identificación las características de interés en las MRI.
- Facilitar la carga de imágenes de resonancia magnética por parte del usuario.
- Presentar de forma visual los resultados de las segmentaciones obtenidas.

2.7.2 Requerimientos no funcionales

Los requerimientos no funcionales del prototipo son:

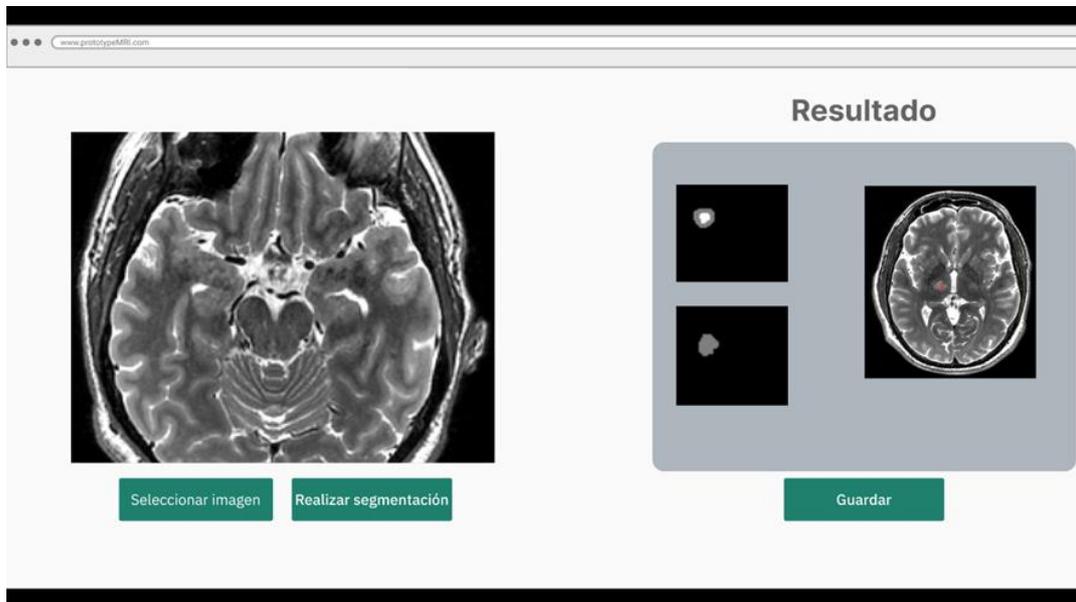
- Usabilidad – Interfaz sencilla, intuitiva, amigable y con pocos botones
- Rendimiento - Tiempo de respuesta para la segmentación razonable.
- Precisión - Procesar y segmentar MRIs de forma correcta y en un tiempo razonable.

2.7.3 Diseño del prototipo

El prototipo cuenta con dos roles principales, el del usuario estándar que empleará el prototipo y el de administrador. El primer rol solo puede acceder a la característica principal del prototipo que consiste en subir una RMI y obtener la segmentación. Para el rol de administrador se propuso que se pudiera agregar nuevas RMI's y realizar entrenamientos del modelo de segmentación en base a los ejemplos ingresados.

Para el diseño inicial del prototipo se propuso implementar dos ventanas de visualización, una para la carga de imagen y otra para mostrar el resultado de la segmentación. Tres botones los cuales servirían para la carga de imágenes de MR, realización de segmentación y descarga del resultado de la segmentación realizada. El diseño inicial del prototipo se muestra a continuación en la Figura 2.8:

Figura 2. 5 – Prototipo inicial



El diseño inicial del prototipo cuenta con dos secciones principales, la primera es la de carga de imágenes y la segunda la que muestra los resultados. Para la primera sección se cuenta con una zona de pre-visualización y dos botones “Seleccionar imagen” y “Realizar segmentación”, el primero permite que el usuario especifique la imagen de MRI adicionalmente para esto se mostrará en la zona de previsualización de imagen la cargada; y el segundo permite enviar una petición para que se proceda a realizar la segmentación de la imagen seleccionada. La segunda sección cuenta con una zona de visualización de las segmentaciones y un botón que permite guardar estos resultados.

2.7.4 Elección de tecnología

El lenguaje de programación empleado para este proyecto es Python debido a la integración sencilla de sus librerías y amplia documentación en la web para cada una de ellas.

Para la parte del pre-procesamiento de las RMI's se empleó librerías como Antspy (versión 0.3.8) [44] y Numpy (versión 1.21.0) [50], las cuales permiten visualizar, analizar y modificar de forma sencilla imágenes 3D. Para la construcción y entrenamiento de los modelos

de segmentación se tomó de base los modelos que se encuentran en los repositorios de GitHub de Keras, construidos con mediante la librería de Tensorflow (versión 2.11.0) [51].

Por último, para el desarrollo del prototipo se decidió emplear Django (versión 4.2.3) [52], una librería de Python para el desarrollo web con la finalidad tener una integración más sencilla al crear un servidor con el modelo de segmentación final, y React (versión 17.0.2) [53] para la creación de la interfaz del prototipo, en el cual se empleó Axios (versión 1.4.0) [54] para el control de peticiones, Material-UI (versión 5.3.1) [55] para control de componentes web y estilos y Niivue (versión 0.36.0) [56] una librería web que permite el renderizado de MRI's en el navegador web.

Se empleó 2 tipos de recursos computacionales: una computadora de Espol con acceso a una gráfica, así como recursos de Google Colab Pro; con estos recursos se realizó operaciones del pre-procesamiento, data augmentation y entrenamiento de los modelos de segmentación.

Capítulo 3

3.1 Resultados y análisis

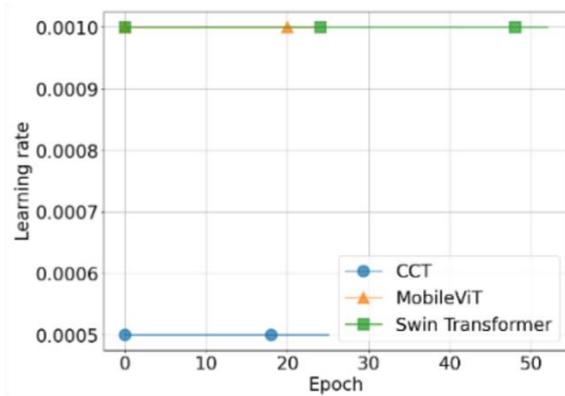
3.1.1 Modelos base

Durante el proceso de entrenamiento y validación Figura 3.1 y Figura 3.2, se puede observar que el mejor modelo capaz de aprender las características más relevantes al problema de segmentación fue Swin Transformer, donde se evidencia mejores métricas en los valores de DICE en la Figura 3.1b, Precision en la Figura 3.1e, F1-score en la Figura 3.1d y Recall en la Figura 3.1f. Con el modelo MobileViT se obtuvo resultados similares pero inferiores a Swin Transformer en la mayoría de las métricas, por otro lado con CCT se observa que no existe rápida convergencia a diferencia de los otros dos modelos.

