

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL
Facultad de Ingeniería en Electricidad y Computación

“SEGMENTACIÓN DEL NÚCLEO SUBTALÁMICO Y SUSTANCIA
NEGRA CEREBRAL A PARTIR DE IMÁGENES DE
RESONANCIA MAGNÉTICA USANDO TÉCNICAS DE
INTELIGENCIA ARTIFICIAL”

EXAMEN COMPLEXIVO - COMPONENTE PRÁCTICO
INFORME PROFESIONAL

Previo a la obtención del Título de:

MAGISTER EN INGENIERÍA BIOMÉDICA

WALTER WLADIMIR PACHACAMA ESTRELLA

GUAYAQUIL – ECUADOR

AÑO: 2025

AGRADECIMIENTOS

Agradezco profundamente a mi familia y amigos por su apoyo incondicional a lo largo de todo este proceso. Su comprensión, paciencia y palabras de aliento han sido fundamentales en cada etapa de mi formación. Gracias por estar siempre presentes, por confiar en mí y por brindarme la fuerza necesaria para superar los desafíos que se presentaron en este proceso de maestría. Este logro no habría sido posible sin su compañía y respaldo constante.

DEDICATORIA

Dedico este trabajo con todo mi cariño y gratitud a mi mamá, mi papá y mi hermano. Gracias por su amor, su apoyo incondicional y por ser mi inspiración constante. Cada uno de ustedes ha sido un pilar fundamental en mi vida y en la realización de este logro. Este esfuerzo y este sueño alcanzado también son de ustedes.

DECLARACIÓN EXPRESA

Yo Walter Wladimir Pachacama Estrella acuerdo y reconozco que: La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá al autor, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor. El estudiante deberá procurar en cualquier caso de cesión de sus derechos patrimoniales incluir una cláusula en la cesión que proteja la vigencia de la licencia aquí concedida a la ESPOL.

La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, secreto empresarial, derechos patrimoniales de autor sobre software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por mí durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que me corresponda de los beneficios económicos que la ESPOL reciba por la explotación de mi innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique al autor que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 30 de junio de 2025

WALTER WLADIMIR PACHACAMA
ESTRELLA

EVALUADORES

PhD. Francis Loayza Paredes

Miembro Principal

PhD. Enrique Peláez Jarrin

Miembro Principal

RESUMEN

El presente trabajo nace de una inquietud muy concreta: ¿es posible ayudar a mejorar el tratamiento del Parkinson desde el campo de la inteligencia artificial? Partiendo de esa pregunta, se desarrolló este proyecto enfocado en la segmentación automática del núcleo subtalámico y la sustancia negra, dos estructuras cerebrales clave en el tratamiento quirúrgico de esta enfermedad, a partir de IRM.

Para alcanzar este objetivo, se construyó un flujo de trabajo completo que inició con un cuidadoso preprocesamiento de las imágenes, incluyendo corrección de inhomogeneidades del campo magnético, alineación AC-PC y recorte centrado en la región de interés. Posteriormente, se entrenaron dos modelos distintos de segmentación: uno basado en redes neuronales convolucionales tridimensionales (CNN U-Net 3D) y otro empleando una arquitectura más reciente conocida como Swin Transformer 3D.

Durante el proceso, se adaptaron los datos al contexto clínico y técnico del problema, se optimizó el entrenamiento en un entorno de cómputo limitado, y se evaluaron los resultados mediante métricas específicas de segmentación médica como Dice Score, precisión, recall y exactitud. El modelo CNN mostró un mejor desempeño general, logrando una mayor coincidencia entre sus predicciones y las segmentaciones reales. Sin embargo, el modelo Swin Transformer también ofreció resultados prometedores, especialmente en la detección de las regiones objetivo, aunque con un nivel más alto de falsos positivos.

ÍNDICE GENERAL

AGRADECIMIENTOS	ii
DEDICATORIA.....	iii
DECLARACIÓN EXPRESA.....	¡Error! Marcador no definido.
EVALUADORES	¡Error! Marcador no definido.
RESUMEN	vi
INTRODUCCIÓN	1
Capítulo 1	3
1. DESCRIPCIÓN DE LOS DATOS	3
2. PREPROCESAMIENTO DE DATOS.....	4
2.1 Corrección de inhomogeneidades del campo magnético.....	5
2.2 Reorientación ACPC	6
2.3 Recorte de la región de interés.....	7
2.4 Aumento de datos	8
Capítulo 2	9
3. Diseño de la arquitectura para segmentación.....	9
3.1. Redes neuronales convolucionales 3D	9
3.2. Vision Transformers 3D	10
3.2.1. Swin Transformer	10
3.3. Entrenamiento.....	11
3.4. Evaluación de LOS algoritmoS DE SEGMENTACIÓN	12
3.4.1. Dice Score	12
3.4.3. Sensibilidad o Recall	13
3.4.4. Exactitud (Accuracy).....	13
4. RESULTADOS.....	14
4.1. Modelo a partir de redes neuronales convolucionales.....	14
4.2. Modelo de transformers: Swin Transformer	17
4.3. Análisis de los resultados	20
CONCLUSIONES Y RECOMENDACIONES	22
CONCLUSIONES	22
RECOMENDACIONES	22

ÍNDICE DE FIGURAS

<i>Figura 1. IRM vista axial [7]</i>	2
<i>Figura 2. Region mesencefálica ampliada [8]</i>	2
<i>Figura 2.1: Comparación de imagen original vs imagen con "bias field correction" [autoría propia]</i>	5
<i>Figura 2.2: Comparación de imagen antes del registro vs imagen después del registro [autoría propia]</i>	6
<i>Figura 2.3: Comparación de imagen antes del recorte vs imagen después del recorte [autoría propia]</i>	8
<i>Figura 4.1: Comparación de la imagen general vs predicción vs máscara real obtenido con U-Net 3D [autoría propia]</i>	15
<i>Figura 4.2 Gráfica pérdida vs épocas [autoría propia]</i>	16
<i>Figura 4.3 Gráfica dice score vs épocas [autoría propia]</i>	17
<i>Figura 4.4: Comparación de la imagen general vs predicción vs máscara real obtenido con Swin Transformer [autoría propia]</i>	18
<i>Figura 4.5 Gráfica Dice Score vs Épocas [autoría propia]</i>	19
<i>Figura 4.6 Gráfica pérdidas vs épocas [autoría propia]</i>	20

ÍNDICE DE TABLAS

<i>Tabla 1-1. Exploración estadística inicial de las IRM</i>	4
<i>Tabla 4-1: Resultados de las métricas para el dataset de pruebas del modelo U-Net 3D</i>	14
<i>Tabla 4-2: Resultados de las métricas para el dataset de pruebas del modelo Swin Transformers</i>	20
<i>Tabla 4-3: Tabla comparativa de las métricas obtenidas en los modelos implementados</i>	20

INTRODUCCIÓN

La enfermedad de Parkinson es una condición que, lamentablemente, afecta a muchas personas, especialmente a partir de los cincuenta años [1]. Esta enfermedad va mucho más allá de lo que la mayoría de la gente imagina: no solo se trata de temblores o movimientos rígidos, sino que también implica complicaciones en el equilibrio y el control corporal [2]. Todo esto ocurre porque, poco a poco, ciertas neuronas encargadas de producir dopamina en el cerebro, específicamente en una zona llamada sustancia negra, comienzan a deteriorarse. Esta pérdida afecta áreas cerebrales importantes para el movimiento, como el putamen, el núcleo subtalámico y el estriado, que trabajan juntas para permitirnos movernos con precisión y fluidez [3].

