



**Implementación de un modelo de segmentación automática de órganos abdominales  
mediante técnicas de Deep Learning en imágenes de Tomografía Computarizada  
contrastada**

Jhoan Sebastian Mora Colorado

Trabajo de grado presentado para optar al título de Bioingeniero

Asesores

Yesika Alexandra Agudelo, Bioingeniera, Estudiante de maestría en neuroingeniería

Isabella Ariza Cuberos, Bioingeniera, Estudiante de maestría en neuroingeniería

Universidad de Antioquia  
Facultad de Ingeniería  
Bioingeniería  
Medellín, Antioquia, Colombia  
2024

| Cita                                    | Mora Colorado [1]  |
|---|--|
| <b>Referencia</b><br>Estilo IEEE (2020) | [1] J.S Mora Colorado, "Implementación de un modelo de segmentación automática de órganos abdominales mediante técnicas de Deep Learning en imágenes de Tomografía Computarizada contrastada", Trabajo de grado profesional, Bioingeniería, Universidad de Antioquia, Medellín, Antioquia, Colombia, 2024. |



Centro de Documentación de la Facultad de Ingeniería (CENDOI)

**Repositorio Institucional:** <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - [www.udea.edu.co](http://www.udea.edu.co)

Rector: John Jairo Arboleda Céspedes.

Decano/Director: Julio Cesar Saldarriaga.

Jefe departamento: John Fredy Ochoa Gomez.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

## **Dedicatoria**

Dedico este proyecto a mi familia, cuyo amor y apoyo han sido mi mayor soporte durante toda mi carrera.

A mis amigos, cuyo apoyo y compañía han sido fundamentales para mantenerme motivado y enfocado. Sus palabras de ánimo y su confianza en mí han sido una fuente de fuerza constante.

Agradezco también a todos aquellos que, de alguna manera, contribuyeron a este logro, ya sea con su amistad, consejos, o simplemente estando presentes en los momentos más importante.

## **Agradecimientos**

Quiero expresar mi mayor agradecimiento a mi familia, especialmente a mis padres y hermanos, por su incondicional apoyo y aliento durante toda mi carrera académica.

Mi gratitud más también va dirigida a mis líderes Yesika Alexandra Londoño, Jon Edinson Duque, Isabella Ariza Cuberos y el Dr. Jorge Mario Vélez. Su apoyo constante, colaboración, guía y paciencia durante el desarrollo de este proyecto han sido indispensables.

A mis amigos aprendices y demás líderes, quienes estuvieron a mi lado durante este proceso, les agradezco por su compañía, consejo y apoyo. Sus palabras de aliento y motivación fueron importantes para seguir adelante.

A la Universidad de Antioquia por brindarme un entorno académico de calidad que ha sido fundamental en mi formación profesional.

Agradezco a SURA por confiar en mis capacidades y permitirme colaborar con el equipo de dirección en imágenes e inteligencia artificial en salud. Estoy muy agradecido por haberme brindado la oportunidad de trabajar con ustedes.

## TABLA DE CONTENIDO

### Contenido

|   |    |
|---|----|
| RESUMEN   | 9  |
| ABSTRACT  | 10 |
| I. INTRODUCCIÓN   | 11 |
| II. PLANTEAMIENTO DEL PROBLEMA  | 13 |
| III. JUSTIFICACIÓN  | 15 |
| IV. OBJETIVOS   | 17 |
| A. Objetivo general   | 17 |
| B. Objetivos específicos  | 17 |
| V. MARCO TEÓRICO  | 18 |
| VI. METODOLOGÍA   | 20 |
| A. Selección de la muestra para entrenamiento con imágenes de CT                    | 20 |
| B. Entrenamiento del modelo   | 22 |
| C. Validación cuantitativa de los modelos con segmentaciones clínicamente validadas | 24 |
| VII. RESULTADOS   | 25 |
| VIII. DISCUSIÓN   | 34 |
| IX. CONCLUSIONES  | 36 |
| X. RECOMENDACIONES  | 36 |
| REFERENCIAS   | 38 |

## LISTA DE TABLAS

|  |    |
|--|----|
| Tabla 1 Configuración de diferentes modelos de segmentación entrenados | 24 |
| Tabla 2 Distribución de sujetos por fabricante del equipo y género     | 27 |
| Tabla 3 Resultados de validación de modelos de segmentación            | 30 |
| Tabla 4 Métricas de rendimiento por órgano del modelo 142 SURA 8       | 31 |
| Tabla 5 Métricas de clasificación por órgano del modelo 142 SURA 8     | 31 |

## LISTA DE FIGURAS

|  |    |
|--|----|
| Figura 1 Distribución de frecuencia de sujetos por edad y sexo   | 27 |
| Figura 2 Distribución de volúmenes de los órganos abdominales según el equipo de adquisición   | 28 |
| Figura 3 Relación entre IMC y volumen del hígado   | 28 |
| Figura 4 Relación entre edad y volumen del hígado  | 29 |
| Figura 5 Muestra de 30 sujetos seleccionada a partir de criterios de selección   | 29 |
| Figura 6 Sujeto desde el visor OHIF mostrando las segmentaciones convertidas en formato DICOM-SEG  | 30 |
| Figura 7 Matriz de confusión del modelo UNETR con 8 etiquetas  | 32 |
| Figura 8 Pérdida promedio por iteración para cada configuración evaluada   | 33 |
| Figura 9 Evaluación cuantitativa de la métrica de DICE y Hausdorff en sujetos de prueba  | 34 |
| Figura 10 Comparación cuantitativa de las segmentaciones generadas por los modelos model 8, 30 model 13, y las segmentaciones de referencia (ground truth) | 35 |

## LISTA DE ECUACIONES

|   |    |
|---|----|
| Ecuación 1 Coeficiente de Dice                | 21 |
| Ecuación 2 Intersection over Union            | 21 |
| Ecuación 3 Distancia de Hausdorff             | 21 |
| Ecuación 4 Average Symmetric Surface Distance | 21 |

## **SIGLAS, ACRÓNIMOS Y ABREVIATURAS**

|              |                                    |
|--------------|------------------------------------|
| <b>TC</b>    | Tomografía Computarizada           |
| <b>IA</b>    | Inteligencia Artificial            |
| <b>IM</b>    | Imágenes médicas                   |
| <b>DL</b>    | Deep Learning                      |
| <b>UdeA</b>  | Universidad de Antioquia           |
| <b>MONAI</b> | Medical Open Network for AI        |
| <b>CNN</b>   | Convolutional Neural Network       |
| <b>ASSD</b>  | Average Symmetric Surface Distance |
| <b>DSC</b>   | Dice Similarity Coefficient        |
| <b>HD</b>    | Hausdorff Distance                 |
| <b>IoU</b>   | Intersection over Union            |
| <b>SemS</b>  | Segmentación semántica             |
| <b>UNETR</b> | U-shaped Network with Transformers |
| <b>BTCV</b>  | Beyond The Cranial Vault           |

---

## RESUMEN

La segmentación automática de órganos abdominales ha representado un desafío significativo en la radiología, especialmente en aplicaciones clínicas donde se requiere precisión y eficiencia. A lo largo de la última década, las redes neuronales completamente convolucionales (FCNNs) han demostrado ser una solución eficaz para la mayoría de las tareas de segmentación de imágenes médicas. Sin embargo, presenta limitaciones para capturar dependencias espaciales a largo alcance. En este trabajo, se aborda la segmentación de órganos abdominales utilizando el framework MONAI, que facilita la implementación de flujos de trabajo de deep learning en imágenes médicas. Se emplea la arquitectura UNETR (U-Net con Transformers), que utiliza transformers para captar información global a diversas escalas. Para mejorar el rendimiento del modelo, se aplicaron técnicas avanzadas de regularización y optimización, aprovechando las herramientas proporcionadas por MONAI. El modelo fue validado con la base de datos de la población SURA, lo que permite evaluar su desempeño en un contexto colombiano, abordando el sesgo presente en muchos estudios que utilizan datos europeos. Los resultados superaron los modelos de referencia, con un Coeficiente de Dice de 0.94365 y una pérdida de 0.5684, destacándose en la segmentación de órganos complejos como el páncreas y las venas esplénica y porta. Este trabajo contribuye a la generalización de modelos de segmentación en datos locales, estableciendo un nuevo estándar en la región.

***Palabras clave*** — Imágenes médicas, TC, Segmentación, Deep learning, MONAI.

---

## ABSTRACT

The automatic segmentation of abdominal organs remains a significant challenge in radiology, particularly in clinical applications where both accuracy and efficiency are essential. Over the past decade, Fully Convolutional Neural Networks (FCNNs) have proven to be effective for most medical image segmentation tasks. However, they present limitations in capturing long-range spatial dependencies. In this work, we address the segmentation of abdominal organs using the MONAI framework, which facilitates the implementation of deep learning workflows in medical imaging. The architecture employed is UNETR (U-Net with Transformers), which leverages transformers to capture global information at multiple scales. To enhance the model's performance, advanced regularization and optimization techniques were applied, utilizing MONAI's powerful tools. The model was validated using the SURAB population database, allowing for an assessment of its performance in a Colombian context and addressing the bias found in many studies relying on European data. The results outperformed baseline models, achieving a Dice Coefficient of 0.94365 and a loss of 0.5684, excelling in the segmentation of complex organs such as the pancreas and the splenic and portal veins. This work contributes to the generalization of segmentation models on local data, establishing a new standard in the region.

**Keywords** — Medical imaging, TC, Segmentation, Deep Learning, MONAI.

---

## I. INTRODUCCIÓN

La segmentación de órganos abdominales a partir de imágenes de tomografía computarizada es una tarea crucial en el ámbito de la imagen médica, que se utiliza en múltiples aplicaciones clínicas de gran relevancia. Este proceso no solo facilita al personal médico en el diagnóstico preciso de enfermedades, sino que también brinda apoyo considerable a los cirujanos en la planificación terapéutica y a los investigadores en el análisis cuantitativo para estudios poblacionales [1].

La segmentación precisa de órganos a partir de imágenes médicas, particularmente en el contexto de TC abdominales, representa un desafío significativo en la práctica clínica contemporánea. Aunque la segmentación manual de radiólogos o con herramientas semiautomáticas ha sido tradicional, su adopción se ve restringida por la exigente demanda de tiempo y recursos humanos [2].