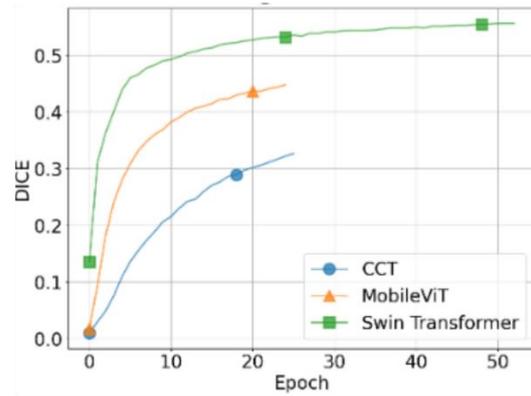
En la Figura 3.1c se puede observar la función de pérdida, la cual en los modelos va disminuyendo con el paso de las épocas; sin embargo, se observa una disminución notable para Swin Transformer luego de las épocas iniciales. De igual forma en la Figura 3.1d para la métrica de F1-score, Figura 3.1e para la métrica de Precision y Figura 3.1f para la métrica de Recall se observa que para el modelo de Swin Transformer existe una mejora temprana y que escala de forma rápida a diferencia de CCT y MobileViT, los cuales tienen valores iniciales bajos y mejoran muy poco con el paso de las épocas.

Durante la etapa de entrenamiento se emplearon 3500 MRI's obtenidas de las operaciones realizadas por el data augmentation, las cuales fueron descritas en el capítulo anterior.

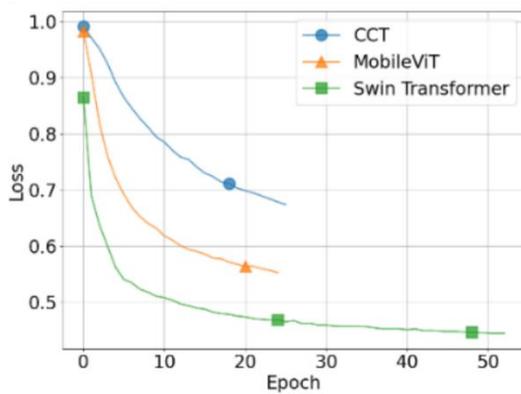
Figura 3. 1 - Resultado de métricas del dataset de entrenamiento en los modelos base [autoría propia]



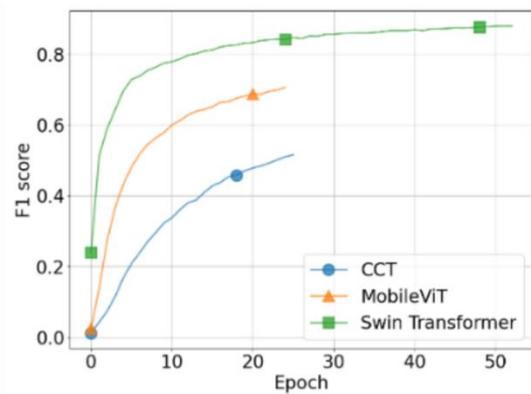
a) Aprendizaje vs. épocas



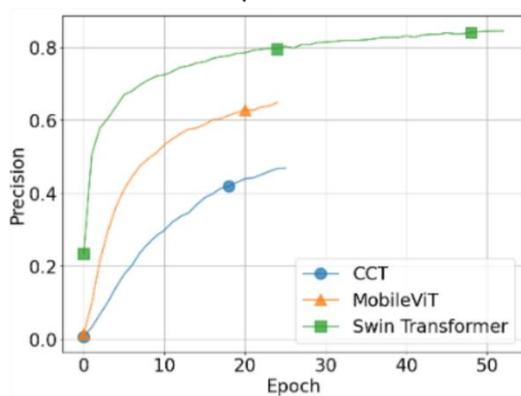
b) DICE vs. épocas



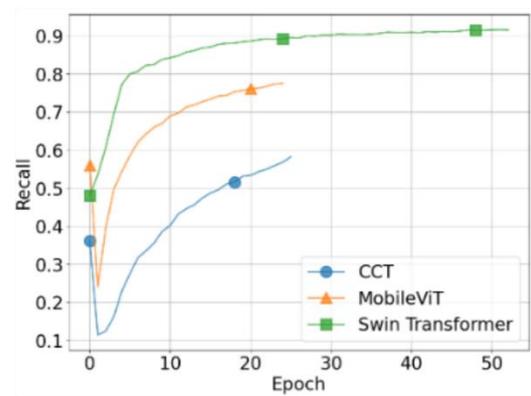
c) Función de pérdida vs. épocas



d) F1 Score vs. épocas



e) Precision vs. épocas

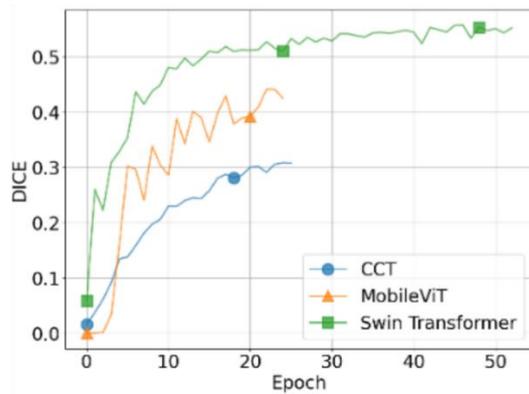


f) Recall vs. épocas

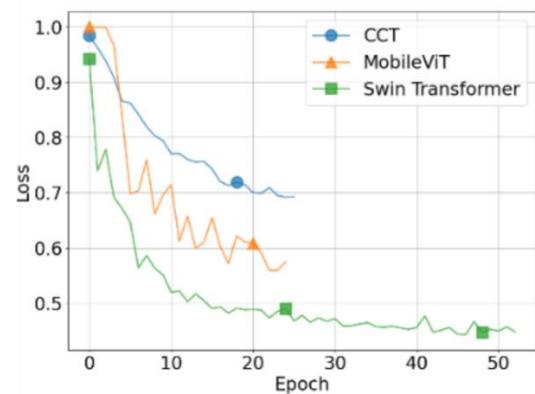
Luego se realizaron pruebas sobre el conjunto de validación, para verificar que los modelos no estuviesen memorizando las características de las RMI's (overfitting), lo que se

puede evidenciar al realizar la comparación entre los valores obtenidos durante en el entrenamiento Figura 3.1 y los valores obtenidos durante la validación en la Figura 3.2. Durante esta etapa de validación se usaron 400 RMI's.