La sustancia negra y el núcleo subtalámico son regiones pequeñas y, en las imágenes de resonancia magnética (IRM), a veces es difícil distinguirlas de las áreas vecinas [4], especialmente en ciertas secuencias. Esto complica mucho el trabajo de los médicos y cirujanos, ya que una localización precisa es fundamental para que los tratamientos funcionen bien y no causen daños innecesarios.

Cuando hablamos de tratamientos quirúrgicos, como la estimulación cerebral profunda, la exactitud en la localización de estas zonas es clave. Implantar electrodos o aplicar técnicas de ablación térmica o ultrasonido focalizado requiere saber exactamente dónde actuar, para así mejorar los síntomas motores y evitar efectos secundarios [5]. Si el procedimiento se realiza correctamente, se pueden reducir los temblores y la rigidez, y hasta mejorar aspectos no motores de la enfermedad, como el dolor, la depresión o los cambios en la expresión facial y el comportamiento social [6].

Por todo esto, en este proyecto me propongo explorar los modelos de segmentación en Deep Learning (DL) para identificar y delimitar estas estructuras cerebrales tan importantes. El objetivo es lograr que los tratamientos sean más seguros y efectivos, y así ayudar a quienes viven con esta enfermedad a tener una vida más plena y con menos limitaciones.

Las imágenes mostradas corresponden a cortes axiales de resonancia magnética cerebral ponderadas en T2. En la figura 1 se observa una visualización general del encéfalo, sin resaltar estructuras profundas de manera clara.

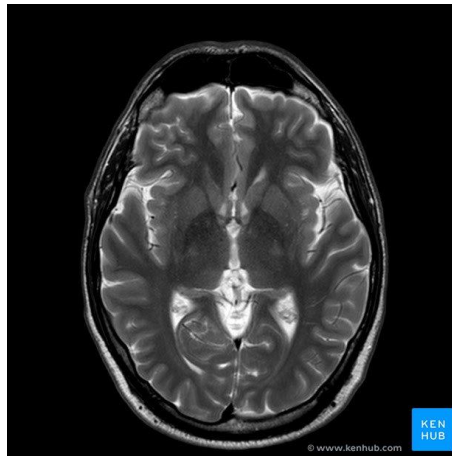


Figura 1. IRM vista axial [7]

La figura 2 permite una vista ampliada de la región mesencefálica, donde se localizan estructuras clave como la sustancia negra y el núcleo subtalámico. Estas estructuras aparecen como áreas más oscuras, lo que dificulta su delimitación manual y automática.

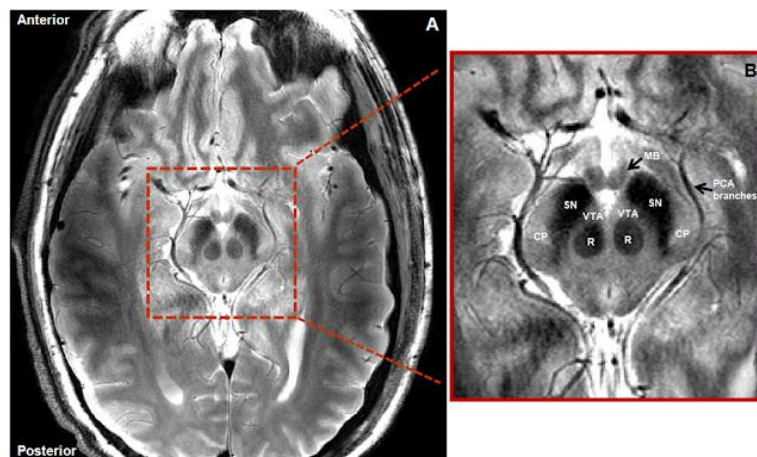


Figura 2. Region mesencefálica ampliada [8]

CAPÍTULO 1

Para el desarrollo del proyecto se utilizó Python como lenguaje de programación conjuntamente con diversas librerías para el análisis de IRM como son torch, numpy, nibabel y os.

La metodología general del proyecto consta de las siguientes partes:

1. Preprocesamiento de datos, esta etapa permite mejorar la calidad de las imágenes de resonancia magnética mediante corrección de artefactos, alineación espacial, normalización de intensidades y recorte centrado, asegurando condiciones uniformes para el entrenamiento de los modelos.
2. Diseño de la arquitectura para segmentación, aquí se desarrollan y entrenan modelos de DL (CNN y Swin Transformer) capaces de identificar y segmentar el núcleo subtalámico y la sustancia negra.
3. Evaluación de resultados, analiza cuantitativamente las predicciones utilizando métricas como Dice Score, precisión y exactitud, además de realizar visualizaciones comparativas.

1. DESCRIPCIÓN DE LOS DATOS

Las imágenes utilizadas en este proyecto corresponden a estudios de resonancia magnética cerebral en formato NIfTI, ya preseleccionados y anonimizados. Debido a la naturaleza del conjunto de datos, no se dispone de información demográfica adicional como edad, sexo o antecedentes clínicos de los sujetos, sin embargo, como primer paso se realizó una exploración estadística de las IRM pertenecientes a cinco pacientes. Todas las imágenes tienen dimensiones uniformes de $182 \times 218 \times 182$ voxels, lo que asegura una resolución constante entre las imágenes. En cuanto a las intensidades de los voxels, se observó una alta variabilidad entre pacientes como se observa en la tabla 1, tanto en la media como en la desviación estándar, la imagen IXI002_Guys2 presenta una intensidad promedio de 45.16 con una desviación estándar de 83.19, mientras que IXI014-HH2 alcanza una media de 145.87 y una desviación estándar de 265.62. Además, se identificaron valores máximos de

intensidad que superan los 1800 en algunos casos, lo cual sugiere la presencia de valores atípicos o artefactos de adquisición de la imagen lo que refuerza la necesidad del preprocesamiento.

Tabla 1-1. Exploración estadística inicial de las IRM

Nombre	Dimensiones	Media Intensidad	Desv. Estándar	Mínimo	Máximo
IXI002_Guys2	(182, 218, 182)	45.16	83.19	-0.000145	654.09
IXI012-HH2	(182, 218, 182)	117.47	218.70	-0.000727	1507.64
IXI013-HH2	(182, 218, 182)	119.24	223.06	-0.000202	1478.26
IXI014-HH2	(182, 218, 182)	145.87	265.62	-0.000620	1837.01
IXI015-HH2	(182, 218, 182)	109.76	204.07	-0.000455	1493.15

2. PREPROCESAMIENTO DE DATOS.

El preprocesamiento de imágenes es un paso fundamental para garantizar que los datos estén en condiciones adecuadas antes de ser analizados por un modelo de inteligencia artificial. En el caso de las IRM, este proceso permite corregir variaciones de intensidad, alinear anatómicamente los volúmenes, normalizar el contraste y enfocar regiones específicas. Estas mejoras no solo aumentan la calidad de los datos, sino que también permiten que el modelo aprenda patrones relevantes de manera más precisa, ayudando a reducir errores y mejorando el desempeño del sistema de segmentación.