Los avances recientes en el aprendizaje profundo, específicamente el uso de redes neuronales convolucionales (CNN), han cambiado de paradigma en la segmentación de órganos, ofreciendo soluciones automatizadas que mejoren la eficiencia y la precisión del proceso. Sin embargo, la aplicación efectiva de modelos basados en CNN en la segmentación de imágenes médicas abdominales presenta dificultades que requieren un enfoque cuidadoso y específico [3].

Un aspecto fundamental que considerar es el equilibrio entre los recursos computacionales disponibles y la precisión del modelo. Los enfoques basados en CNN pueden clasificarse en 2D y 3D, cada uno con sus ventajas y limitaciones inherentes. Por ejemplo, los enfoques 2D, que utilizan cortes de imágenes separadas para el entrenamiento, ofrecen una ventaja en términos de velocidad computacional, pero pueden perder información espacial crucial [4].

Por otro lado, los enfoques 3D, que operan directamente en volúmenes tridimensionales, representan más la anatomía abdominal, pero a costa de una mayor demanda computacional y la necesidad de resolver problemas relacionados con la memoria [5].

En el contexto de la segmentación abdominal, se presentan dificultades adicionales, como la falta de un método de registro bien establecido para el espacio estándar debido a las mayores variaciones en los tejidos blandos entre sujetos. Esto hace que el recorte de imágenes conduzca a una pérdida de conocimientos relevantes para la segmentación. Además, las estrategias convencionales de reducción de resolución introducen operaciones de interpolación que pueden

---

comprometer la integridad de las estructuras biológicas en las imágenes médicas, lo que resulta en una pérdida de información relevante para la segmentación abdominal.

En este contexto de desafíos en la segmentación de imágenes médicas, MONAI se ha consolidado como una herramienta que supera muchas de las limitaciones existentes. Este framework de código abierto, especializado en imágenes médicas, ofrece soluciones innovadoras para optimizar la segmentación de órganos abdominales. MONAI destaca por su capacidad para implementar arquitecturas híbridas como ResUNet y UNETR, entre otras, logrando un rendimiento notable en tareas de segmentación. Además, facilita el desarrollo y la optimización de modelos, eliminando la necesidad de scripts manuales y permitiendo una implementación más rápida y eficiente [5].

Estudios recientes han demostrado la eficacia de MONAI en estos desafíos. Diversas arquitecturas de MONAI se han implementado para la segmentación de órganos abdominales en imágenes pediátricas de TC, logrando un rendimiento superior en comparación con métodos convencionales [6]. Los modelos de MONAI no solo demostraron alta precisión y eficiencia computacional, procesando imágenes en apenas 2.9 segundos, sino que también exhibieron una notable capacidad de adaptación a través del aprendizaje por transferencia. Esto permite abordar eficazmente la variabilidad anatómica entre sujetos y preservar la información espacial en la segmentación abdominal [6].

El objetivo principal de este estudio fue diseñar un modelo de segmentación automática de órganos abdominales, adaptado para imágenes de TC contrastadas de población adulta SURA. Para abordar las limitaciones del modelo actual, se propuso un enfoque totalmente automatizado utilizando una red neuronal convolucional profunda.

Este enfoque no solo permite la segmentación de órganos abdominales, sino que también ofrece beneficios adicionales, como la mejora en la lectura de imágenes, una mayor precisión en la identificación de estructuras anatómicas, y una capacidad de adaptación a las variaciones individuales en la anatomía abdominal. Además, la automatización tiene el potencial de reducir significativamente el tiempo requerido para el análisis de imágenes, lo que podría aumentar la eficiencia y la productividad en entornos clínicos y de investigación.

Para el desarrollo del modelo, se utilizó el framework MONAI junto con técnicas de Deep Learning, aplicadas a la población adulta SURA. Este enfoque no solo permite una mejor

---

generalización a la población colombiana, sino que también proporciona una herramienta más adecuada para el contexto local

---

## II. PLANTEAMIENTO DEL PROBLEMA

El uso de modelos normativos en el análisis de imágenes médicas ha demostrado ser una herramienta eficaz para capturar la variabilidad interindividual en áreas como la neuroimagen, donde ha facilitado la identificación de biomarcadores relevantes para trastornos psiquiátricos. Sin embargo, aunque estas técnicas han sido implementadas con éxito en otros contextos, su aplicación en el análisis de imágenes abdominales, particularmente en la población colombiana, sigue siendo un área poco explorada [7]. Dado el creciente desarrollo de tecnologías de imagen no invasivas, se presentan nuevas oportunidades para estudiar estructuras anatómicas abdominales y sus variaciones.

La segmentación automática es fundamental para el análisis de imágenes abdominales, con aplicaciones que abarcan el diagnóstico asistido por computadora, la planificación quirúrgica y la extracción de índices cuantitativos. La delineación de órganos viscerales sólidos, como el hígado, los riñones y el bazo, es crucial para la localización, evaluación volumétrica y seguimiento. Sin embargo, el análisis de imágenes abdominales mediante tomografía computarizada (TC) y resonancia magnética (RM) es un proceso complejo y que demanda mucho tiempo al personal clínico, debido a la diversidad anatómica. Por ello, contar con una segmentación automática no solo facilita la interpretación de imágenes, sino que también optimiza la toma de decisiones clínicas y mejora la atención al paciente, evitando los esfuerzos manuales de delineación.

A lo largo de los años, se han desarrollado múltiples métodos interactivos, semiautomáticos y automáticos para la segmentación, incluidas técnicas basadas en modelos estadísticos y aprendizaje automático. Más recientemente, el aprendizaje profundo ha logrado avances significativos en la segmentación de órganos abdominales, especialmente en su capacidad para adaptarse a la variabilidad en forma, tamaño y textura de los órganos. La segmentación del tejido adiposo visceral, por ejemplo, es particularmente desafiante debido a la gran variabilidad anatómica y la presencia de gas en los contenidos intestinales. En pacientes con baja adiposidad, los métodos convencionales, como el umbral simple, tienden a fallar, mientras que los enfoques basados en aprendizaje profundo han demostrado ser tan precisos como la segmentación manual, superando incluso la concordancia entre observadores [8].

---

El uso de redes completamente convolucionales (FCN) ha sido clave en este avance, permitiendo la extracción de características a múltiples escalas y la recuperación de la resolución original a través de convoluciones ascendentes. En la comunidad de procesamiento de imágenes médicas, la arquitectura UNET-R es ampliamente reconocida por su capacidad para aprender de conjuntos de datos relativamente pequeños, lo que le permite capturar conocimiento de alto nivel de manera similar a los radiólogos.

A pesar de estos avances, la evaluación de la efectividad de las redes neuronales en la segmentación de múltiples órganos sigue siendo limitada, ya que la mayoría de los estudios se enfocan en órganos específicos, utilizan una sola modalidad (como la TC) o se basan en conjuntos de datos pequeños y privados. Además, una limitación importante radica en que muchos de los modelos pre entrenados disponibles han sido desarrollados utilizando poblaciones predominantemente europeas o estadounidenses, lo que introduce sesgos en el análisis de datos cuando se aplican a otras poblaciones, como la colombiana. Esta falta de representatividad limita la generalización de los modelos y subraya la necesidad de desarrollar modelos más inclusivos y representativos de diversas poblaciones.

---

### III. JUSTIFICACIÓN

La segmentación automática de órganos abdominales en imágenes de TC es importante debido a su impacto en diversos aspectos del cuidado de la salud, mejorando la precisión de aplicaciones clínicas como el diagnóstico, la planificación y la entrega de tratamientos. Esta técnica es relevante en el contexto de enfermedades complejas, como el cáncer, donde una delineación precisa de los órganos y tumores puede influir significativamente en las decisiones terapéuticas. La capacidad de segmentar automáticamente los órganos abdominales permite un análisis detallado y consistente de las imágenes, reduciendo la variabilidad Inter observador y el tiempo requerido para la evaluación.

Además, la segmentación automática proporciona un apoyo adicional en procedimientos quirúrgicos y terapias guiadas por imágenes, tales como la radioterapia y la cirugía mínimamente invasiva. Al ofrecer modelos anatómicos detallados, esta tecnología ayuda a los cirujanos y oncólogos a visualizar las relaciones espaciales entre los órganos y las áreas afectadas, optimizando las estrategias de tratamiento y reduciendo el riesgo de complicaciones. Asimismo, la automatización de la segmentación de imágenes médicas, que tradicionalmente ha sido un proceso laborioso y que consume una cantidad considerable de tiempo y recursos humanos, contribuye significativamente a mejorar la eficiencia y productividad en el entorno clínico.

Esta mejora en la eficiencia permite a los profesionales de la salud centrarse en tareas más relevantes que requieren de su experiencia, lo que resulta valioso en situaciones de emergencia donde el tiempo es un factor crucial. Además de su impacto en la práctica clínica, la segmentación automática tiene un gran potencial en la investigación biomédica y los estudios poblacionales. Al proporcionar datos segmentados de manera precisa y consistente, se facilita la realización de análisis cuantitativos, lo que a su vez posibilita el descubrimiento de patrones epidemiológicos que podrían ser difíciles de identificar sin esta tecnología.

Además, los recientes avances en el aprendizaje profundo y el aumento de la disponibilidad de datos médicos han impulsado el desarrollo de modelos más precisos y eficientes. Utilizando frameworks especializados como MONAI, es posible crear modelos que no solo mejoran la precisión de la segmentación, sino que también permiten la personalización de la medicina, adaptando los tratamientos a las necesidades específicas de cada paciente. En este contexto, los modelos normativos emergen como una herramienta valiosa para establecer patrones de

---

referencia en la anatomía, facilitando la identificación de variaciones patológicas y promoviendo un enfoque más personalizado en la atención médica

---

## IV. OBJETIVOS

### ***A. Objetivo general***

Diseñar un modelo de segmentación automática de múltiples órganos abdominales utilizando técnicas de Deep Learning, adaptado para imágenes de tomografía computarizada (TC) contrastadas basados en población SURA

### ***B. Objetivos específicos***

- Establecer un flujo de trabajo para la conversión y carga de etiquetas médicas de formato NIFTI a DICOM-SEG, garantizando su compatibilidad con el visor OHIF mediante su integración a la plataforma XNAT.
- Ejecutar estrategias de aumento de datos y optimización de hiperparámetros con el fin de mejorar la generalización del modelo UNETR entrenado con el framework MONAI en el contexto de segmentación automática de órganos abdominales en imágenes procedentes tanto de repositorios públicos como propios
- Validar el modelo de segmentación abdominal entrenado con el fin de evaluar su utilidad y fiabilidad en el ámbito clínico

---

## V. MARCO TEÓRICO

El estudio y análisis de imágenes médicas ha sido fundamental en el diagnóstico y tratamiento de enfermedades, permitiendo a los proveedores de atención médica una visión interna del cuerpo humano que de otro modo sería inaccesible. Durante décadas, los radiólogos han sido los intérpretes principales de estas imágenes, utilizando su conocimiento y experiencia para identificar patologías y guiar tratamientos.