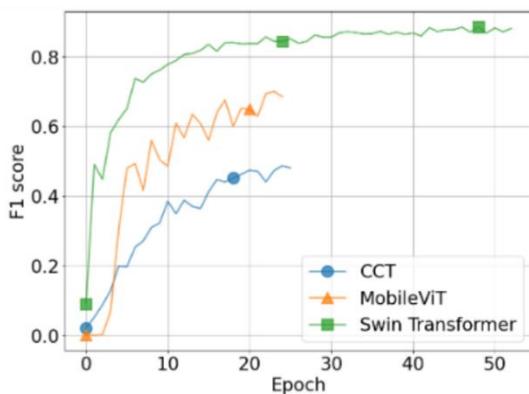
Figura 3. 2 - Resultado de métricas del dataset de validación en los modelos base [autoría propia].



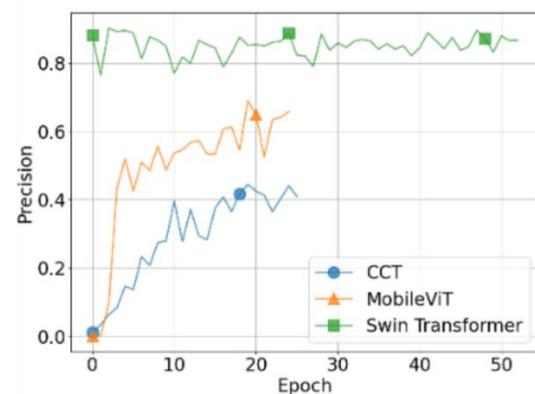
a) DICE vs. épocas



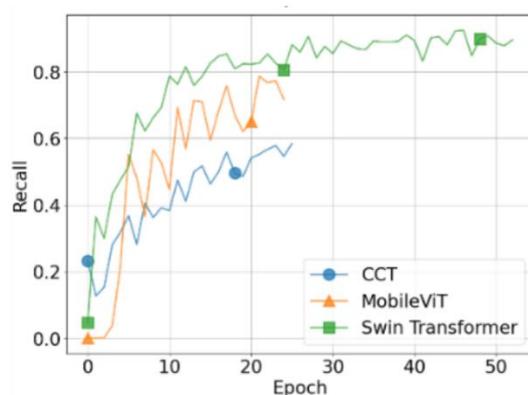
b) Función de pérdida vs. épocas



c) F1 Score vs. épocas



d) Precision vs. épocas



e) Recall vs. épocas

Es importante resaltar que estos modelos iniciales contenían la lesión mezclada con el edema; es decir, una máscara binaria. Por lo que se realizó dos modificaciones: la primera fue aplicada al dataset y la segunda al mejor modelo de segmentación de los 3 modelos base. La primera modificación consistió en realizar una segunda evaluación del dataset para verificar que se estuvieran cubriendo casos clínicos reales. La segunda modificación consistió en emplear al mejor modelo de segmentación Swin Transformer, al cual se le realizaron modificaciones para evitar interpolaciones de las clases presentes en las máscaras de segmentación.

3.1.2 Dataset mejorado

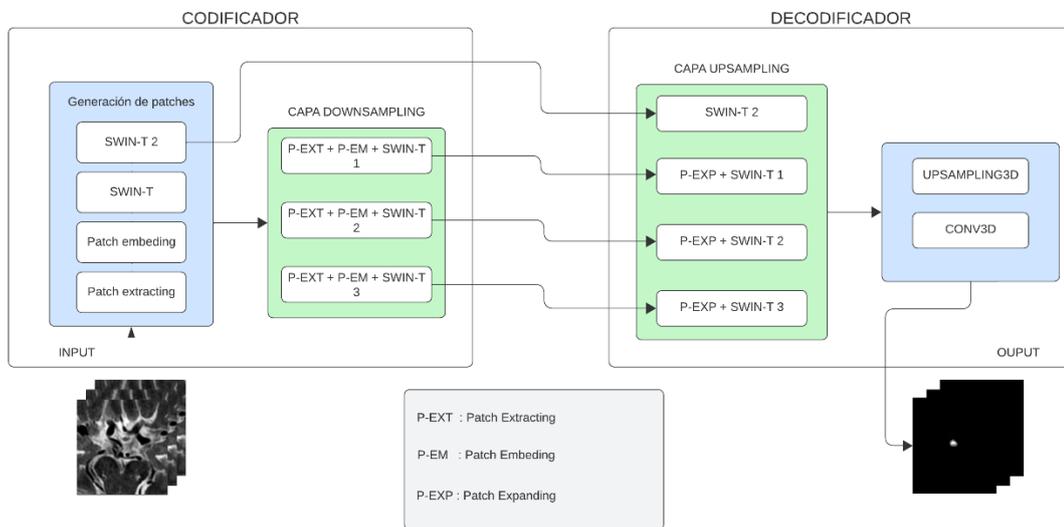
Los cambios se realizaron en el conjunto de validación y en el conjunto de test excluyendo aquellas imágenes que posean alguna modificación realizada durante el data augmentation, esto fue realizado con la finalidad de obtener resultados más cercanos a casos clínicos reales y evitar confusiones durante el proceso de validación y prueba del modelo.

3.1.3 Modelo mejorado

Los cambios realizados al modelo Swin Transformer consistieron en una adaptación del modelo inicial, tal como se muestra en la Figura 3.3, donde se tomó como backbone (red principal) al modelo inicial y se adaptó a una arquitectura UNet. Para esto se consideró lo siguiente:

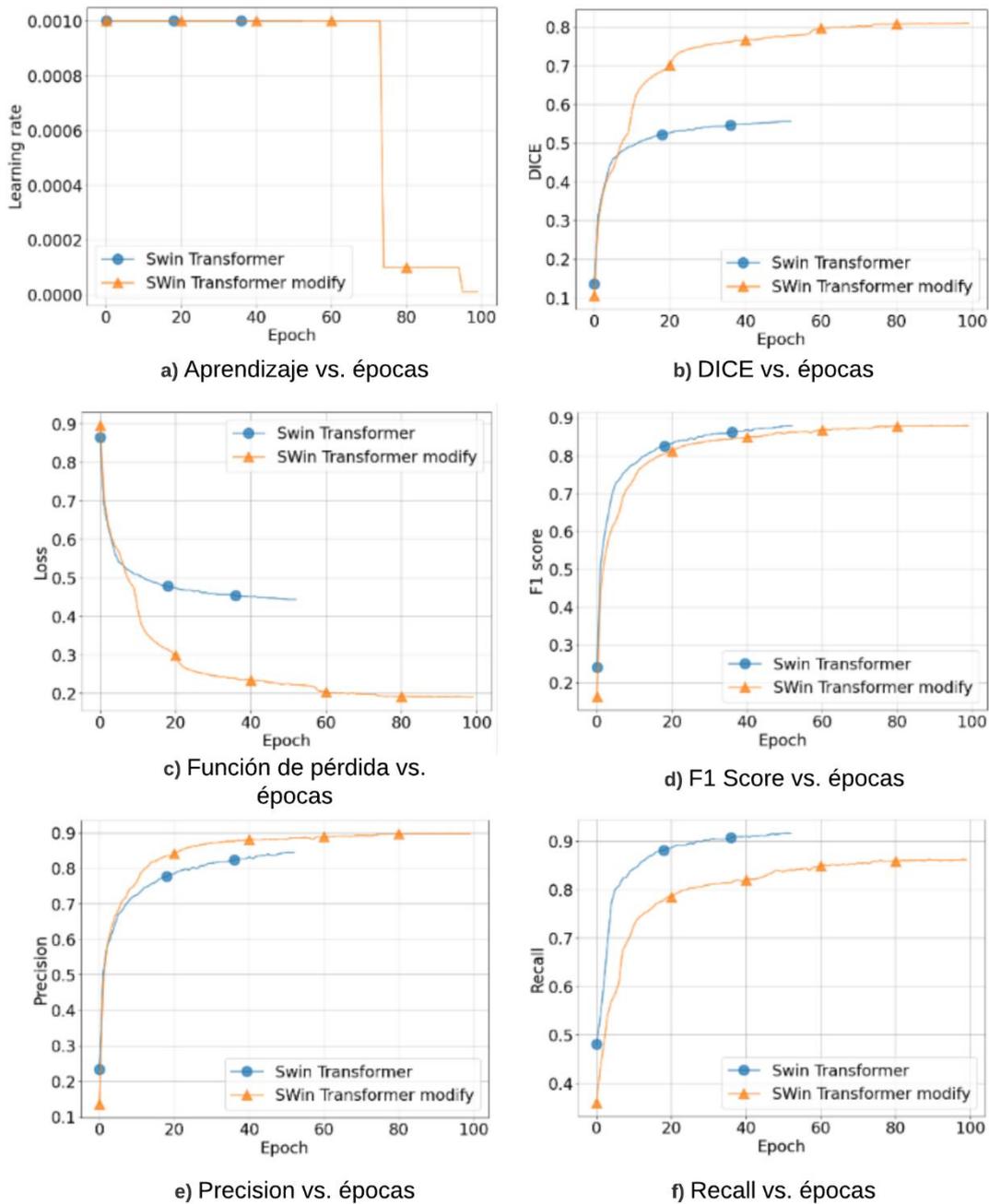
- Se estableció una sección encoder y una sección decoder en la cual sus componentes principales emplearon la arquitectura Swin Transformer.
- Adaptación de las modificaciones realizadas anteriormente para las capas de entrada y salida.
- Adaptación para las capas ocultas del modelo.
- El resto de hiperparámetros fueron conservados.

Figura 3. 3 - Arquitectura mejorada para el modelo base de Swin Transformer [autoría propia]



Después de aplicar estas modificaciones al modelo original Swin Transformer, se realizó nuevamente el entrenamiento. En la Figura 3.4 se puede ver como en las métricas DICE, Precisión, F1-score, Recall se obtuvo mejores resultados con el modelo Swin Transformer modificado en comparación a los resultados obtenidos con el modelo Swin Transformer base durante la fase de entrenamiento. Las mejoras principalmente están en las métricas DICE, que pasó de 0.55 (base) a 0.80 (modificado) aproximadamente, tal como se muestra en la Figura 3.4c; y la métrica de Precisión, que pasó desde 0.85 (base) a 0.90 (modificado), tal como se muestra en la Figura 3.4e. El F1-score en la Figura 3.4d no posee diferencias notorias tan grandes, sin embargo se observa un descenso en la métrica Recall en la Figura 3.4f lo cual se debe principalmente a la relación que guarda con la métrica de Precisión y F1-score, donde la primera ha aumentado ligeramente y la segunda ha disminuido.

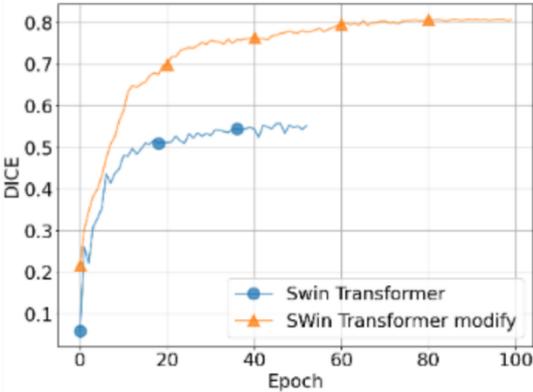
Figura 3. 4 - Resultado de métricas del dataset de entrenamiento del modelo Swin Transformer base vs Swin Transformer modificado [autoría propia].



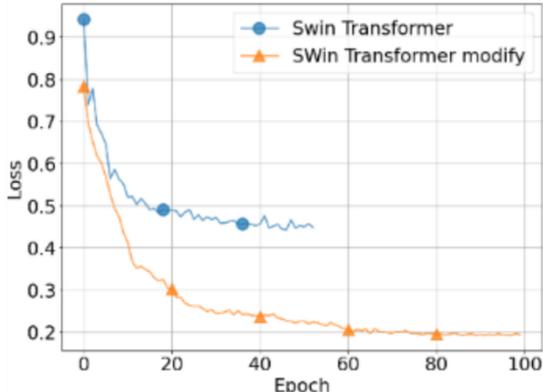
De igual forma se realizó la respectiva validación, donde se observa que al igual que el entrenamiento las métricas DICE, Precisión, F1-score y Recall se obtuvieron mejores resultados con el modelo Swin Transformer modificado en comparación al modelo base. Donde en las

métricas DICE en la Figura 3.5a y Precision en la Figura 3.5d se observa una mejora significativa. Y las métricas F1-score en la Figura 3.5c y Recall en la Figura 3.5e se mantienen con valores similares.

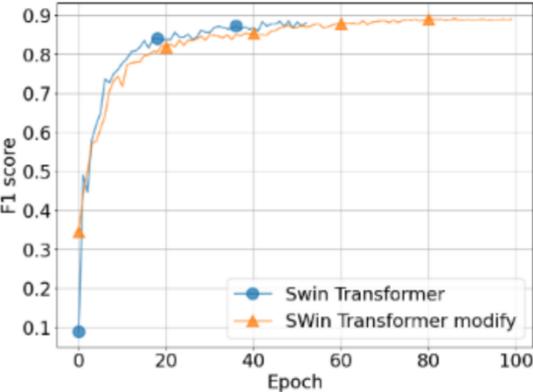
Figura 3. 5 - Resultado de métricas del dataset de validación del modelo Swin Transformer base vs Swin Transformer modificado [autoría propia].