Para el preprocesamiento se tuvo dos bases de datos de imágenes, en las cuales la primera contiene 60 imágenes de baja resolución en total, segmentadas el núcleo subtalámico y la sustancia negra, las cuales se utilizarán para el primer entrenamiento del algoritmo, mientras que la segunda base de datos, que contiene 35 imágenes en total pero de mayor resolución, servirá para el entrenamiento del modelo final, además se utilizó “bias field correction” que es la corrección del campo de polarización, es decir, ajustar señales indeseables de baja frecuencia que corrompe las IRM[9]. Así como también la orientación “ACPC, comisura anterior y posterior” usando la

transformación de cuerpo rígido “rigid body transformation”. Donde todas las modificaciones espaciales serán aplicadas en las imágenes T2 y al ground truth.

Posteriormente se recortó las imágenes de acuerdo a la región de interés. A continuación, se procedió a realizar las transformaciones necesarias para realizar el aumento de los datos. Estas transformaciones consistieron en rotaciones y aplicación de filtros, aumento o estiramiento y cizallamiento en los ejes X, Y, Z de forma independiente.

2.1 CORRECCIÓN DE INHOMOGENEIDADES DEL CAMPO MAGNÉTICO.

En el trabajo con IRM, muchas veces nos encontramos con que las fotos del cerebro no salen perfectamente uniformes. Hay zonas que pueden verse más claras o más oscuras de lo que realmente son, simplemente porque el campo magnético del resonador no es completamente uniforme [4]. Esto puede confundir a los médicos y a los programas de computadora cuando intentan identificar, por ejemplo, la sustancia negra o el núcleo subtalámico.

Aquí es donde entra en juego “Bias field correction” o corrección del campo de sesgo, como se observa en la figura 2.1 lo que hace este proceso es “suavizar” esas diferencias de brillo en la imagen, para que el cerebro se vea lo más parecido posible a la realidad [10].

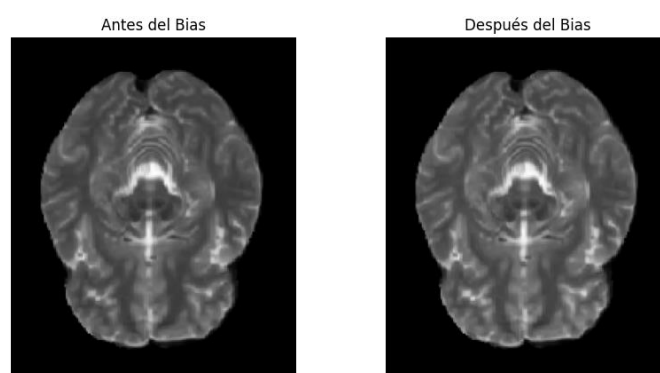


Figura 2.1: Comparación de imagen original vs imagen con “bias field correction” [autoría propia]

2.2 REORIENTACIÓN ACPC

El alineamiento ACPC (Anterior Commissure–Posterior Commissure) es un procedimiento fundamental en el análisis neuroanatómico basado en IRM. Consiste en orientar y estandarizar los volúmenes cerebrales tomando como referencia dos estructuras anatómicas claramente identificables: la comisura anterior y la comisura posterior. Estas comisuras son puntos clave del sistema nervioso central y su conexión define una línea virtual que atraviesa el cerebro, permitiendo un posicionamiento anatómico consistente entre sujetos y estudios [11].

En el contexto de nuestro proyecto, el alineamiento ACPC resulta esencial por varias razones. En primer lugar, la variabilidad anatómica natural entre pacientes, así como las diferencias en la posición de la cabeza durante la adquisición de las imágenes, pueden introducir sesgos significativos en la localización y comparación de estructuras cerebrales de interés, como la sustancia negra o el núcleo subtalámico. Como se puede apreciar en la figura 2.2 al alinear todas las imágenes según el plano ACPC, logramos una base común de referencia espacial, lo que facilita la comparación objetiva de resultados y la integración de datos provenientes de diferentes pacientes o momentos de seguimiento.

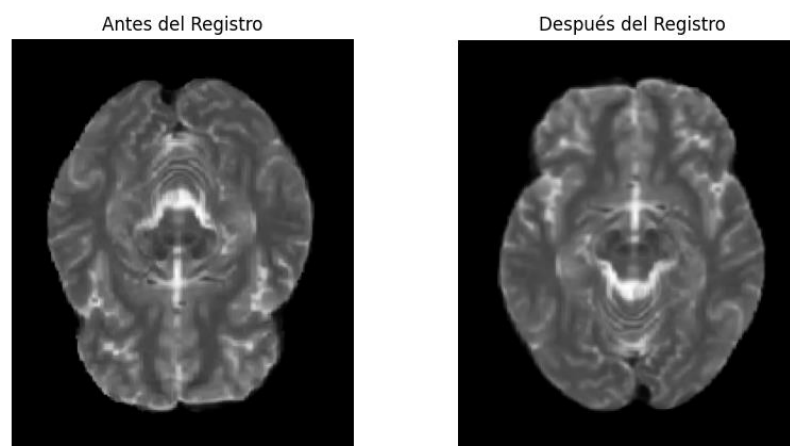


Figura 2.2: Comparación de imagen antes del registro vs imagen después del registro [autoría propia]

2.3 RECORTE DE LA REGIÓN DE INTERÉS

El recorte de imágenes es un proceso esencial cuando trabajamos con resonancia magnética cerebral, especialmente si nuestro objetivo es analizar estructuras específicas, como la sustancia negra o el núcleo subtalámico. En la práctica, una imagen de RM abarca todo el cerebro y, aunque es muy útil para tener un panorama general, muchas veces solo necesitamos enfocarnos en una región particular para nuestros análisis o procedimientos [12].

Para enfocar el análisis únicamente en la sustancia negra y núcleo subtalámico, se realizó un recorte tridimensional en cada imagen, extrayendo un volumen de interés (VOI) centrado en el núcleo subtalámico y la sustancia negra. Este recorte se aplicó de manera automática mediante un algoritmo en Python, utilizando la función “center_crop” de la librería TorchIO, la cual permite generar un recorte simétrico a partir del centro de una imagen. Se definió un tamaño fijo de 64x64x64 vóxeles, que resultó adecuado tras analizar visualmente que este volumen cubría por completo la región de interés en todos los casos, como se muestra en la figura 2.3.

Para establecer el centro de cada recorte, primero se calculó el centro de masa de la máscara asociada a cada imagen mediante la función “center_of_mass” de la librería SciPy, la cual nos permite analizar los vóxeles que tienen valores positivos y devuelve las coordenadas (x, y, z) del punto central promedio. Este punto fue usado como referencia para ubicar el recorte de forma precisa sobre las estructuras objetivo. La función “center_crop” generó entonces un nuevo volumen, más pequeño, pero centrado exclusivamente en la región relevante, sin necesidad de seleccionar coordenadas manuales. Este procedimiento no solo redujo el tamaño de los datos, sino que también permitió que los modelos de segmentación implementados, específicamente una arquitectura CNN tipo U-Net 3D y un modelo basado en Swin Transformer 3D, enfocaran su capacidad de aprendizaje directamente en el área de interés, mejorando la eficiencia del procesamiento y la precisión en la identificación de estructuras anatómicas pequeñas y de bajo contraste.