Sin embargo, la interpretación visual presenta limitaciones inherentes, ya que es un proceso subjetivo. Además, en ocasiones las imágenes no proporcionan la resolución espacial necesaria debido a restricciones como el tiempo de adquisición, la dosis de radiación o las limitaciones del hardware disponible. Estas limitaciones tecnológicas pueden comprometer la precisión con la que los radiólogos interpretan los detalles tridimensionales de las estructuras internas del cuerpo humano [9].

La necesidad de superar estas barreras llevó al desarrollo de sistemas asistidos por computadora, combinando el procesamiento digital de imágenes con técnicas de inteligencia artificial, como el aprendizaje automático, la lógica difusa y el reconocimiento de patrones. Estas herramientas ofrecen una solución a la subjetividad y a las limitaciones inherentes a la interpretación humana, proporcionando una asistencia importante en la tarea de analizar imágenes médicas [10].

Dentro de este contexto, la TC surgió como una tecnología revolucionaria, ofreciendo una forma de visualizar el interior del cuerpo sin las limitaciones de superposición de estructuras anatómicas que presentaban otras técnicas como la radiografía convencional. La capacidad de la TC para generar imágenes con mayor contraste y detalle ha hecho de esta tecnología un fundamento en la práctica clínica moderna [11].

Sin embargo, la gran cantidad de información que proporcionan las imágenes de TC plantea nuevos desafíos, como la necesidad de identificar y delinear estructuras específicas, tales como órganos, tejidos o lesiones, para obtener información cuantitativa en el diagnóstico y tratamiento. En este contexto, la segmentación de imágenes médicas se ha convertido en un proceso computacional necesario para el análisis de imágenes radiológicas [11].

La segmentación no solo permite el reconocimiento de un órgano o lesión, sino que también delimita su extensión espacial, uniendo así los procesos de reconocimiento de alto nivel y de delimitación de bajo nivel [12].

---

En los últimos años, el aprendizaje profundo ha demostrado ser una herramienta eficaz para abordar los retos en la segmentación de imágenes médicas. Las redes neuronales convolucionales (CNN), en particular, han demostrado un rendimiento sobresaliente en tareas desafiantes como el reconocimiento visual de objetos y la segmentación de órganos. Estas redes, compuestas por múltiples capas de procesamiento, permiten capturar mapeos altamente no lineales entre las entradas y salidas, ofreciendo una alta capacidad para procesar imágenes complejas de manera eficiente [12].

La arquitectura de las CNN se basa en principios como los campos receptivos locales y el submuestreo, lo que garantiza cierto grado de invarianza a cambios de posición, escala y distorsión. Estas características han hecho que las CNN sean especialmente adecuadas para la segmentación de órganos y lesiones en imágenes médicas, capturando información sutil y compleja que puede ser vital para el diagnóstico [13].

Para llevar estas capacidades a un nivel aún más avanzado, se han desarrollado frameworks especializados como MONAI, diseñados con la finalidad de aprovechar al máximo el potencial de las CNN en el ámbito médico. MONAI proporciona herramientas y flujos de trabajo optimizados para entrenar modelos de IA en salud, integrándose perfectamente con PyTorch y facilitando la adopción de estas tecnologías en el ámbito médico y de investigación [14].

Dentro de la gran diversidad de arquitecturas de CNN, el modelo UNETR ha ganado prominencia por su capacidad para segmentar imágenes médicas de manera precisa. UNETR combina la eficiencia de la arquitectura U-Net, conocida por su estructura en forma de U que facilita la segmentación, con el poder de los Transformers, que capturan relaciones a larga distancia en secuencias y modelan estructuras complejas [15,16].

A medida que estas tecnologías avanzan, la necesidad de gestionar y analizar grandes volúmenes de datos de imágenes médicas ha llevado al desarrollo de plataformas como XNAT. Esta herramienta se ha convertido en un estándar en la investigación médica, proporcionando una solución integral para el almacenamiento, análisis y compartición de datos de imágenes médicas. XNAT no solo facilita la colaboración entre investigadores de diversas disciplinas al integrar otros programas y herramientas de análisis, sino que también juega un papel importante en el proceso de entrenamiento de modelos de aprendizaje profundo. Al almacenar grandes volúmenes de imágenes de TC y otros datos relevantes, XNAT proporciona la base de datos para entrenar

---

modelos de segmentación y diagnóstico automatizado, garantizando al mismo tiempo la privacidad y seguridad de la información del paciente [17,18]

Para garantizar que los modelos predictivos desarrollados a partir de estas imágenes sean precisos y generalizables, se emplean técnicas como la validación cruzada k-fold. Este método de evaluación divide el conjunto de datos en varios pliegos, permitiendo entrenar y evaluar el modelo en diferentes subconjuntos. Esto ayuda a identificar problemas como el sobreajuste o subajuste y proporciona una estimación confiable del rendimiento del modelo [19,20].

Es importante destacar que los modelos de segmentación se dividen en diferentes categorías de problemas. En particular, los modelos de segmentación de órganos se consideran dentro del ámbito de la segmentación semántica. Esta categoría de segmentación realiza clasificaciones a nivel de píxel, y para evaluar su rendimiento, se emplean diversas métricas.

Para medir la superposición entre la anotación de referencia y la predicción del algoritmo, se utilizan métricas como el Coeficiente de Dice (DSC) y el IoU. El Coeficiente de Dice se define como:

$$DSC = \frac{2TP}{2TP+FP+FN}$$

*Ecuación 1 Coeficiente de Dice*

donde TP representa el número de verdaderos positivos, FP el número de falsos positivos, y FN el número de falsos negativos. Por su parte, el Intersección sobre Unión (IoU) se calcula mediante:

$$IoU = \frac{TP}{TP+FP+FN}$$

*Ecuación 2 Intersection over Union*

Aunque estas métricas son útiles para evaluar la superposición, no capturan todas las características de las estructuras, especialmente en casos con alta variabilidad en forma o tamaño. Por ello, se complementan con métricas basadas en distancia. La Distancia de Hausdorff (HD), por ejemplo, se define como:

$$HD = \max(\max_{p \in A} \min_{q \in B} \|p - q\|, \max_{q \in B} \min_{p \in A} \|p - q\|)$$

*Ecuación 3 Distancia de Hausdorff*

---

donde A y B son las regiones segmentadas y verdaderas, respectivamente. Además, el Average Symmetric Surface Distance (ASSD) calcula la distancia promedio entre las superficies de las regiones segmentadas y las verdaderas, y se define como:

$$ASSD = \frac{1}{2} \left( \frac{1}{|A|} \sum_{q \in B} \min \|p - q\| + \frac{1}{|B|} \sum_{p \in A} \min \|p - q\| \right)$$

*Ecuación 4 Average Symmetric Surface Distance*

donde |A| y |B| son los números de puntos en las superficies de las regiones segmentadas y verdaderas, respectivamente. Estas métricas basadas en distancia permiten una evaluación más precisa de los contornos y las superficies de las estructuras segmentadas [21].

---

## VI. METODOLOGÍA

### ***A. Selección de la muestra para entrenamiento con imágenes de TC***

Para el entrenamiento del modelo se realizó un estudio de la base de datos descriptivo e inferencial. La población del estudio consistió en imágenes de TC contrastadas de sujetos entre 18 y 80 años, provenientes del proyecto volumétrico de abdomen de la población SURA. La muestra total fue de 2706 sujetos, distribuidos de la siguiente manera:

- 1041 sujetos sanos y validados
- 1163 sujetos sanos aún no validados
- 91 sujetos patológicos, incluyendo:
  - 42 casos de hepatomegalia
  - 35 casos de esplenomegalia
  - 14 casos con otras patologías

Para la selección de la muestra se tuvo en cuenta la cantidad de datos utilizados en el modelo previamente entrenado, que en ese caso fueron 30 sujetos. Estos datos no contenían información sociodemográfica de las imágenes de TC, por lo que la selección de nuestra muestra se realizó considerando la diversidad volumétrica de los órganos abdominales considerados como sanos. Se tuvo en cuenta variables de interés como el equipo de adquisición, edad, sexo, IMC y volúmenes de órganos para el análisis y selección de la muestra.

Se seleccionó una muestra homogénea de 30 sujetos, divididos equitativamente en 15 hombres y 15 mujeres, todos ellos sanos y validados por radiólogos especializados. Los sujetos se encontraban en un rango de edad entre 18 a 80 años. La selección fue dirigida a obtener una diversidad en los volúmenes de órganos abdominales, específicamente bazo, riñón derecho, riñón izquierdo, hígado, aorta, vena cava inferior, venas porta y esplénica, y páncreas. Se puso especial énfasis en garantizar que los volúmenes de los órganos seleccionados se encontrarán dentro de los rangos considerados clínicamente normales.

Con la muestra seleccionada, se procedió a descargar las imágenes TC de los 30 sujetos seleccionados desde la plataforma XNAT. Se implementó un script bash para automatizar la

---

conversión de las etiquetas del formato NIFTI a DICOM-SEG, necesaria para asegurar la compatibilidad con el visor OHIF integrado en XNAT. Las imágenes convertidas se cargaron nuevamente en XNAT, permitiendo su visualización y edición a través del visor OHIF.

Se realizó la capacitación a un equipo de radiólogos en el uso del visor OHIF para revisar y, de ser necesario, corregir las segmentaciones directamente en la plataforma XNAT. Con el flujo de trabajo implementado se esperaba obtener un conjunto de datos de alta calidad y validados por expertos, necesarios para el entrenamiento del modelo de segmentación automática de órganos abdominales.