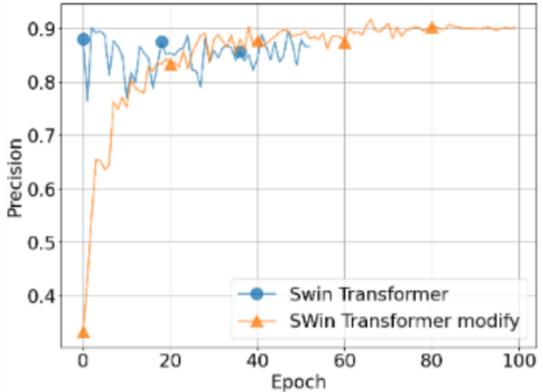
a) DICE vs. épocas



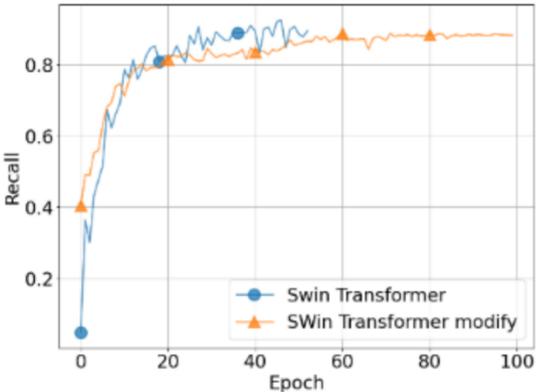
b) Función de pérdida vs. épocas



c) F1 Score vs. épocas



d) Precision vs. épocas



e) Recall vs. épocas

Finalmente, después de obtener los resultados del modelo mejorado, se procedió a realizar el cálculo para las métricas complementarias: Hausdorff Distance, Average Volumen Distance y Cross-entropy. Esto se puede observar en la Tabla 3.1, la cual muestra los resultados obtenidos por cada uno de los modelos que se emplearon durante este trabajo.

Tabla 3. 1 - Resultados de las métricas para el dataset de pruebas [autoría propia].

Modelo	DICE	Recall	Precision	F1_Score	Loss	AVD	Hausdorff distance	Cross entropy
MOBILEVIT	0.4767	0.9014	0.5439	0.6784	0.5233	0.5619	3.0000	0.0057
CCT	0.2810	0.5878	0.3668	0.4517	0.7190	0.4314	84.2971	0.0434
SWIN	0.5542	0.9254	0.8251	0.8715	0.4458	0.1017	3.1633	0.0090
SWIN MODIFICADO	0.7590	0.9037	0.9037	0.9037	0.2410	0.0019	2.0000	0.0056

En la Tabla 3.1 se puede observar que los modelos basados en Swin Transformer tienen mejores resultados para las métricas: Hausdorff Distance, que indica la proporción de volumen predicho que se debe expandir o contraer para poder encapsular al volumen objetivo; Average Volumen Distance, que indica la similitud entre el volumen predicho por el modelo y el volumen objetivo empleando porcentajes; y por último, Cross-entropy que indica la capacidad del modelo para etiquetar correctamente la clase de cada vóxel, donde valores cercanos a cero indican que se está etiquetando con mayor precisión.

En la Tabla 3.2 se tiene una matriz de confusión, calculada a partir de las MRI's del conjunto de prueba, con porcentajes de presencia de vóxeles clasificados como una clase o fondo. Se muestra una escala de colores azulados, donde las celdas más claras indican mayor presencia de valores y las celdas más oscuras indican mayor presencia de valores. La Tabla 3.2 muestra que la mayor parte de valores se están clasificando de forma correcta y una cantidad aceptable de vóxeles que están siendo considerados como otras clases o parte del fondo, lo que es evidenciado al observar la diagonal donde se aprecian valores mayores a 75, lo que complementa a lo obtenido en la Figura 3.4, Figura 3.5 y la Tabla 3.1.

Tabla 3. 2 - Matriz de confusión del modelo Swin Transformer [autoría propia].

Matriz de confusión	Fondo	Clase 1 (Edema)	Clase 2 (Lesión)
Fondo	99.98	0.02	0
Clase 1 (Edema)	12.4	84.36	3.24
Clase 2 (Lesión)	4.97	18	77.02

A continuación se muestra los resultados de las segmentaciones realizadas por los dos mejores modelos de base; en la Figura 3.6 se muestra el resultado de MobileViT, y en la Figura 3.7 el resultado de Swin Transformer, donde se aprecia que ambos modelos identifican de forma correcta el área circundante a la lesión, sin embargo también se puede apreciar que se coloca un mismo valor tanto para la lesión como el edema, representándolos con un mismo valor de intensidad.

Figura 3. 6 - Resultado de la segmentación con el modelo MobileViT [autoría propia].



Figura 3. 7 - Resultado de la segmentación con el modelo Swin Transformer [autoría propia].



En la Figura 3.8 se puede ver un ejemplo de las segmentaciones realizadas por el modelo Swin Transformer modificado, donde se empleó una imagen del dataset corregido. La Figura 3.8 muestra ambas clases representadas por dos colores, el color gris señala la región del edema mientras que el color azul representa la región de la lesión.

Figura 3. 8 Resultado de la segmentación con el modelo de Swin Transformer modificado [autoría propia].



3.1.4 Prototipo

El prototipo desarrollado contiene el mejor modelo de segmentación. En la pantalla inicial se permite al usuario seleccionar dos opciones: “Ingresar como médico” o “Ingresar como administrador”. La primera opción lo lleva a la sección que se aprecia en la Figura 3.8, la cual le permite cargar y visualizar una imagen MRI, y luego realizar el proceso de segmentación, tal como se observa en la Figura 3.9, incluso colocar observaciones respecto al resultado de las segmentaciones como se aprecia en la Figura 3.10.

Figura 3. 8 – Pantalla inicial del prototipo [autoría propia].



Figura 3. 9 – Visualización y segmentación de MRI [autoría propia].