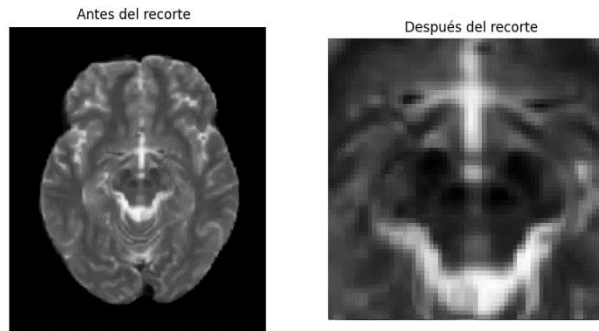


Figura 2.3: Comparación de imagen antes del recorte vs imagen después del recorte [autoría propia]

2.4 AUMENTO DE DATOS

Como parte del desarrollo del modelo de segmentación, se consideró aplicar técnicas de aumento de datos (data augmentation) para mejorar el rendimiento del entrenamiento, así como también para mejorar el desbalance de los datos. Esta estrategia busca generar nuevas variaciones a partir de las imágenes originales, de manera que el modelo pueda aprender con mayor diversidad y adaptarse mejor a diferentes formas anatómicas sin necesidad de ampliar el conjunto real de imágenes [13].

Durante el proceso de aumento de datos se aplicaron transformaciones aleatorias tanto a las imágenes como a sus respectivas máscaras, con el fin de mejorar la capacidad del modelo para generalizar frente a nuevos casos. Se utilizaron tres tipos principales de transformaciones: volteo aleatorio en los tres ejes (flip), rotación affine aleatoria y ligeras deformaciones espaciales. Estas modificaciones se implementaron utilizando la librería TorchIO, diseñada específicamente para el procesamiento y aumento de datos médicos en formato 3D, como IRM. Se eligió TorchIO debido a que ofrece un conjunto robusto y fácilmente integrable de transformaciones compatibles con PyTorch, permitiendo mantener la coherencia espacial entre imágenes y máscaras durante el aumento de datos. En total se partió de 95 imágenes originales, y a cada una se le generaron 3 nuevas versiones transformadas, resultando en un conjunto total de 380 imágenes. Este aumento permitió al modelo entrenarse con un conjunto más diverso, mejorando su robustez frente a variaciones anatómicas y de adquisición.

CAPÍTULO 2

3. Diseño de la arquitectura para segmentación

3.1. REDES NEURONALES CONVOLUCIONALES 3D

Para la segmentación del núcleo subtalámico y la sustancia negra, se empleó una red neuronal convolucional tridimensional (3D-CNN) basada en la arquitectura U-Net 3D, adaptada para tareas de segmentación multiclase [14]. Esta arquitectura ha demostrado ser especialmente eficaz para delimitar estructuras anatómicas pequeñas gracias a su capacidad para integrar características espaciales locales y globales [14].

El diseño implementado presenta una estructura en forma de "U", compuesta por dos fases simétricas: una etapa de codificación (encoder) y otra de decodificación (decoder), conectadas por un bloque central (bridge). En la fase de codificación, se utilizan bloques secuenciales de dos convoluciones 3D (con kernel $3 \times 3 \times 3$), cada una seguida por una normalización por lotes (BatchNorm3d) y una función de activación ReLU [15]. Esta fase también incluye operaciones de reducción espacial mediante MaxPooling3D, lo cual permite extraer representaciones profundas de la imagen [15].

El bloque central conecta ambos extremos y profundiza la extracción de características mediante convoluciones adicionales. Posteriormente, la fase de decodificación emplea convoluciones transpuestas (ConvTranspose3D) para recuperar progresivamente la resolución espacial. A través de concatenaciones tipo "skip connection", se incorporan características de la fase de codificación que ayudan a preservar la información de borde. Finalmente, una capa convolucional $1 \times 1 \times 1$ proyecta el volumen final al número de clases objetivo, generando un mapa de segmentación multiclase.

Durante el entrenamiento, se utilizaron volúmenes 3D normalizados y centrados. Las máscaras de segmentación incluyeron cuatro etiquetas, correspondientes a estructuras diferenciadas. Para mejorar la generalización del modelo, se aplicaron técnicas de aumento de datos mediante deformaciones geométricas como rotaciones aleatorias, escalados y volteos.

Como función de pérdida, se utilizó la combinación de CrossEntropyLoss ponderada por clase y Dice Loss, lo que permitió equilibrar el desbalance entre clases y mejorar la segmentación de estructuras pequeñas. El entrenamiento se llevó a cabo en Google Colab, aprovechando aceleración por GPU, y se guardó automáticamente el mejor modelo en base a la pérdida de validación.

3.2. VISION TRANSFORMERS 3D

En los últimos años, el campo de la inteligencia artificial ha experimentado avances notables con la aparición de modelos basados en *Transformers*, inicialmente diseñados para el procesamiento de lenguaje natural [16]. Sin embargo, su estructura versátil ha permitido adaptarlos también al análisis de imágenes, dando lugar a una nueva categoría conocida como *Vision Transformers (ViT)*. A diferencia de las redes convolucionales, que extraen patrones locales a través de filtros, los Vision Transformers procesan la información dividiendo la imagen en pequeñas secciones (o *patches*) y analizando las relaciones entre ellas mediante mecanismos de atención. Esto les permite capturar contextos más amplios, lo cual es particularmente valioso en tareas como la segmentación de IRM, donde las estructuras cerebrales pueden presentar variaciones anatómicas sutiles pero relevantes [17].

3.2.1. SWIN TRANSFORMER

En este proyecto se empleó una arquitectura basada en Swin Transformer tridimensional (Swin 3D). Esta alternativa moderna a las redes convolucionales permite aprovechar mejor las relaciones espaciales de largo alcance dentro del volumen cerebral, lo que puede resultar especialmente útil para estructuras pequeñas y de bordes poco definidos [18].

El modelo utilizado se estructuró en dos fases principales: una fase de codificación que extrae características jerárquicas del volumen mediante ventanas deslizantes (window-based self-attention), y una fase de decodificación que reconstruye la segmentación a través de operaciones de interpolación y convoluciones. Además, se integran mecanismos de atención que permiten que el modelo aprenda a identificar patrones contextuales complejos sin depender de filtros convolucionales locales [19].

Para este trabajo, se adaptó una versión ligera del Swin Transformer 3D, ajustada a las restricciones de memoria disponibles en Google Colab. Las imágenes fueron recortadas y redimensionadas a volúmenes de $48 \times 48 \times 48$ voxeles, centrados en la región anatómica de interés. Antes del entrenamiento, los datos fueron corregidos por campo de bias, alineados a la orientación AC-PC, normalizados en intensidad y etiquetados en cuatro clases diferenciadas.

El entrenamiento se llevó a cabo con un conjunto mixto de imágenes de baja y alta resolución, usando aumento de datos que incluye rotaciones aleatorias, escalamiento y volteos tridimensionales, con el fin de robustecer el aprendizaje del modelo. La función de pérdida combinó Dice Loss y Cross Entropy ponderada por clase, para mitigar el desequilibrio entre regiones segmentadas y fondo.

El modelo fue entrenado durante 100 épocas, utilizando optimización con Adam (técnica utilizada para actualizar los pesos de una red neuronal durante el entrenamiento, con el objetivo de minimizar la función de pérdida) y un aprendizaje supervisado completo. Una vez finalizado, se aplicó el modelo entrenado para generar predicciones sobre volúmenes no vistos. Las salidas fueron evaluadas utilizando métricas cuantitativas como Dice Score, precisión, recall y exactitud, y se realizaron visualizaciones de las predicciones superpuestas sobre las imágenes originales para su análisis cualitativo.