### ***B. Entrenamiento del modelo***

Para la implementación del modelo de segmentación, se utilizó el framework MONAI, optimizado para procesar imágenes de TC con contraste. El conjunto de datos fue dividido en subconjuntos de entrenamiento, validación y prueba, y se siguieron las siguientes etapas en el preprocesamiento de las imágenes y la optimización del modelo:

- **Normalización de la intensidad:** Se aplicó la transformación ‘ScaleIntensityRanged’ para estandarizar el rango de valores de las imágenes, asegurando la homogeneidad en la intensidad.
- **Remuestreo:** Las imágenes fueron muestreadas a un tamaño de vóxel de 1.5 mm x 1.5 mm x 2.0 mm, con el fin de mantener un espaciamiento y tamaño equitativo en las imágenes y sus etiquetas.
- **Aumento de datos:** Durante el entrenamiento, se implementaron diversas técnicas de aumento de datos, incluyendo ‘Orientationd’, ‘Spacingd’, ‘CropForegroundd’, ‘RandCropByPosNegLabeld’, ‘RandFlipd’, ‘RandRotate90d’ y ‘RandShiftIntensityd’, con el objetivo de incrementar la variabilidad del conjunto de datos

### **Optimización del modelo:**

- **Función de pérdida:** Se empleó una combinación de las funciones de pérdida Dice y entropía cruzada (DiceCELoss), que permite optimizar la superposición entre las segmentaciones predichas y las etiquetas de referencia, mejorando la precisión global.

- 
- **Validación:** Para monitorear y ajustar continuamente el rendimiento del modelo, se llevó a cabo una fase de validación cada 500 iteraciones, permitiendo ajustes en los parámetros de acuerdo con los resultados observados.

La arquitectura seleccionada para el modelo fue UNETR (UNet Transformer), basada en estudios previos que demostraron una buena efectividad en la segmentación semántica de imágenes médicas tridimensionales[20]. Esta arquitectura se eligió por su capacidad para capturar las relaciones entre los volúmenes de entrada y mejorar la precisión en la segmentación.

Se decidió implementar los hiper parámetros principales sugeridos en el estudio de referencia para el entrenamiento de los diferentes modelos, los cuales fueron:

- `feature_size = 16`
- `hidden_size = 768`
- `mlp_dim = 3072`
- `num_heads = 12`
- `dropout = 0`

A lo largo del proceso de entrenamiento, se evaluaron diferentes configuraciones del modelo para optimizar tanto su rendimiento como su capacidad de generalización. A continuación, se presenta un resumen de las principales configuraciones y técnicas utilizadas en el entrenamiento de los modelos:

*Tabla 1 Configuración de diferentes modelos de segmentación entrenados*

| Nombre del modelo | Numero de sujetos | Fuente de datos              | Número de órganos | Optimizador             | Técnicas de regularización | Iteraciones |
|-------------------|-------------------|------------------------------|-------------------|-------------------------|----------------------------|-------------|
| 30 model repo 13  | 30                | Repositorios externos (BTCV) | 13                | AdamW, LR=1e-4, WD=1e-5 | Ninguna                    | 25,000      |

|                          |     |                                     |    |                             |         |        |
|--------------------------|-----|-------------------------------------|----|-----------------------------|---------|--------|
| 30 model repo 10         | 30  | Repositorios externos (BTCV)        | 10 | AdamW, LR=1e-4, WD=1e-5     | Ninguna | 25,000 |
| 30 model SURA 8          | 30  | Población SURA                      | 8  | AdamW, LR=1e-4, WD=1e-5     | Steplr  | 25,000 |
| 30 model 8 with LRFinder | 30  | SURA y Repositorios externos (BTCV) | 8  | LR Finder (LR: 1e-4 a 1e-3) | Ninguna | 25,000 |
| 60 model SURA REPO 8     | 60  | SURA y Repositorios externos (BTCV) | 8  | AdamW, LR=1e-4, WD=1e-5     | Ninguna | 40,000 |
| 142 model SURA 8         | 142 | Población SURA                      | 8  | AdamW, LR=1e-4, WD=1e-5     | Ninguna | 35,000 |

Para evaluar la precisión y efectividad de las segmentaciones generadas por el modelo, se emplearon las siguientes métricas

**Evaluación del Modelo:**

- **Métricas de Segmentación:**
  - Coeficiente de Similitud de Dice (DSC)
  - Intersección sobre Unión (IoU)
  - Distancia Simétrica Promedio de Superficie (ASSD)
  - Distancia de Hausdorff (HD)
- **Métricas de Clasificación:**
  - Precisión
  - Recall

- 
- o F1-score
  - o Support

Estas métricas cuantifican la superposición entre las segmentaciones predichas y las etiquetas de referencia además de que evalúa la precisión de las segmentaciones al medir la máxima distancia entre los puntos de superficie de referencia y los puntos de superficie de la predicción.

Como complemento se generaron gráficos para visualizar la evolución de la pérdida y del coeficiente Dice a lo largo de las iteraciones de entrenamiento, así como un gráfico que muestra el coeficiente Dice para cada etiqueta, permitiendo evaluar la precisión de la segmentación por órgano. Además, se creó una matriz de confusión y un gráfico de radar para representar simultáneamente varias métricas de evaluación.

### ***C. Validación cuantitativa de los modelos con segmentaciones clínicamente validadas***

Tras completar el entrenamiento de los modelos, se llevó a cabo una validación clínica para evaluar su desempeño en comparación con segmentaciones manuales validadas por radiólogos. Este proceso de validación se realizó utilizando imágenes de los sujetos seleccionados para pruebas, cuyas segmentaciones manuales sirvieron como referencia.

Se evaluó la precisión y la generalización de las segmentaciones automáticas utilizando las métricas previamente descritas, además de que se crearon gráficos que visualizan el rendimiento de los modelos, permitiendo identificar áreas potenciales de mejora y ajustar los parámetros de entrenamiento si es necesario.

Finalmente, se compararon las segmentaciones obtenidas con las generadas por un modelo pre entrenado en el conjunto de datos BTCV, lo que permitió validar la generalización y la adaptabilidad del modelo entrenado con la población SURA.

## VII. RESULTADOS

A partir del análisis de un total de 2706 sujetos, con edades comprendidas entre 18 y 80 años, se identificaron 1041 sujetos sanos y validados, distribuidos en 779 mujeres y 262 hombres. Para equilibrar la muestra por sexo y asegurar la variabilidad volumétrica, se seleccionó un subconjunto homogéneo de 374 sujetos, con una distribución equitativa entre hombres y mujeres. Esta selección inicial se basó en un filtrado de la calidad de las imágenes disponibles y las segmentaciones realizadas.

Posteriormente, se redujo la muestra a 30 sujetos para la conversión de etiquetas de NIfTI a DICOM-SEG. La selección final de estos 30 sujetos fue basada en la distribución de volúmenes abdominales, asegurando la diversidad necesaria para las pruebas de segmentación.

En la figura 1 se observa la distribución de sujetos por edades.

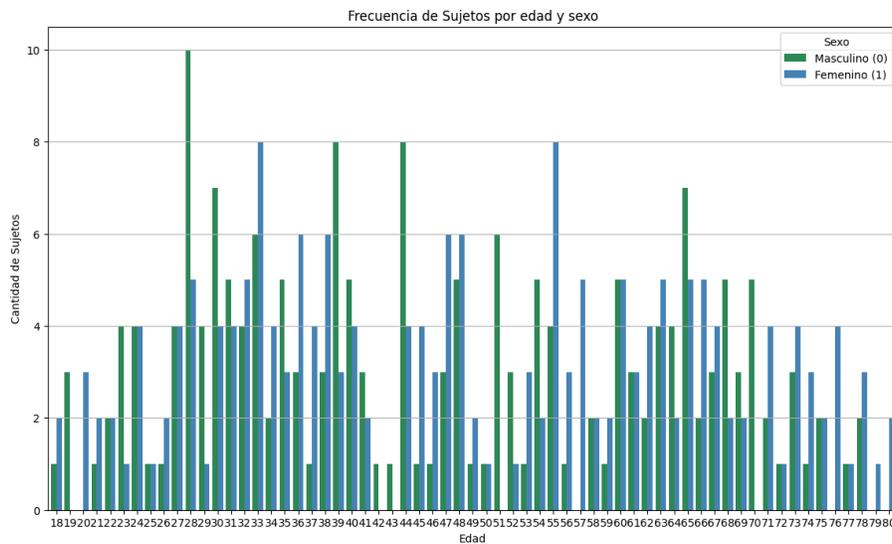


Figura 1 Distribución de frecuencia de sujetos por edad y sexo

Tras analizar la distribución de edades y género en la muestra seleccionada, se procedió a investigar la posible influencia del equipo de adquisición de imágenes en los volúmenes de los órganos abdominales. Los sujetos fueron escaneados utilizando equipos de dos fabricantes principales: GE Medical Systems y Siemens. La distribución de los sujetos según el fabricante del equipo y su género se muestra en la Tabla 2.