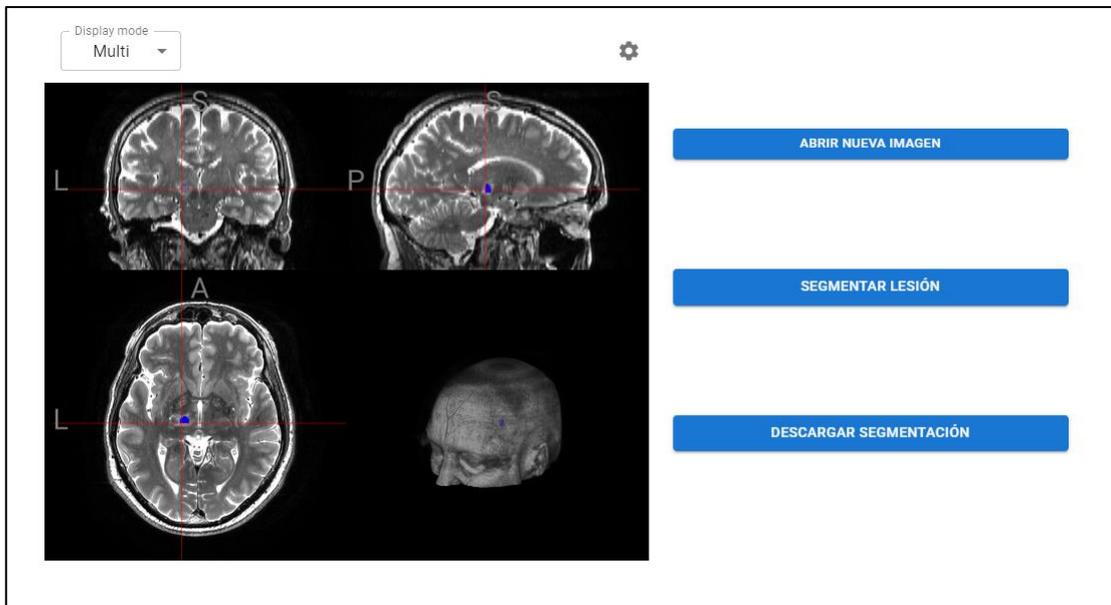


Figura 3. 10 – Sección de observaciones [autoría propia].



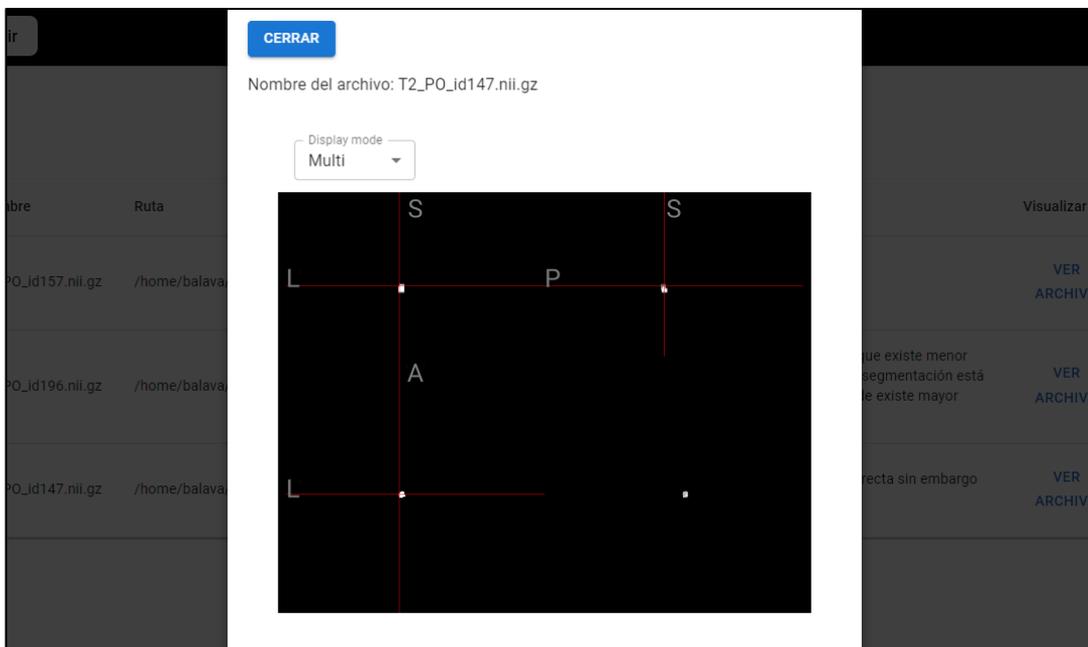
En la Figura 3.11 se aprecia el inicio de la sesión luego de seleccionar “Ingresar como administrador”, posterior en la Figura 3.12 se observa el rol administrador, que le permite visualizar una tabla con los datos de las máscaras de segmentación predichas por el modelo, así como el resultado de las MRI’s cargadas por personal médico, y que han recibido observaciones;

además tal como se muestra en la Figura 3.12, se puede visualizar la imagen al hacer clic en el botón de “Ver archivo”.

Figura 3. 11 – Sección de administrador [autoría propia].

Nombre	Ruta	Observacion	Visualizar
T2_PO_id157.nii.gz	/home/balava/Escritorio/SERVER/Imagenes/prepro/T2_PO_id157.nii.gz	Revisión completa	VER ARCHIVO
T2_PO_id196.nii.gz	/home/balava/Escritorio/SERVER/Imagenes/prepro/T2_PO_id196.nii.gz	En las capas iniciales se evidencia que existe menor presencia de lesión, sin embargo la segmentación está realizada de forma correcta en donde existe mayor presencia de lesión	VER ARCHIVO
T2_PO_id147.nii.gz	/home/balava/Escritorio/SERVER/Imagenes/masks/T2_PO_id147.nii.gz	Parece estar realizada de forma correcta sin embargo debe evaluarse.	VER ARCHIVO

Figura 3. 12 – Visualización de MRI's con observaciones [autoría propia].



Capítulo 4

4.1 Conclusiones y recomendaciones

4.1.1 Conclusiones

Entre los modelos de base el que se ajustó mejor a la tarea de segmentación fue el Swin Transformer en comparación con los modelos CCT y MobileViT, los que durante el entrenamiento, validación y prueba, obtuvieron la métrica de F1-score: 0.87, 0.67 y 0.45 respectivamente.

La modificación del modelo Swin Transformer permitió obtener un mejor desempeño en general del conjunto de métricas. El cambio esencial que permitió obtener un mejor desempeño consistió en adaptar una arquitectura Unet con una sección encoder-decoder, empleando como backbone el modelo Swin Transformer base y adaptando las capas de entrada y salida, así como capas ocultas.

La evaluación y corrección de los datasets permitió obtener métricas más estables durante la validación y prueba, ya que se emplearon casos clínicos reales agregando más confiabilidad a los resultados del modelo.

Finalmente, el prototipo implementado consigue acoplar el mejor modelo de segmentación con una interfaz web amigable y de fácil uso para personal médico, haciendo transparente la configuración e implementación del servidor que maneja tanto el pre-procesamiento de las MRI's así como la segmentación de los volúmenes.

4.1.2 Recomendaciones

En este trabajo se usaron imágenes de resonancia magnética de secuencia T2 con cortes de 1mm de separación, que al igual a lo realizado en trabajos pasados esta característica de separación permite obtener información relevante alrededor del NST y en este caso permite realizar el objetivo de segmentar la lesión y edema.

Se recomienda la adquisición de MRI's con características diferentes y que aborden varios casos clínicos, ya que esto beneficiaría a que durante proceso de data augmentation se generen una cantidad diversa de imágenes que ayuden enriquecer el entrenamiento del modelo y posterior la generalización de casos clínicos al emplear el modelo en un entorno real de producción.