3.3. ENTRENAMIENTO

En este trabajo se emplearon dos enfoques diferentes para la segmentación del núcleo subtalámico y sustancia negra partir de IRM: un modelo basado en redes neuronales convolucionales tridimensionales (CNN 3D) y otro fundamentado en arquitecturas de tipo Transformer, específicamente Swin Transformer 3D. El modelo CNN que se implementó corresponde a una versión de U-Net adaptada a datos volumétricos, con bloques de codificación y decodificación que permiten extraer características espaciales locales y reconstruir el mapa de segmentación. Esta arquitectura es particularmente eficaz para capturar detalles anatómicos finos en regiones pequeñas gracias a sus operaciones de convolución y conexiones de salto, lo que la convierte en una opción robusta para tareas de segmentación médica.

Por otro lado, Swin Transformer 3D introduce una estrategia distinta al aprovechar mecanismos de atención auto-regresiva en ventanas deslizantes jerárquicas, lo que permite modelar relaciones espaciales de largo alcance sin perder la información local. Esta característica le otorga una mayor capacidad de generalización ante variaciones morfológicas complejas. Ambos modelos fueron entrenados desde cero utilizando imágenes T2 preprocesadas, alineadas, normalizadas y recortadas a un volumen de 48x48x48 voxels en el caso de Swin Transformer, mientras que en CNN se pudo entrenar con imágenes de volumen 64x64x64 voxels. El entrenamiento se realizó en la plataforma Google Colab, con soporte GPU (NVIDIA Tesla T4), utilizando la función de pérdida Dice Loss y el optimizador Adam. Para evaluar el desempeño, se aplicaron métricas como Dice Score, precisión y exactitud, lo que permitió realizar una comparación cuantitativa entre ambas arquitecturas en el contexto de la segmentación del núcleo subtalámico y la sustancia negra.

3.4. EVALUACIÓN DE LOS ALGORITMOS DE SEGMENTACIÓN

La evaluación de los modelos se llevará a cabo utilizando métricas específicas para segmentación médica, como el Dice Score, la precisión (precision), la sensibilidad (recall) y la exactitud (accuracy) [20]. Estas métricas permitirán cuantificar la calidad de las segmentaciones generadas por el modelo en comparación con las máscaras reales. La evaluación se aplicará tanto al modelo CNN como al Swin Transformer, con el objetivo de analizar cuál de las dos arquitecturas presenta un mejor desempeño en la segmentación.

3.4.1. DICE SCORE

Esta métrica evalúa la superposición entre la máscara predicha y el ground truth (máscara real). Es especialmente útil en segmentación médica, donde el desbalance entre clases es común [21]. Definida por la ecuación (2.1)

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (2.1)$$

donde:

A: conjunto de vóxeles predichos como positivos (por el modelo).

B: conjunto de vóxeles reales positivos (ground truth).

3.4.2. PRECISIÓN

Indica la proporción de verdaderos positivos entre todos los positivos predichos. Es decir, de todo lo que el modelo dijo que era una estructura [4]. Tal como se aprecia en la ecuación (2.2)

$$Precision = \frac{TP}{TP + FP} \quad (2.2)$$

Donde:

TP: verdaderos positivos

FP: falsos positivos

3.4.3. SENSIBILIDAD O RECALL

Mide la capacidad del modelo para encontrar todos los positivos reales. Es decir, de todo lo que realmente era la estructura [22]. Como se define en la ecuación (2.3)

$$Precision = \frac{TP}{TP + FN} \quad (2.3)$$

Donde:

FN: falsos negativos

3.4.4. EXACTITUD (ACCURACY)

Indica la proporción total de aciertos (positivos y negativos) entre todos los vóxeles analizados [22]. Se define en la ecuación (2.4)

$$Exactitud = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.4)$$

Donde:

TN: verdaderos negativos (cuando el modelo correctamente predice que un voxel no pertenece a la estructura).

4. RESULTADOS

En el presente capítulo se muestran los resultados obtenidos con los dos modelos indicados en la sección anterior:

4.1. MODELO A PARTIR DE REDES NEURONALES CONVOLUCIONALES

Una vez completado el proceso de entrenamiento con el modelo CNN 3D U-Net, se procedió a evaluar su desempeño sobre un conjunto de validación compuesto por imágenes previamente segmentadas del núcleo subtalámico y la sustancia negra. Las métricas obtenidas reflejan una capacidad del modelo para aproximarse de forma razonable a la segmentación manual, aunque con ciertas variaciones dependiendo del caso.

A continuación, se resumen los resultados obtenidos en cinco muestras representativas:

Tabla 4-1: Resultados de las métricas para el dataset de pruebas del modelo U-Net 3D

IMAGEN	DICE	PRECISION	RECALL	ACCURACY
IXI002_GUYS2_NORM	0,555	0,5157	0,6008	0,9973
IXI012-HH2_NORM	0,7632	0,8639	0,6835	0,9983
IXI013-HH2_NORM	0,8072	0,7538	0,8689	0,9985
IXI014-HH2_NORM	0,7815	0,7984	0,7654	0,9985
IXI015-HH2_NORM	0,7532	0,8033	0,709	0,9985

Como se observa, el valor promedio del Dice Score en estas muestras se encuentra en un rango moderado (aproximadamente entre 0.55 y 0.80), lo cual indica una superposición parcial entre las segmentaciones predichas y las reales. Si bien no se

alcanza una coincidencia perfecta, estos resultados reflejan que el modelo ha logrado aprender características relevantes de las estructuras de interés. Además, los valores de precisión y recall se mantienen relativamente equilibrados, lo que sugiere que el modelo presenta un compromiso aceptable entre la detección de regiones verdaderas y la minimización de errores por omisión o exceso.

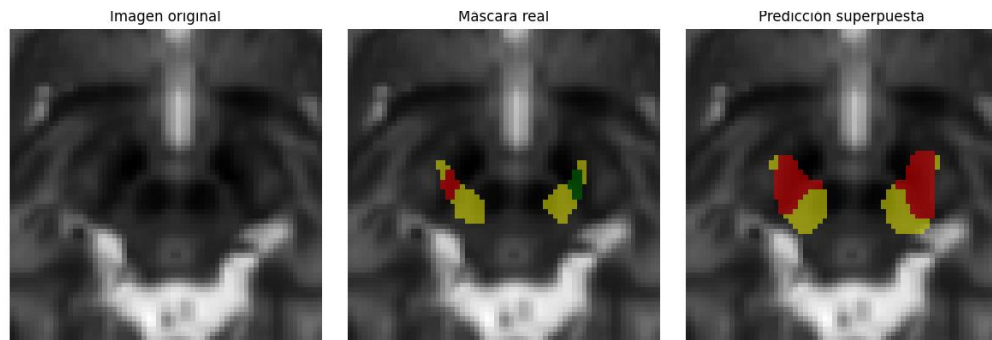


Figura 4.1: Comparación de la imagen general vs predicción vs máscara real obtenido con U-Net 3D [autoría propia]

En términos visuales, como se observa en la figura 4.1 la comparación entre la imagen original, la máscara real y la máscara predicha muestra que el modelo es capaz de capturar correctamente la forma y posición del núcleo subtalámico y la sustancia negra, aunque existen algunos casos en los que se observan predicciones más limitadas o regiones parcialmente subsegmentadas.