Tabla 2 Distribución de sujetos por fabricante del equipo y género

| Fabricante         | Hombres | Mujeres |
|--------------------|---------|---------|
| GE Medical Systems | 151     | 143     |
| Siemens            | 36      | 51      |

Se evaluaron los volúmenes de los ocho órganos abdominales: bazo, riñón derecho, riñón izquierdo, hígado, aorta, vena cava inferior, venas porta y esplénica, y páncreas. Los resultados del análisis no mostraron diferencias significativas en los volúmenes de los órganos entre los sujetos escaneados con equipos de GE Medical Systems y Siemens. A continuación, se presenta el gráfico comparativos para el hígado, donde se observa la distribución de volumen según el equipo de adquisición

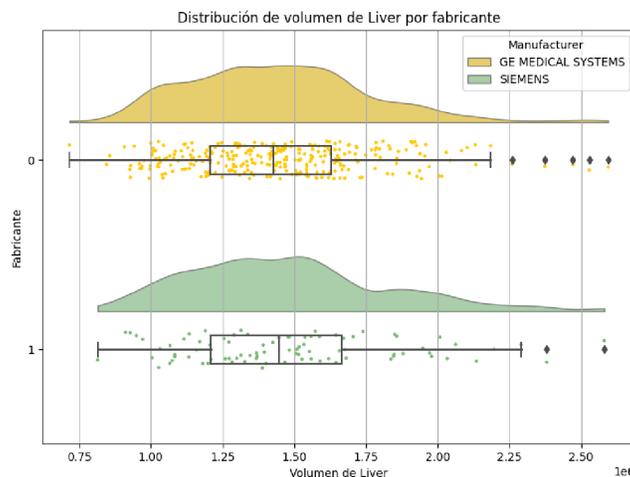


Figura 2 Distribución de volúmenes de los órganos abdominales según el equipo de adquisición

Se realizó un análisis adicional para investigar la relación entre el índice de masa corporal (IMC) y el volumen de los órganos. El análisis reveló que, a mayor IMC, mayor volumen de los órganos. Este comportamiento se ejemplifica con el gráfico del hígado (Figura 3), donde se observa una correlación directa entre IMC y volumen del órgano

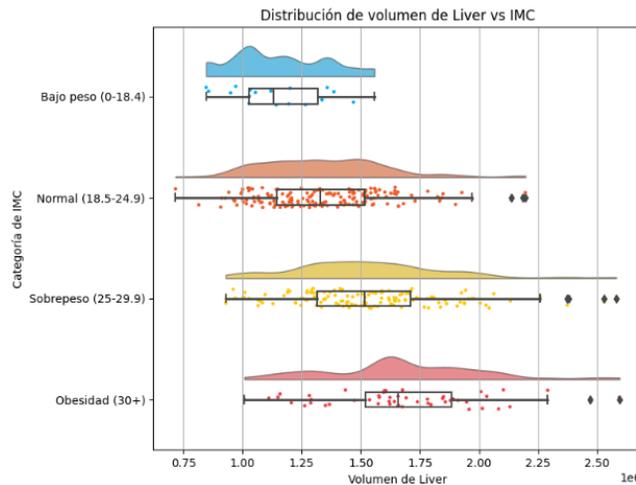


Figura 3 Relación entre IMC y volumen del hígado

Por último, para obtener una comprensión más completa de los límites volumétricos y las características de los órganos, se evaluó también la influencia de la edad en el volumen de los órganos. Los resultados mostraron una variación en los volúmenes de los órganos con la edad, como se observa en el gráfico

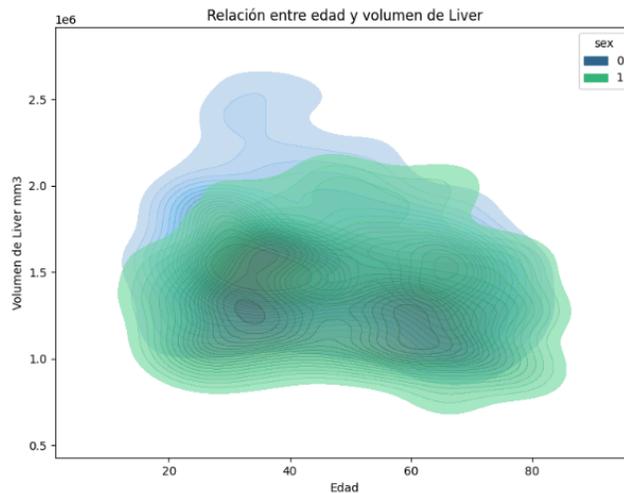


Figura 4 Relación entre edad y volumen del hígado

Con base en los análisis previos, se seleccionó una muestra homogénea de 30 sujetos, buscando la mayor diversidad en los volúmenes de los órganos abdominales. La distribución de los sujetos por grupo de edad se presenta en la figura 5

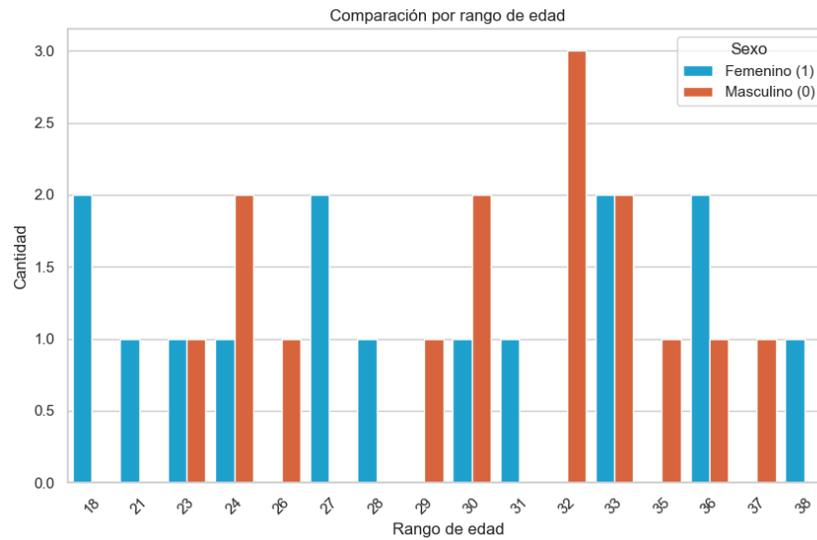


Figura 5 Muestra de 30 sujetos seleccionada a partir de criterios de selección

A cada sujeto seleccionado se le implementó el cambio de formato y la integración al visor de XNAT, a continuación, se ve el resultado del proceso.

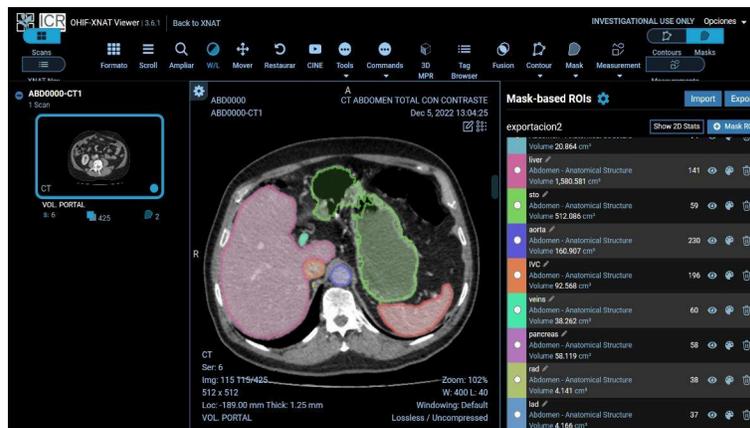


Figura 6 Sujeto desde el visor OHIF mostrando las segmentaciones convertidas en formato DICOM-SEG

Con las segmentaciones validadas desde OHIF, se implementó la etapa de entrenamiento del modelo UNETR. Durante esta fase, se exploraron diversas configuraciones de datos, evaluando el impacto de distintas técnicas de regularización y optimización en el desempeño del modelo. A continuación, se presenta una tabla con los valores de DICE y la función de pérdida obtenidos para cada configuración

*Tabla 3 Resultados de validación de modelos de segmentación*

| <b>Modelo</b>            | <b>DICE Validation</b> | <b>Loss Function</b> |
|--------------------------|------------------------|----------------------|
| 30 model repo 13         | 0.8567                 | 0.8088               |
| 30 model repo 10         | 0.8137                 | 0.6175               |
| 30 model SURA 8          | 0.9136                 | 0.6880               |
| 60 model SURA REPO 8     | 0.8932                 | 0.5881               |
| 30 model 8 with LRFinder | 0.8567                 | 1.6714               |
| 142 model SURA 8         | 0.9437                 | 0.5684               |

El análisis cuantitativo de las segmentaciones generadas durante la etapa de validación por el modelo 142 SURA 8 confirmó su superioridad en comparación con las otras configuraciones evaluadas. Este modelo alcanzó un Coeficiente de Similitud de Dice (DSC) de 0.9436 y una función de pérdida de 0.5684, lo que representa el mejor equilibrio entre precisión de segmentación y minimización del error

Para evaluar el desempeño del modelo en la segmentación de cada órgano individual, se analizó el rendimiento del modelo 142 SURA 8, que obtuvo los mejores resultados en la fase de validación. Las métricas de evaluación se resumen en la Tabla 4

*Tabla 4 Métricas de rendimiento por órgano del modelo 142 SURA 8*

| <b>Órgano</b>           | <b>Coeficiente Dice</b> | <b>Índice Jaccard</b> | <b>Hausdorff (HD)</b> | <b>ASSD</b> |
|-------------------------|-------------------------|-----------------------|-----------------------|-------------|
| Bazo                    | 0.966                   | 0.935                 | 1.0                   | 0.423       |
| Riñón Derecho           | 0.962                   | 0.927                 | 1.0                   | 0.391       |
| Riñón Izquierdo         | 0.964                   | 0.931                 | 1.0                   | 0.395       |
| Páncreas                | 0.857                   | 0.75                  | 2.236                 | 0.789       |
| Hígado                  | 0.978                   | 0.958                 | 1.0                   | 0.434       |
| Aorta                   | 0.875                   | 0.777                 | 1.0                   | 0.492       |
| Vena Cava               | 0.882                   | 0.79                  | 2.236                 | 0.716       |
| Venas porta y esplénica | 0.861                   | 0.756                 | 1.414                 | 0.458       |

El modelo 142 SURA 8 mostró un alto nivel de precisión en la segmentación de órganos como el hígado y el bazo, con Coeficientes Dice de 0.978 y 0.966, respectivamente. Sin embargo, se observó una menor precisión en la segmentación del páncreas y las venas, lo que se reflejó en valores de Dice más bajos y mayores Distancias de Hausdorff.

Además de las métricas mencionadas anteriormente, se calcularon la Precisión, el Recall, el F1-Score y el Support para cada clase. Los resultados obtenidos permitieron entender mejor el comportamiento del modelo en la detección y clasificación de cada órgano segmentado, complementando las métricas de superposición y distancia.

*Tabla 5 Métricas de clasificación por órgano del modelo 142 SURA 8*

| <b>Órgano</b>           | <b>Precisión</b> | <b>Recall</b> | <b>F1-Score</b> | <b>Support</b> |
|-------------------------|------------------|---------------|-----------------|----------------|
| Bazo                    | 0.97             | 0.962         | 0.966           | 41723          |
| Riñón Derecho           | 0.962            | 0.962         | 0.962           | 24907          |
| Riñón Izquierdo         | 0.953            | 0.975         | 0.964           | 26477          |
| Páncreas                | 0.801            | 0.922         | 0.857           | 12518          |
| Hígado                  | 0.981            | 0.976         | 0.978           | 296013         |
| Aorta                   | 0.838            | 0.915         | 0.875           | 4633           |
| Vena Cava               | 0.830            | 0.942         | 0.882           | 13544          |
| Venas porta y esplénica | 0.904            | 0.822         | 0.861           | 4950           |

Los resultados revelan un patrón en el cual el modelo enfrenta dificultades para generalizar en la segmentación de órganos más pequeños, una tendencia que se refleja en el menor número de vóxeles disponibles, como lo indica el 'support'.