En data augmentation los algoritmos propuestos pueden ser modificados y empleados dependiendo de las necesidades, para el algoritmo de resalte de bordes se recomienda emplear no más de 2 factores diferentes ya que se tiende a modificar demasiado la imagen original, en las rotaciones se puede emplear ángulos más pequeños simulando casos clínicos reales. También, existe la posibilidad de agregar otro tipo de operaciones de data augmentation que modifiquen niveles de brillo y contraste de las MRI's, para aumentar el tamaño del dataset en el conjunto de validación.

Los modelos iniciales no son suficientes para determinar de forma clara tanto la lesión como el edema, por lo que aumentando la complejidad de los modelos base no seleccionados podría ayudar a encontrar de forma correcta la lesión y el edema, lo que podría conllevar a obtener resultados similares o superiores al presentado por el modelo Swin Transformer modificado, y de esta forma evitar que se realicen interpolaciones del edema en la lesión y se obtengan máscaras "incompletas" o "binarias".

El modelo modificado Swin Transformer demuestra una mejoría significativa en comparación al modelo base, bajo esta premisa explorar otro tipo de modificaciones al modelo podría asegurar la obtención de mejores resultados durante las fases de entrenamiento, validación y prueba.

Por último, dentro del contexto de la medicina es posible explorar nuevas tareas de identificación de lesiones en otras partes del cuerpo empleando los modelos propuestos durante este trabajo, siendo herramientas que pueden aportar al área de neuroimagen o similares, siempre y cuando exista un proceso regulatorio y se logre estándares necesarios para su aplicación.

Bibliografía

- [1] Jankovic, J., & Lang, A. E. (2021). Diagnosis and assessment of Parkinson disease and other movement disorders. *Bradley's Neurology in Clinical Practice E-Book*, 310(1).
- [2] Obeso, J. A., Jahanshahi, M., Alvarez, L., Macias, R., Pedroso, I., Wilkinson, L., Pavon, N., Day, B., Pinto, S., Rodríguez-Oroz, M. C., Tejero, J., Artieda, J., Talelli, P., Swayne, O., Rodríguez, R., Bhatia, K., Rodríguez-Díaz, M., Lopez, G., Guridi, J., & Rothwell, J. C. (2009). What can man do without basal ganglia motor output? The effect of combined unilateral subthalamotomy and pallidotomy in a patient with Parkinson's disease. *Experimental Neurology*, 220(2), 283–292. <https://doi.org/10.1016/j.expneurol.2009.08.030>
- [3] Aristieta, A., Azkona, G., Sagarduy, A., Miguelez, C., Ruiz-Ortega, J. Á., Sanchez-Pernaute, R., & Ugedo, L. (2012). The role of the subthalamic nucleus in L-DOPA induced dyskinesia in 6-hydroxydopamine lesioned rats. *PloS one*, 7(8), e42652. <https://doi.org/10.1371/journal.pone.0042652>
- [4] McGowan, J. C. (2008). Basic principles of magnetic resonance imaging. *Neuroimaging Clinics of North America*, 18(4), 623-636.
- [5] Bakshi, R., Ariyaratana, S., Benedict, R. H., & Jacobs, L. (2001). Fluid-attenuated inversion recovery magnetic resonance imaging detects cortical and juxtacortical multiple sclerosis lesions. *Archives of neurology*, 58(5), 742-748.
- [6] Kawahara, D., & Nagata, Y. (2021). T1-weighted and T2-weighted MRI image synthesis with convolutional generative adversarial networks. *reports of practical Oncology and radiotherapy*, 26(1), 35-42.
- [7] Kuperman V. 2000. *Magnetic Resonance Imaging: Physical Principles and Applications*. San Diego, CA: Academic.
- [8] Magallanes, F.; Nebel, J. (2022). Segmentación automática de la Sustancia Negra cerebral y Núcleo Subtalámico a partir de imágenes 3D de resonancia magnética.
- [9] Blandini, F., Nappi, G., Tassorelli, C., & Martignoni, E. (2000). Functional changes of the basal ganglia circuitry in Parkinson's disease. *Progress in Neurobiology*, 62(1), 63–88. doi:10.1016/s0301-0082(99)00067-2
- [10] Permezel, F. (2021). Brain MRI-guided focused ultrasound conceptualised as a tool for brain network intervention. *Journal of Clinical Neuroscience*, 90, 370–379. <https://doi.org/https://doi.org/10.1016/j.jocn.2021.05.062>
- [11] Martínez-Fernández, R., Rodríguez-Rojas, R., del Álamo, M., Hernández-Fernández, F., Pineda-Pardo, J. A., Dileone, M., Alonso-Frech, F., Foffani, G., Obeso, I., Gasca-Salas, C., de Luis-Pastor, E., Vela, L., & Obeso, J. A. (2018). Focused ultrasound subthalamotomy in patients with asymmetric Parkinson's disease: a pilot study. *The Lancet Neurology*, 17(1), 54–63. [https://doi.org/https://doi.org/10.1016/S1474-4422\(17\)30403-9](https://doi.org/https://doi.org/10.1016/S1474-4422(17)30403-9)