A continuación, se muestra la gráfica obtenida de pérdida vs época que muestra la evolución de la función de pérdida (*loss*) durante el proceso de entrenamiento y validación del modelo U-Net 3D multiclase. En el eje **X** se representan las épocas (de 1 a 100), y en el eje **Y** se muestra el valor de la pérdida, se puede observar que la línea azul corresponde a la pérdida calculada sobre el conjunto de entrenamiento, mientras que la línea naranja representa la pérdida sobre el conjunto de validación.

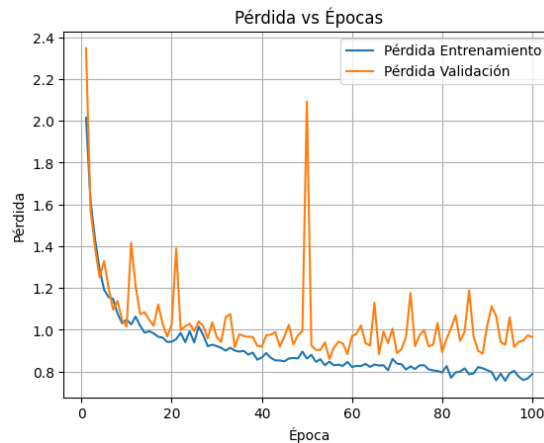


Figura 4.2 Gráfica pérdida vs épocas [autoría propia]

En la figura 4.2 se aprecia que, al inicio del entrenamiento, las pérdidas de entrenamiento y validación presentan valores altos, reflejo de la incertidumbre típica de un modelo recién inicializado. Conforme avanzan las primeras épocas, ambas curvas descienden con rapidez debido a los ajustes iniciales de los parámetros. Más adelante, la pérdida de entrenamiento continúa disminuyendo de forma gradual, mientras que la de validación exhibe pequeñas oscilaciones atribuibles a la variabilidad entre muestras. Aun así, las dos curvas se mantienen próximas entre sí, lo que sugiere que no se produce un sobreajuste significativo y que el modelo optimiza de manera adecuada la función objetivo a lo largo del proceso de aprendizaje.

En la figura 4.3 se presenta la gráfica de Dice Score vs Épocas la cual muestra la evolución del Dice Score promedio en el conjunto de validación durante el entrenamiento. Este valor mide la superposición entre las máscaras predichas y las reales, siendo un indicador directo del rendimiento del modelo en la tarea de segmentación. El eje X representa las épocas, mientras que el eje Y indica el valor del Dice Score.

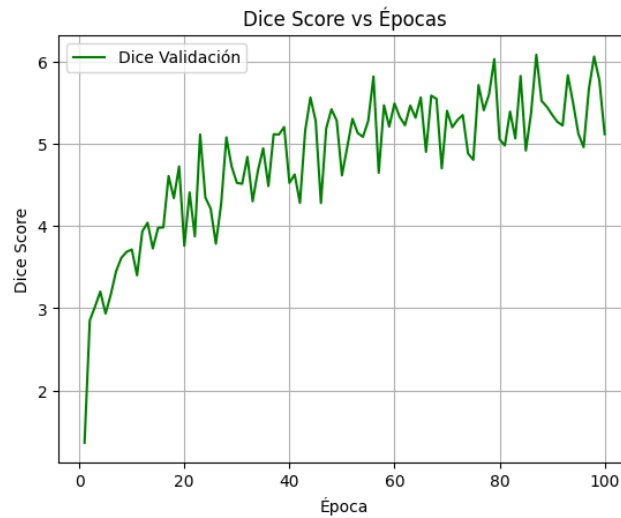


Figura 4.3 Gráfica dice score vs épocas [autoría propia]

Se puede apreciar que a medida que avanzan las épocas, se observa un incremento progresivo del Dice Score, lo que indica que el modelo mejora su capacidad para segmentar el núcleo subtalámico y sustancia negra. Aunque se presentan pequeñas variaciones, la tendencia ascendente es clara, lo que sugiere que el modelo está aprendiendo patrones relevantes. Esto valida que la red convolucional ha sido eficaz en la tarea de segmentación.

4.2. MODELO DE TRANSFORMERS: SWIN TRANSFORMER

Para esta parte del proyecto, decidimos trabajar con un modelo Swin Transformer en 3D, adaptado especialmente para nuestras imágenes médicas. A diferencia de los modelos tradicionales, este enfoque nos permite analizar los volúmenes de resonancia magnética desde distintas perspectivas, reconociendo patrones espaciales más complejos gracias al uso de una técnica llamada “atención”, que actúa como si el modelo observara distintas partes de la imagen con mayor detalle.

El proceso comenzó dividiendo las imágenes en pequeños bloques para que el modelo pueda analizarlos más fácilmente. Luego, se fue entrenando poco a poco para aprender a identificar y resaltar las regiones más importantes en cada imagen, en este caso el núcleo subtalámico y la sustancia negra.

Diseñamos el modelo para que primero recoja la información, la resuma, y luego la reconstruya como una especie de “máscara” que marca exactamente las zonas que queremos segmentar. Para que el aprendizaje fuera efectivo, usamos una función especial que mide qué tan bien la predicción del modelo se parece a la realidad.

Además, como las imágenes que usamos ocupan bastante memoria, trabajamos con versiones más pequeñas (48x48x48), lo que permitió que el entrenamiento fuera más rápido y sin errores.

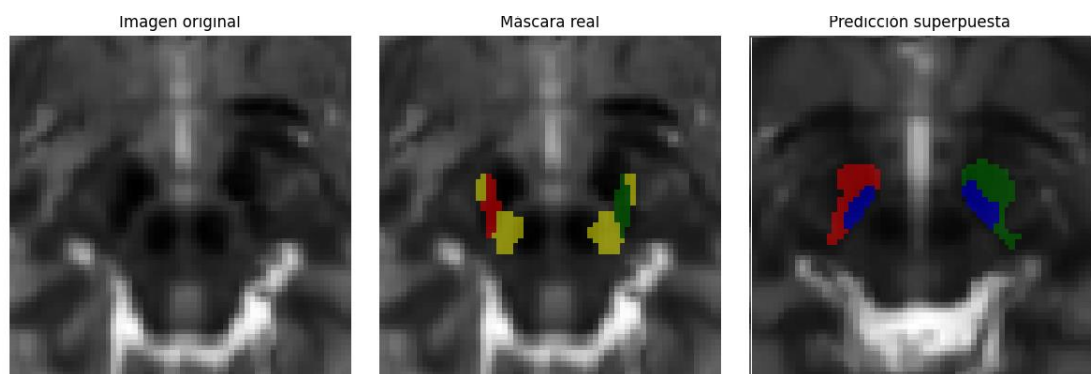


Figura 4.4: Comparación de la imagen general vs predicción vs máscara real obtenido con Swin Transformer [autoría propia]

En términos visuales, como se observa en la figura 4.4 la comparación entre la imagen original, la máscara real y la máscara predicha muestra que el modelo es capaz de capturar la forma y posición del núcleo subtalámico y la sustancia negra, aunque existen casos en los que se observan predicciones más limitadas o regiones parcialmente subsegmentadas.