Además de las métricas cuantitativas previamente mencionadas, se generó una matriz de confusión para evaluar cómo el modelo 142 SURA 8 distingue entre las diferentes clases de órganos. La matriz de confusión nos permitió evaluar el desempeño del modelo al identificar posibles confusiones entre clases, especialmente en la segmentación de órganos que presentan características similares o que están menos representados en el conjunto de datos de entrenamiento. El resultado se puede observar en la figura 7

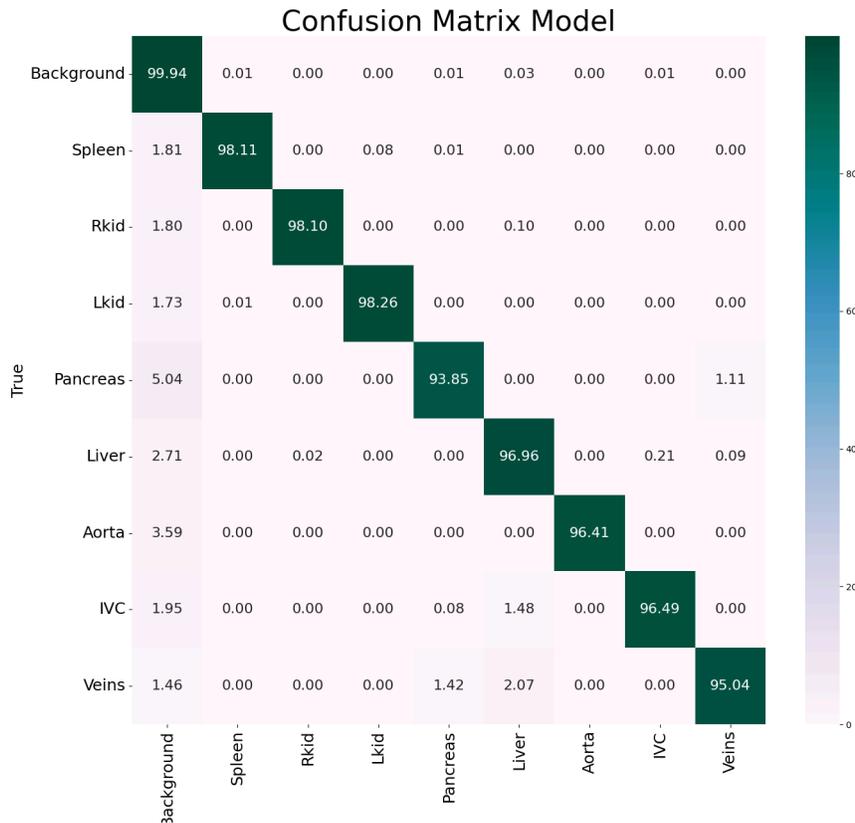


Figura 7 Matriz de confusión del modelo UNETR con 8 etiquetas

La matriz de confusión del modelo seleccionado muestra una alta precisión en la clasificación de la mayoría de las clases. En particular, se observa una segmentación precisa de órganos como el hígado y el bazo, con mínimos errores de clasificación, tal como se mencionó anteriormente. Sin embargo, en la segmentación de órganos más pequeños, como las venas y el páncreas, se identifican errores de clasificación cruzada, donde se confunden partes del fondo y de otros órganos con las estructuras de interés

Tras completar la fase de entrenamiento y evaluación de los modelos con los datos de validación, se procedió a evaluar su desempeño utilizando datos de prueba que no habían sido vistos previamente por los modelos. Para esta etapa, se emplearon datos que previamente habían sido validados por radiólogos, con el objetivo de evaluar la capacidad de generalización de los modelos y su rendimiento en datos que no formaban parte del conjunto de entrenamiento ni del conjunto de validación

Los resultados del entrenamiento de los diversos modelos de segmentación automática de órganos abdominales se presentan en la Figura 8.

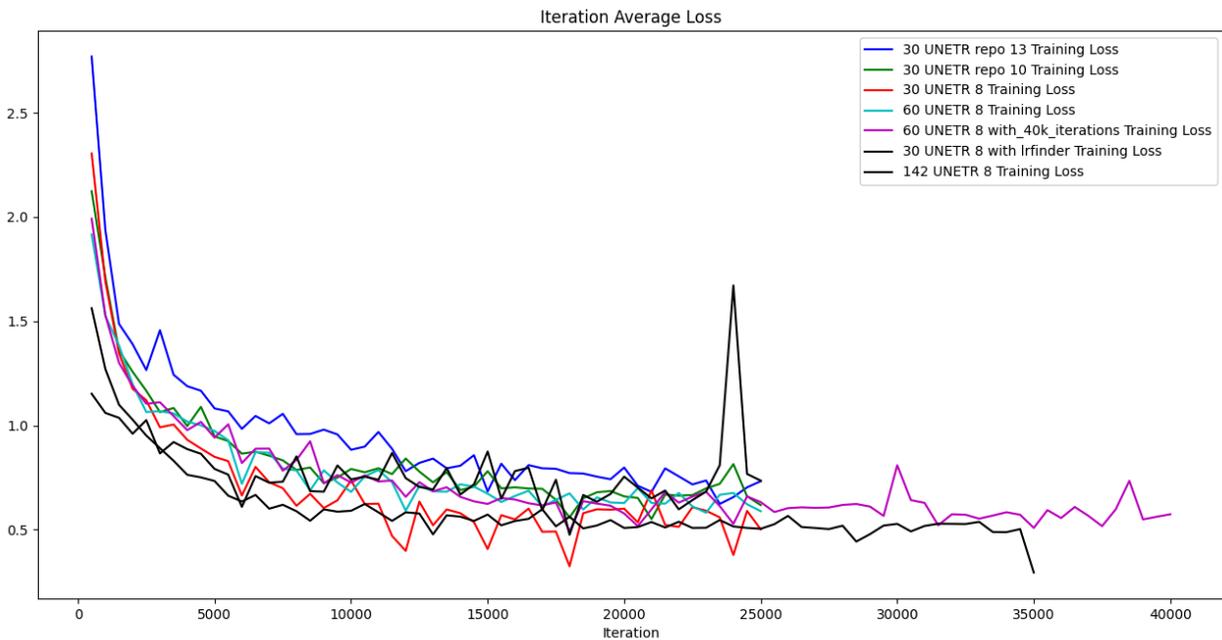


Figura 8 Pérdida promedio por iteración para cada configuración evaluada

Se observó una tendencia general de disminución en la pérdida de entrenamiento en todos los modelos a medida que aumentaba el número de iteraciones, lo cual es indicativo de un aprendizaje efectivo. Sin embargo, se evidencian variaciones en la velocidad de convergencia y estabilidad entre los diferentes modelos, con algunos mostrando fluctuaciones más pronunciadas que otros.

Los modelos con mayor número de épocas y arquitecturas más profundas tendieron a mostrar un mejor rendimiento, con una convergencia más estable y pérdidas más bajas hacia el final del entrenamiento. En particular, el modelo "142 UNETR 8" destacó por su rendimiento superior, indicando que una arquitectura más profunda podría ser más adecuada para la tarea de segmentación de órganos abdominales

Se procedió a evaluar el rendimiento de los modelos seleccionados mediante inferencia en datos de prueba. La Figura 9 presenta los resultados de esta evaluación, mostrando el coeficiente Dice y la distancia de Hausdorff para diferentes órganos abdominales en cuatro modelos: 142 model 8, 30 model 13 (el modelo base), 30 model 8 y 60 model 8

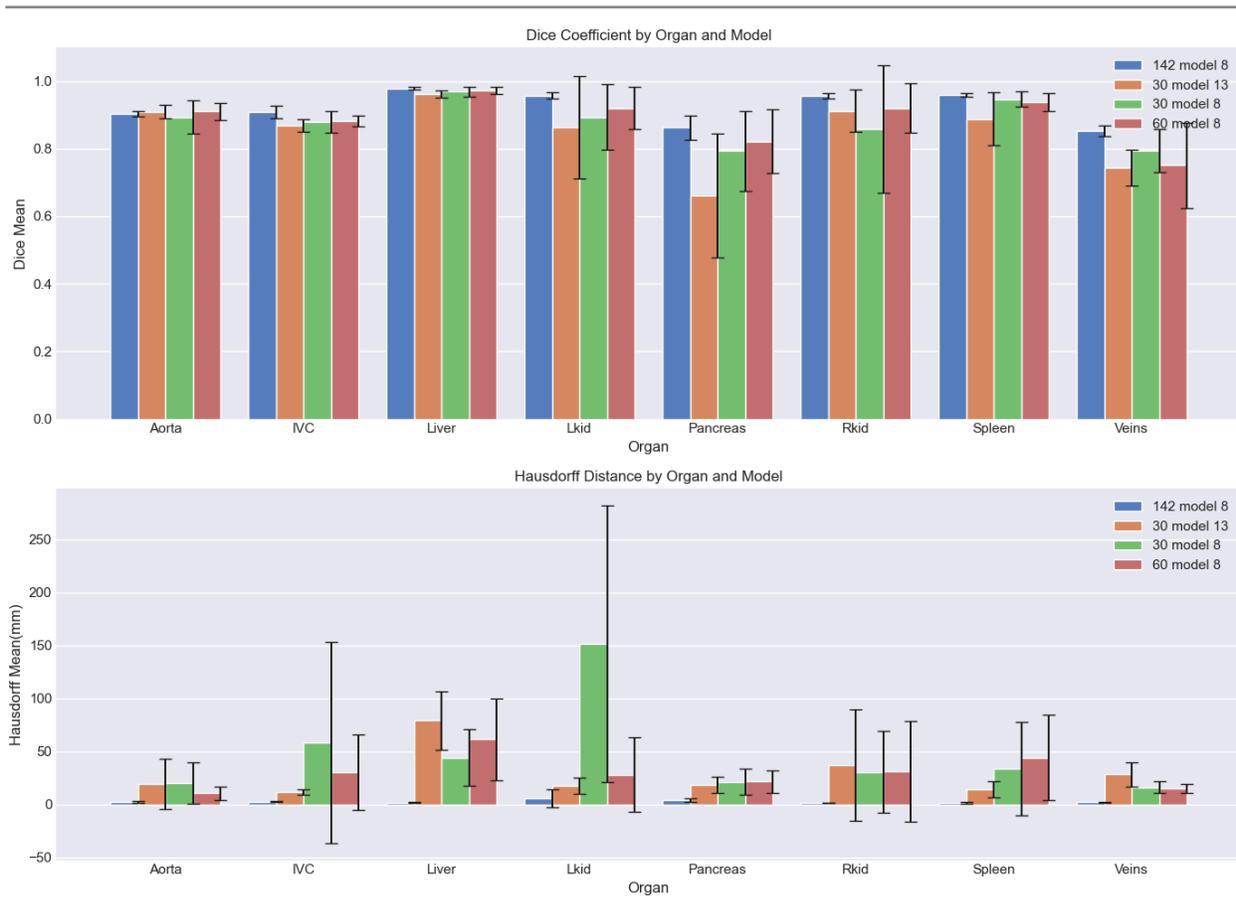


Figura 9 Evaluación cuantitativa de la métrica de DICE y Hausdorff en sujetos de prueba