- [12] Rohani, M., & Fasano, A. (2017). Focused Ultrasound for Essential Tremor: Review of the Evidence and Discussion of Current Hurdles. *Tremor and other hyperkinetic movements* (New York, N.Y.), 7, 462. <https://doi.org/10.7916/D8Z89JN1>
- [13] Shen, W., Zhou, M., Yang, F., Yang, C., & Tian, J. (2015). Multi-scale Convolutional Neural Networks for Lung Nodule Classification. *Information Processing in Medical Imaging*, 588–599. doi:10.1007/978-3-319-19992-4_46
- [14] Park, C. W., Seo, S. W., Kang, N., Ko, B., Choi, B. W., Park, C. M., Chang, D. K., Kim, H., Kim, H., Lee, H., Jang, J., Ye, J. C., Jeon, J. H., Seo, J. B., Kim, K. J., Jung, K. H., Kim, N., Paek, S., Shin, S. Y., Yoo, S., ... Yoon, H. J. (2020). Artificial Intelligence in Health Care: Current Applications and Issues. *Journal of Korean medical science*, 35(42), e379. <https://doi.org/10.3346/jkms.2020.35.e379>
- [15] Karimaghloo, Z., Rivaz, H., Arnold, D. L., Collins, D. L., & Arbel, T. (2015). Temporal Hierarchical Adaptive Texture CRF for Automatic Detection of Gadolinium-Enhancing Multiple Sclerosis Lesions in Brain MRI. *IEEE Transactions on Medical Imaging*, 34(6), 1227–1241. <https://doi.org/10.1109/TMI.2014.2382561>
- [16] Galimzianova, A., Pernuš, F., Likar, B., & Špiclin, Ž. (2016). Stratified mixture modeling for segmentation of white-matter lesions in brain MR images. *NeuroImage*, 124, 1031–1043. <https://doi.org/https://doi.org/10.1016/j.neuroimage.2015.09.047>
- [17] Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., & Erickson, B. J. (2017). Deep learning for brain MRI segmentation: state of the art and future directions. *Journal of digital imaging*, 30, 449-459.
- [18] Lee, A., & Gilbert, R. M. (2016). Epidemiology of Parkinson Disease. *Neurologic Clinics*, 34(4), 955–965. doi:10.1016/j.ncl.2016.06.012
- [19] Morris, M. E. (2000). Movement disorders in people with Parkinson disease: a model for physical therapy. *Physical therapy*, 80(6), 578-597.
- [20] Gupta, H. v, Pahwa, R., Dowell, P., Khosla, S., & Lyons, K. E. (2021). Exploring essential tremor: Results from a large online survey. *Clinical Parkinsonism & Related Disorders*, 5, 100101. <https://doi.org/https://doi.org/10.1016/j.prdoa.2021.100101>
- [21] Jenner, P. (2008). Preventing and controlling dyskinesia in Parkinson's disease—a view of current knowledge and future opportunities. *Movement Disorders: Official Journal of the Movement Disorder Society*, 23(S3), S585-S598.
- [22] Solmi, M., Pigato, G., Kane, J. M., & Correll, C. U. (2018). Clinical risk factors for the development of tardive dyskinesia. *Journal of the Neurological Sciences*, 389, 21-27.
- [23] Emmi, A., Antonini, A., Macchi, V., Porzionato, A., & De Caro, R. (2020). Anatomy and connectivity of the subthalamic nucleus in humans and non-human primates. *Frontiers in neuroanatomy*, 14, 13.
- [24] Dovrou, A., Nikiforaki, K., Zaridis, D., Manikis, G. C., Mylona, E., Tachos, N., ... & Marias, K. (2023). A segmentation-based method improving the performance of N4 bias field correction on T2weighted MR imaging data of the prostate. *Magnetic Resonance Imaging*, 101, 1-12.

- [25] Méndez-Herrera, C. R. (2011). El núcleo subtalámico en la fisiopatología de la enfermedad de Parkinson y su rol como diana quirúrgica. *Revista chilena de neuro-psiquiatría*, 49(1), 62-68.
- [26] Stover, N. P., Okun, M. S., Evatt, M. L., Raju, D. V., Bakay, R. A., & Vitek, J. L. (2005). Stimulation of the subthalamic nucleus in a patient with Parkinson disease and essential tremor. *Archives of neurology*, 62(1), 141-143.
- [27] Bellon, E. M., Haacke, E. M., Coleman, P. E., Sacco, D. C., Steiger, D. A., & Gangarosa, R. E. (1986). MR artifacts: a review. *American Journal of Roentgenology*, 147(6), 1271-1281.
- [28] Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging*, 29(6), 1310-1320..
- [29] Nyúl, L. G., & Udupa, J. K. (1999). On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(6), 1072-1081.
- [30] Shah, M., Xiao, Y., Subbanna, N., Francis, S., Arnold, D. L., Collins, D. L., & Arbel, T. (2011). Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. *Medical image analysis*, 15(2), 267-282.
- [31] Wyawahare, M. V., Patil, P. M., & Abhyankar, H. K. (2009). Image registration techniques: an overview. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2(3), 11-28.
- [32] Hill, D. L., Batchelor, P. G., Holden, M., & Hawkes, D. J. (2001). Medical image registration. *Physics in medicine & biology*, 46(3), R1.
- [33] Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., & Karnath, H. O. (2012). Age-specific CT and MRI templates for spatial normalization. *Neuroimage*, 61(4), 957-965.
- [34] Hu, W., Cai, B., Zhang, A., Calhoun, V. D., & Wang, Y. P. (2019). Deep collaborative learning with application to the study of multimodal brain development. *IEEE Transactions on Biomedical Engineering*, 66(12), 3346-3359.
- [35] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Hounsfield, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- [36] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).
- [37] Kawahara, D., & Nagata, Y. (2021). T1-weighted and T2-weighted MRI image synthesis with convolutional generative adversarial networks. *Reports of practical oncology and radiotherapy : journal of Great Poland Cancer Center in Poznan and Polish Society of Radiation Oncology*, 26(1), 35–42. <https://doi.org/10.5603/RPOR.a2021.0005>

- [38] Han, X. (2017). MR-based synthetic CT generation using a deep convolutional neural network method. *Medical physics*, 44(4), 1408-1419.
- [39] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (pp. 234-241). Springer International Publishing.
- [40] Russell, S., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. Pearson.
- [41] Mehta, S., & Rastegari, M. (2021). Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178*.
- [42] Hassani, A., Walton, S., Shah, N., Abuduweili, A., Li, J., & Shi, H. (2021). Escaping the big data paradigm with compact transformers. *arXiv preprint arXiv:2104.05704*.
- [43] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10012-10022).
- [44] Avants, B. B., Tustison, N., & Song, G. (2009). Advanced normalization tools (ANTS). *Insight j*, 2(365), 1-35.
- [45] Reinhold, J. C., Dewey, B. E., Carass, A., & Prince, J. L. (2019, March). Evaluating the impact of intensity normalization on MR image synthesis. In *Medical Imaging 2019: Image Processing* (Vol. 10949, pp. 890-898). SPIE
- [46] Ardekani, B. A., & Bachman, A. H. (2009). Model-based automatic detection of the anterior and posterior commissures on MRI scans. *Neuroimage*, 46(3), 677-682.
- [47] Tun, N. M., Gavrilov, A. I., & Paing, P. P. (2019, March). Human face recognition using combination of ZCA feature extraction method and deep neural network. In *2019 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)* (pp. 1-5). IEEE.
- [48] 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.
- [49] 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.
- [50] Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. *Nature* 585, 357–362 (2020). <https://doi.org/10.1038/s41586-020-2649-2>
- [51] TensorFlow. (2023). TensorFlow: A library for machine learning. (Version 2.11.0) [Software]. Retrieved from <https://www.tensorflow.org>
- [52] Desarrolladores de Django. (2023). Django: The Web framework for perfectionists with deadlines. Versión 4.2.3. Recuperado de <https://www.djangoproject.com/>

[53] React. (2023). React: A JavaScript library for building user interfaces (Version 17.0.2) [Software]. Facebook. <https://reactjs.org/>

[54] Axios. (2023). Axios: Promise based HTTP client for the browser and node.js (Version 1.4.0) [Software]. GitHub. <https://github.com/axios/axios>

[55] Material-UI. (2023). Material-UI: A popular React UI framework (Version 5.3.1) [Software]. GitHub. <https://github.com/mui-org/material-ui>

[56] Niivue. (2023). niivue/niivue: un visor de imágenes médicas basado en WebGL2. GitHub. <https://github.com/niivue/niivue>