La figura 4.5 presenta la evolución del Dice Score durante las 100 épocas de entrenamiento del modelo Swin Transformer 3D. Inicialmente, se observa un aprendizaje progresivo con valores bajos de segmentación, lo cual es esperable dado que el modelo aún no ha identificado patrones relevantes. A partir de la época 55, se produce un cambio notable con una mejora rápida y sostenida del rendimiento. Este comportamiento podría atribuirse a la capacidad del optimizador Adam para ajustar dinámicamente los gradientes durante el entrenamiento, permitiendo que el modelo

escape de mínimos locales y mejore la generalización. Adicionalmente, el uso de técnicas de aumento de datos puede haber contribuido a exponer al modelo a representaciones más diversas y útiles para el aprendizaje, favoreciendo un avance más efectivo en las últimas fases del entrenamiento.

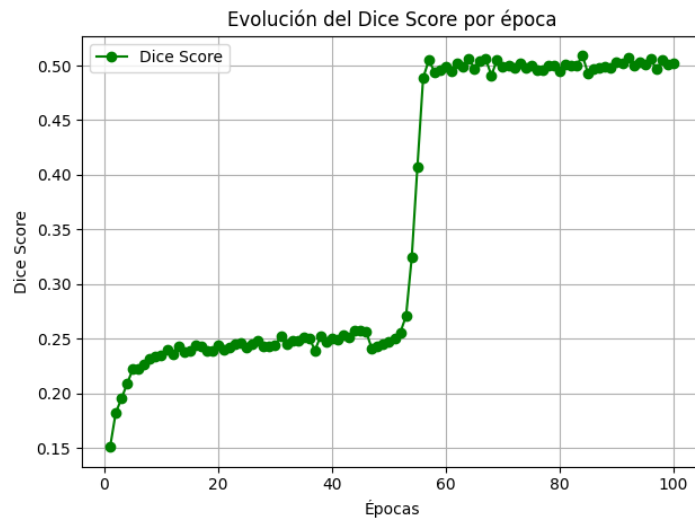


Figura 4.5 Gráfica Dice Score vs Épocas [autoría propia]

En la figura 4.6 se representa la evolución de la pérdida (loss) durante el proceso de entrenamiento y validación. La pérdida combina dos funciones: *CrossEntropyLoss* y *DiceLoss*, las cuales permiten al modelo ajustar sus parámetros para minimizar el error en la predicción de clases segmentadas. Se puede observar que, desde las primeras épocas, ambas curvas (azul para entrenamiento y naranja para validación) disminuyen rápidamente, lo que indica que el modelo está aprendiendo patrones relevantes en los datos. A lo largo del entrenamiento, ambas curvas siguen reduciéndose de manera paralela, lo cual sugiere un comportamiento estable sin evidencia de sobreajuste. La cercanía entre ambas curvas también muestra que el modelo generaliza relativamente bien a los datos de validación, lo cual es fundamental en tareas de segmentación.

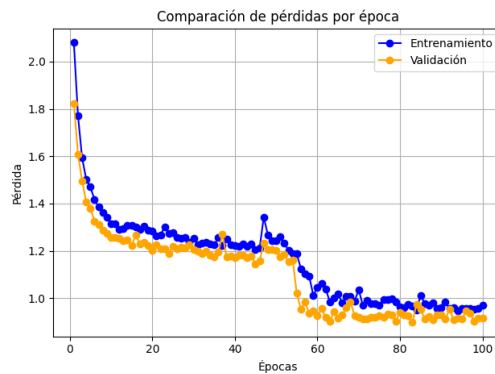


Figura 4.6 Gráfica pérdidas vs épocas [autoría propia]

A continuación, se resumen los resultados obtenidos en cinco muestras representativas:

Tabla 4-2: Resultados de las métricas para el dataset de pruebas del modelo Swin Transformers

IMAGEN	DICE	PRECISION	RECALL	ACCURACY
IXI012-HH2_NORM	0,5656	0,4593	0,7359	0,9893
IXI013-HH2_NORM	0,6015	0,4921	0,7736	0,9914
IXI014-HH2_NORM	0,6099	0,5268	0,724	0,9925
IXI015-HH2_NORM	0,5435	0,42	0,7701	0,9902
IXI034-HH2_NORM_READY	0,5177	0,403	0,7237	0,9904

4.3. ANÁLISIS DE LOS RESULTADOS

Tabla 4-3: Tabla comparativa de las métricas obtenidas en los modelos implementados

IMAGEN	U-NET 3D				SWIN TRANSFORMER			
	dice	precision	recall	accuracy	dice	precision	recall	accuracy
IXI012-HH2_NORM	0,7632	0,8639	0,6835	0,9983	0,5656	0,4593	0,7359	0,9893
IXI013-HH2_NORM	0,8072	0,7538	0,8689	0,9985	0,6015	0,4921	0,7736	0,9914
IXI014-HH2_NORM	0,7815	0,7984	0,7654	0,9985	0,6099	0,5268	0,724	0,9925
IXI015-HH2_NORM	0,7532	0,8033	0,709	0,9985	0,5435	0,42	0,7701	0,9902

Una vez completado el entrenamiento de ambos modelos, se procedió a evaluar su rendimiento utilizando un mismo grupo de imágenes y aplicando las mismas métricas: Dice Score, precisión, recall y exactitud.

Los resultados muestran que el modelo basado en redes neuronales convolucionales (CNN) obtuvo un mejor desempeño general. En casi todos los casos analizados, la CNN presentó valores más altos tanto en la coincidencia de las predicciones con las máscaras reales (medido con el Dice Score), como en la precisión y exactitud. Por ejemplo, en la imagen IXI013-HH2_norm, la CNN logró un Dice de 0,8072, mientras que el modelo Swin Transformer apenas alcanzó 0,6015.

Esto sugiere que, en este caso, la CNN fue capaz de aprender de manera más efectiva la forma y localización del núcleo subtalámico y la sustancia negra, mientras que el modelo basado en Swin Transformer tuvo más dificultades para ajustarse al patrón de los datos. Una posible razón es que este último modelo, aunque más moderno, necesita muchos más datos para entrenarse correctamente. Como en nuestro caso los datos son limitados, la CNN logró adaptarse mejor.

Sin embargo, el modelo Swin no es del todo malo. De hecho, en varias imágenes obtuvo un recall relativamente alto, lo que significa que fue capaz de identificar bien las regiones de interés, aunque también cometió más errores al incluir zonas que no correspondían (de ahí su menor precisión). Esto sugiere que tiene potencial, pero requiere más datos o ajustes para mostrar lo que realmente puede hacer.

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES

- El modelo CNN resultó ser el más efectivo para el proyecto. A pesar de ser una arquitectura más tradicional, logró aprender con mayor precisión la forma y ubicación del núcleo subtalámico y la sustancia negra, superando claramente al modelo Swin Transformer en las métricas evaluadas.
- El modelo Swin Transformer, no alcanzó el rendimiento esperado. Su desempeño fue más bajo en casi todos los casos, lo que podría deberse a la necesidad de más datos o mayor capacidad de cómputo para explotar todo su potencial.
- El preprocesamiento tuvo un papel clave. Pasos como la normalización, la corrección del bias y la orientación AC-PC facilitaron que los modelos trabajaran sobre datos más homogéneos y comparables, mejorando la calidad del entrenamiento y las predicciones.
- El uso de imágenes médicas requiere enfoques cuidadosamente diseñados. A diferencia de otros tipos de imágenes, aquí cada pequeño error tiene un peso importante, por lo que es fundamental adaptar tanto los modelos como los parámetros al tipo de datos y al problema específico.