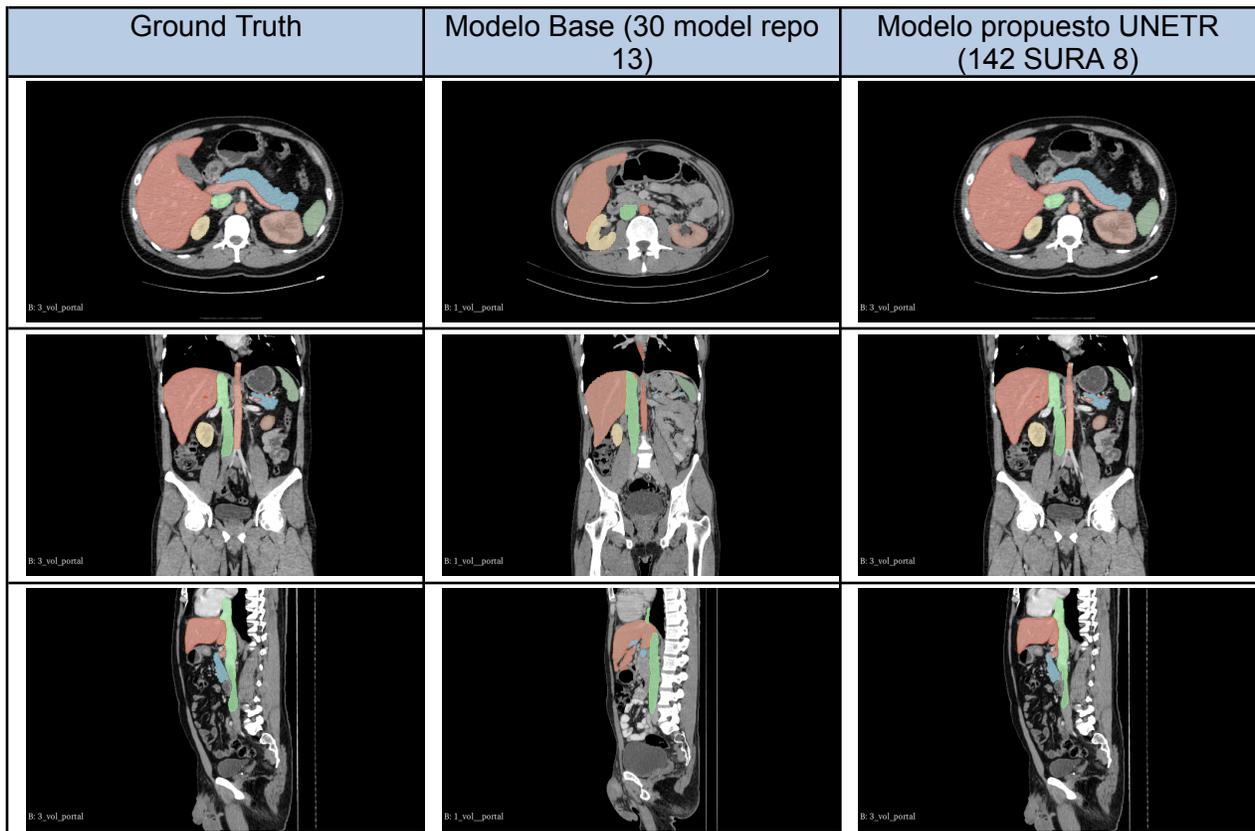
En cuanto al coeficiente Dice, el modelo 142 model 8 destacó por su rendimiento superior en la mayoría de los órganos, con coeficientes Dice consistentemente altos e intervalos de confianza estrechos. La segmentación del hígado fue particularmente notable, con todos los modelos logrando coeficientes Dice cercanos a 1. Sin embargo, el páncreas presentó los coeficientes Dice más bajos y los intervalos de confianza más amplios en todos los modelos, lo que indica que es el órgano más desafiante para la segmentación.

Respecto a la distancia de Hausdorff, el modelo 142 model 8 nuevamente demostró un rendimiento superior, con distancias generalmente más bajas e intervalos de confianza más estrechos. Se observaron variaciones significativas en la distancia de Hausdorff entre órganos y modelos, con intervalos de confianza más amplios en comparación con el coeficiente Dice. El hígado y el bazo mostraron distancias de Hausdorff relativamente bajas en todos los modelos, sugiriendo una segmentación más precisa de estos órganos.

En comparación con el modelo base (30 model 13), el modelo propuesto 142 model 8, mostró mejoras significativas en la precisión de segmentación para la mayoría de los órganos.

Esta mejora se evidencia tanto en los coeficientes Dice más altos como en las distancias de Hausdorff más bajas.

Para complementar las evaluaciones cuantitativas, se realizó un análisis visual de las segmentaciones generadas por los modelos en comparación con las segmentaciones de referencia (ground truth). La Figura 10 muestra ejemplos de las segmentaciones obtenidas por el modelo 142 model 8, el modelo base (30 model 13), y las segmentaciones manuales realizadas por los radiólogos (ground truth), a lo largo de los cortes axial, sagital y coronal para los órganos abdominales



*Figura 10 Comparación cuantitativa de las segmentaciones generadas por los modelos 142 model 8, 30 model 13, y las segmentaciones de referencia (ground truth)*

Esta presentación permite apreciar visualmente la precisión de las segmentaciones en distintos órganos, destacando la capacidad del modelo 142 para delinear estructuras anatómicas complejas con mayor exactitud en comparación con el modelo base

---

## VIII. DISCUSIÓN

Los resultados obtenidos con el modelo 142 SURA 8 muestran un rendimiento superior en la segmentación automática de órganos abdominales en comparación con el modelo de referencia 30 model repo 13, destacándose especialmente en la precisión de la segmentación de órganos como el páncreas y las venas. Este incremento en el rendimiento puede atribuirse a varios factores que se abordarán a continuación.

En primer lugar, la cantidad y calidad de los datos utilizados para entrenar el modelo 142 SURA 8 fueron factores determinantes en la mejora de su desempeño. A diferencia del modelo de referencia, que fue entrenado con un conjunto de datos más reducido y menos variado, el modelo 142 SURA 8 se benefició de un conjunto de datos ampliado que incluyó imágenes de TC de alta calidad y con contraste, lo que mejoró significativamente la capacidad del modelo para generalizar y segmentar con mayor precisión las estructuras anatómicas más complejas. Estudios previos han demostrado que el volumen y la diversidad de datos son determinantes en la capacidad de generalización de los modelos de segmentación automática, destacando que modelos entrenados con conjuntos de datos más completos tienden a tener un mejor rendimiento en escenarios clínicos diversos [22].

Además de la cantidad y calidad de los datos, la arquitectura del modelo jugó un papel crucial en la mejora del rendimiento. El modelo 142 SURA 8 se basa en la arquitectura UNETR, que combina lo mejor de las redes convolucionales con los mecanismos de atención de transformers, permitiendo una mejor captura de las relaciones espaciales a lo largo de las imágenes de TC en 3D. La capacidad de la UNETR para procesar volúmenes completos en lugar de cortes 2D aislados ha demostrado ser efectiva en la segmentación de estructuras anatómicas complejas como el páncreas y las venas, que presentan desafíos particulares debido a su variabilidad en forma y tamaño. La arquitectura UNETR es especialmente adecuada para abordar la complejidad de estas estructuras, ya que los transformers permiten captar contextos globales en las imágenes, mejorando la coherencia espacial y la precisión en la segmentación [23].

Estudios recientes han destacado la efectividad de los transformers en tareas de segmentación 3D, indicando que estos modelos son capaces de aprender representaciones más completas y detalladas en comparación con las arquitecturas convencionales basadas únicamente en convoluciones [24]. En el caso del modelo 142 SURA 8, la implementación de UNETR permitió

---

un mejor manejo de la información contextual, lo que resultó en segmentaciones más precisas, incluso en órganos y estructuras con bajo contraste o bordes menos definidos. Esta ventaja se refleja especialmente en la segmentación del páncreas y las venas, donde la variabilidad entre sujetos y la dificultad en la diferenciación de los límites anatómicos representan desafíos significativos.

Otro aspecto importante a considerar es la estrategia de optimización utilizada en el entrenamiento del modelo. La implementación de diversas técnicas de ajuste de hiperparámetros, como el StepLR y optimizadores adaptativos como AdamW, facilitó la exploración de configuraciones que permitieran una convergencia más rápida y estable del modelo. Estas técnicas optimizaron el proceso de aprendizaje, permitiendo que el modelo alcanzara un punto de equilibrio donde la precisión se maximiza y la pérdida se minimiza. Además, la combinación de estos ajustes con técnicas de regularización, como el early stopping y la validación cruzada k-fold, contribuyó a prevenir el sobreajuste, mejorando la capacidad del modelo para generalizar a nuevos datos no vistos durante el entrenamiento.