RECOMENDACIONES

- Se sugiere incrementar la cantidad de imágenes de resonancia magnética (IRM) reales, es decir, incorporar datos adicionales provenientes de más sujetos o pacientes. Esto permitiría ampliar la diversidad anatómica y mejorar la generalización del modelo, particularmente en el caso de los Vision Transformers, que suelen beneficiarse significativamente de grandes volúmenes de datos reales para alcanzar un rendimiento óptimo. Cabe aclarar que, si bien en este proyecto se utilizó *data augmentation* para generar variaciones sintéticas a partir de las imágenes originales, estas técnicas no sustituyen completamente la necesidad de contar con un conjunto amplio y variado de imágenes reales.

BIBLIOGRAFÍA

- [1] L. M. Ariza-Serrano, J. Guerrero-Vega, P. Ortiz, y C. L. Moreno-Lopez, «Caracterización de pacientes con enfermedad de Parkinson en un centro de referencia de la ciudad de Bogotá, Colombia», *Acta Neurológica Colomb.*, vol. 32, n.º 3, pp. 203-208, sep. 2016.
- [2] K. P. Chaves Morales, D. S. Padilla Elizondo, y R. Vargas Fernández, «Enfermedad de Parkinson», *Rev. Médica Sinerg.*, vol. 7, n.º 2, p. 5, 2022.
- [3] A. T. Vertinsky *et al.*, «Localization of the Subthalamic Nucleus: Optimization with Susceptibility-Weighted Phase MR Imaging», *Am. J. Neuroradiol.*, vol. 30, n.º 9, pp. 1717-1724, oct. 2009, doi: 10.3174/ajnr.A1669.
- [4] F. E. Magallanes Pinargote, J. F. Nebel Dunn, y C. E. D. Peláez Jarrín, «Segmentación automática de la Sustancia Negra cerebral y Núcleo Subtalámico a partir de imágenes 3D de resonancia magnética», Thesis, ESPOL. FIEC., 2022. Accedido: 27 de mayo de 2025. [En línea]. Disponible en: <http://www.dspace.espol.edu.ec/handle/123456789/57043>
- [5] V. D. Sharma, M. Patel, y S. Miocinovic, «Surgical Treatment of Parkinson's Disease: Devices and Lesion Approaches», *Neurotherapeutics*, vol. 17, n.º 4, pp. 1525-1538, oct. 2020, doi: 10.1007/s13311-020-00939-x.
- [6] A. Avena *et al.*, «Spectral Topography of the Subthalamic Nucleus to Inform Next-Generation Deep Brain Stimulation», *Mov. Disord.*, vol. 38, n.º 5, pp. 818-830, 2023, doi: 10.1002/mds.29381.
- [7] «Resonancia magnética cerebral normal», Kenhub. Accedido: 27 de junio de 2025. [En línea]. Disponible en: <https://www.kenhub.com/es/library/anatomia-es/rm-cerebral-normal>
- [8] J. E. L. Salazar, «NeuroAnatomía Dr. William Ganem: Parálisis Supranuclear Progresiva - Síndrome de Richardson», NeuroAnatomía Dr. William Ganem. Accedido: 27 de junio de 2025. [En línea]. Disponible en: https://pgmiucna.blogspot.com/2012/05/paralisis-supranuclear-progresiva_27.html
- [9] T. Goldfryd, S. Gordon, y T. R. Raviv, «Deep Semi-Supervised Bias Field Correction Of Mr Images», en *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, Nice, France: IEEE, abr. 2021, pp. 1836-1840. doi: 10.1109/ISBI48211.2021.9433889.
- [10] J. Lu, Z. Wang, E. Bier, S. Leewiwatwong, D. Mummy, y B. Driehuys, «Bias field correction in hyperpolarized 129 Xe ventilation MRI using templates derived by RF-depolarization mapping», *Magn. Reson. Med.*, vol. 88, n.º 2, pp. 802-816, ago.

2022, doi: 10.1002/mrm.29254.

[11] A. L. Alaminos-Bouza, «Imaging, Stereotactic Space and Targeting», en *Functional Neurosurgery*, A. Cukiert, Ed., Alaúde Editorial., 2014, pp. 67-79. Accedido: 27 de mayo de 2025. [En línea]. Disponible en: <https://hal.science/hal-02867315>

[12] D. S. Chaves Burbano, «Automatic segmentation of parapharyngeal fat in magnetic resonance imaging using U-net 3D», mar. 2024, Accedido: 27 de mayo de 2025. [En línea]. Disponible en: <http://hdl.handle.net/10554/66844>

[13] J. Nalepa, M. Marcinkiewicz, y M. Kawulok, «Data Augmentation for Brain-Tumor Segmentation: A Review», *Front. Comput. Neurosci.*, vol. 13, dic. 2019, doi: 10.3389/fncom.2019.00083.

[14] S. Esmaeilzadeh, Y. Yang, y E. Adeli, «End-to-End Parkinson Disease Diagnosis using Brain MR-Images by 3D-CNN», 13 de junio de 2018, *arXiv: arXiv:1806.05233*. doi: 10.48550/arXiv.1806.05233.

[15] N. U. Islam, R. Khanam, y A. Kumar, «Using 3D CNN for classification of Parkinson's disease from resting-state fMRI data», *J. Eng. Appl. Sci.*, vol. 70, n.º 1, p. 89, ago. 2023, doi: 10.1186/s44147-023-00236-2.

[16] C. J. Farfán, A. B. Chonillo, A. Constantine, E. Valarezo, F. R. Loayza, y E. P. Jarrín, «A Sharper Look: Edge Detection and Light-Weight 3D Vision Transformers for Parkinsonian Disorder Classification on MRI Imaging», en *2024 IEEE ANDESCON*, sep. 2024, pp. 01-06. doi: 10.1109/ANDESCON61840.2024.10755869.

[17] E. U. Henry, O. Emebob, y C. A. Omonhinmin, «Vision Transformers in Medical Imaging: A Review», 18 de noviembre de 2022, *arXiv: arXiv:2211.10043*. doi: 10.48550/arXiv.2211.10043.

[18] S. S. Hussain *et al.*, «A swin transformer and CNN fusion framework for accurate Parkinson disease classification in MRI», *Sci. Rep.*, vol. 15, n.º 1, p. 15117, abr. 2025, doi: 10.1038/s41598-025-93671-5.

[19] D. Yoo, J. Kim, y J. Yoo, «FSwin Transformer: Feature-Space Window Attention Vision Transformer for Image Classification», *IEEE Access*, vol. 12, pp. 72598-72606, 2024, doi: 10.1109/ACCESS.2024.3394539.

[20] A. W. Setiawan, «Image Segmentation Metrics in Skin Lesion: Accuracy, Sensitivity, Specificity, Dice Coefficient, Jaccard Index, and Matthews Correlation Coefficient», en *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, nov. 2020, pp. 97-102. doi: 10.1109/CENIM51130.2020.9297970.

[21] S. Gul, M. S. Khan, A. Bibi, A. Khandakar, M. A. Ayari, y M. E. H. Chowdhury, «Deep learning techniques for liver and liver tumor segmentation: A review», *Comput. Biol. Med.*, vol. 147, p. 105620, ago. 2022, doi: 10.1016/j.combiomed.2022.105620.

[22] «Magician's Corner: 9. Performance Metrics for Machine Learning Models». Accedido: 21 de junio de 2025. [En línea]. Disponible en: <https://pubs.rsna.org/doi/epdf/10.1148/ryai.2021200126>