Sin embargo, uno de los principales desafíos fue el tiempo necesario para entrenar el modelo, que generalmente oscilaba entre 2 y 3 días, pero que podía extenderse hasta 4 o 5 días si el servidor estaba ocupado con otros procesos. Además, los intentos de entrenar el modelo con arquitecturas adicionales, como SegResNet, o con conjuntos de datos de mayor tamaño, se encontraron con limitaciones de memoria, lo que provocaba que el entrenamiento se detuviera en iteraciones avanzadas. Estos fallos no sólo resultaban en la pérdida de tiempo, ya que tras un día de entrenamiento un fallo implicaba empezar de nuevo, sino que también llevaban a un uso ineficiente de recursos, pues se requería ajustar la configuración del modelo y reiniciar el proceso desde el principio. Esta situación puso de manifiesto la necesidad de desarrollar modelos más optimizados para manejar conjuntos de datos grandes y arquitecturas complejas.

---

## IX. CONCLUSIONES

En este estudio, se ha desarrollado y evaluado un modelo de segmentación automática de órganos abdominales utilizando la arquitectura UNETR. Los resultados obtenidos indican que el modelo final supera al modelo de referencia en la segmentación de órganos, como el páncreas y las venas, mostrando una mayor precisión y coherencia espacial en las segmentaciones

Las principales conclusiones que se pueden extraer de este estudio son las siguientes:

- El modelo 142 SURA 8, gracias a su arquitectura basada en transformers y al uso de un conjunto de datos más amplio y diverso, logró segmentar con mayor precisión estructuras anatómicas complejas, como el páncreas y las venas, en comparación con el modelo de referencia.
- La inclusión de imágenes de TC de alta calidad y con contraste fue determinante para mejorar la capacidad del modelo para generalizar y segmentar estructuras con variabilidad significativa en forma y tamaño.
- La capacidad de UNETR para procesar volúmenes completos de imágenes en 3D y capturar contextos globales ha demostrado ser efectiva en la mejora de la segmentación automática, especialmente en órganos con bordes menos definidos o con bajo contraste.
- El uso de técnicas de ajuste de hiper parámetros, regularización y validación cruzada mejoró la estabilidad y precisión del modelo, previniendo el sobreajuste y optimizando el rendimiento en datos no vistos.
- El tiempo extenso requerido para el entrenamiento del modelo y las limitaciones de memoria fueron desafíos significativos. Además, el proceso de validación fue prolongado debido al tiempo requerido por los radiólogos para corregir las segmentaciones, lo que impidió alcanzar la meta de obtener un mayor número de segmentaciones.

---

## X. RECOMENDACIONES

Con base en los hallazgos de este estudio, se proponen las siguientes recomendaciones para futuras investigaciones y aplicaciones clínicas:

1. Optimización de recursos computacionales: Dado el tiempo requerido para entrenar modelos con arquitecturas complejas y conjuntos de datos grandes, se recomienda explorar opciones para optimizar el uso de recursos computacionales. Esto podría incluir la implementación de técnicas de paralelización, el uso de hardware más potente o la simplificación de modelos para hacerlos más eficientes sin comprometer la precisión.
2. Ampliación del conjunto de datos: Para mejorar aún más la capacidad de generalización del modelo, se sugiere aumentar la muestra de datos, incluyendo imágenes de diferentes poblaciones y escenarios clínicos. Esto no solo contribuiría a la robustez del modelo, sino también a su aplicabilidad en un entorno clínico más amplio.
3. Exploración de nuevas arquitecturas: Aunque UNETR ha demostrado ser efectiva, es importante seguir explorando otras arquitecturas que podrían ofrecer ventajas adicionales, como una mejor eficiencia computacional o mayor precisión en la segmentación de ciertos órganos. Modelos como SegResNet u otras variantes podrían ser evaluados en futuros estudios.
4. Optimización del proceso de validación: Dado que la validación de las segmentaciones fue un proceso lento debido a la disponibilidad y compromiso del personal médico, se recomienda considerar la incorporación de personal dedicado exclusivamente a esta tarea o buscar alternativas para asegurar una revisión más ágil y eficiente. La asignación de un equipo de validación dedicado podría acelerar el proceso y permitir la obtención de un mayor número de segmentaciones válidas.
5. Implementación clínica y validación: Finalmente, para que estos modelos sean adoptados en la práctica clínica, es importante realizar validaciones en escenarios clínicos reales. Esto incluiría pruebas de rendimiento en imágenes de pacientes reales y comparaciones con la segmentación manual realizada por expertos médicos.

---

## REFERENCIAS

- [1] “Semi-Supervised Abdomen Extraction and Organ Segmentation in CT Images.” [Online].  
Available: <https://docs.monai.io/en/stable/transforms.html#histogramnormalize>
- [2] S. K. Jeon, I. Joo, J. Park, J. M. Kim, S. J. Park, and S. H. Yoon, “Fully-automated multi-organ segmentation tool applicable to both non-contrast and post-contrast abdominal CT: deep learning algorithm developed using dual-energy CT images,” *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-55137-y.
- [3] C. S. Perone and J. Cohen-Adad, “Promises and limitations of deep learning for medical image segmentation,” *Journal of Medical Artificial Intelligence*, vol. 2, no. January. AME Publishing Company, Jan. 01, 2019. doi: 10.21037/jmai.2019.01.01.
- [4] Marina Castelló, M. (2023). Universidad San Jorge Escuela de Arquitectura y Tecnología Grado en Ingeniería Informática Proyecto Final Desarrollo de un sistema de reconocimiento de poses 3D en personas utilizando ML.
- [5] S. Muhammad and J. Zhang, “Segmentation of Liver Tumors by Monai and PyTorch in CT Images with Deep Learning Techniques,” *Applied Sciences (Switzerland)*, vol. 14, no. 12, Jun. 2024, doi: 10.3390/app14125144.
- [6] Y. Tang et al., “High-resolution 3D Abdominal Segmentation with Random Patch Network Fusion,” 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1361841520302589>
- [7] Saraiva, D. F. O. (2021). *Normative model for the diagnosis of neuropsychiatric disorders using deep learning methods* (Doctoral dissertation)
- [8] A. D. Weston et al., “Automated Abdominal Segmentation of CT Scans for Body Composition Analysis Using Deep Learning,” *Radiology*, vol. 290, no. 3, pp. 669–679, Mar. 2019, doi: 10.1148/radiol.2018181432.
- [9] E. Somasundaram et al., “Deep-Learning Models for Abdominal CT Organ Segmentation in Children: Development and Validation in Internal and Heterogeneous Public Datasets,” *American Journal of Roentgenology*, May 2024, doi: 10.2214/AJR.24.30931.
- [10] D. D. Patil, S. G. Deore, and S. Bhusawal, “International Journal of Computer Science and Mobile Computing Medical Image Segmentation: A Review,” 2013. [Online]. Available: [www.ijcsmc.com](http://www.ijcsmc.com)
- [11] T. M. Buzug, “Computed To 16. Computed Tomography.”

- 
- [12] A. Mansoor et al., "Segmentation and image analysis of abnormal lungs at CT: Current approaches, challenges, and future trends," *Radiographics*, vol. 35, no. 4, pp. 1056–1076, Jul. 2015, doi: 10.1148/rg.2015140232.
- [13] W. Li, F. Jia, and Q. Hu, "Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks," *Journal of Computer and Communications*, vol. 03, no. 11, pp. 146–151, 2015, doi: 10.4236/jcc.2015.311023.
- [14] M. J. Cardoso et al., "MONAI: An open-source framework for deep learning in healthcare," Nov. 2022, [Online]. Available: <http://arxiv.org/abs/2211.02701>
- [15] Gomes, R., Pham, T., He, N., Kamrowski, C., Wildenberg, J. Analysis of Swin-UNet vision transformer for Inferior Vena Cava filter segmentation from CT scans (2023) *Artificial Intelligence in the Life Sciences*, 4, art. no. 100084.
- [16] Zhou, L., Li, Z. Automatic multi-label temporal bone computed tomography segmentation with deep learning (2023) *International Journal of Medical Robotics and Computer Assisted Surgery*, 19 (5), art. no. e2536.
- [17] Nordio, G., Easmin, R., Giacomel, A., Dipasquale, O., Martins, D., Williams, S., Turkheimer, F., Howes, O., Veronese, M., Jauhar, S., Rogdaki, M., McCutcheon, R., Kaar, S., Vano, L., Rutigliano, G., Angelescu, I., Borgan, F., D'Ambrosio, E., Dahoun, T., Kim, E., Kim, S., Bloomfield, M., Egerton, A., Demjaha, A., Bonoldi, I., Nosarti, C., Maccabe, J., McGuire, P., Matthews, J., Talbot, P.S., and the FDOPA PET imaging working group An automatic analysis framework for FDOPA PET neuroimaging (2023) *Journal of Cerebral Blood Flow and Metabolism*, 43 (8), pp. 1285-1300.
- [18] Doran, S.J., D'Arcy, J., Collins, D.J., Andriantsimiavona, R., Orton, M., Koh, D.-M., Leach, M.O. Informatics in radiology: Development of a research PACS for analysis of functional imaging data in clinical research and clinical trials (2012) *Radiographics*, 32 (7), pp. 2135-2150.
- [19] Pérez-Planells, L., Delegido, J., Rivera-Caicedo, J. P., & Verrelst, J. (2015). Análisis de métodos de validación cruzada para la obtención robusta de parámetros biofísicos. *Revista de teledetección*, (44), 55-65.
- [20] Chacón Chamorro, M. V. Estudio de la reducción del sobreajuste en arquitecturas de redes neuronales residuales ResNet en un escenario de clasificación de patrones (Doctoral dissertation, Universidad Nacional de Colombia).

- 
- [21] L. Maier-Hein *et al.*, “Metrics reloaded: Recommendations for image analysis validation,” Jun. 2022, doi: 10.1038/s41592-023-02151-z.
- [22] Bostan, P. A., Heuvelink, G. B., & Akyurek, S. Z. (2012). Comparison of regression and kriging techniques for mapping the average annual precipitation of Turkey. *International Journal of Applied Earth Observation and Geoinformation*, 19, 115-126
- [23] A. Hatamizadeh *et al.*, “UNETR: Transformers for 3D Medical Image Segmentation,” Mar. 2021, [Online]. Available: <http://arxiv.org/abs/2103.10504>
- [24] L. Blankemeier *et al.*, “Merlin: A Vision Language Foundation Model for 3D Computed Tomography,” Jun. 2024