

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO - CHILE



“MÉTODO DE SEGMENTACIÓN AUTOMÁTICA DE
HÍGADO EN IMÁGENES DE RESONANCIA MAGNÉTICA
UTILIZANDO UN MODELO DE APRENDIZAJE
AUTOMÁTICO”

GABRIELA PAZ ACUÑA BENITO

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERA CIVIL EN INFORMÁTICA

Profesor Guía: Julio Sotelo
Profesor Correferente: Carolina Saavedra

Septiembre - 2024

DEDICATORIA

A mis padres, Mónica y Fernando, a mi hermana Isidora y a mi pareja Tomás, quienes me han acompañado y apoyado incondicionalmente a lo largo de estos cinco años. A mi tío Alonso, mis tías Cecilia, Marcela, Andrea, y a mi abuela Mónica, por su constante preocupación por mi bienestar y sus palabras cariñosas ante cada desafío que me he propuesto. Finalmente, a mi Tata Miguel, que siempre está conmigo.

AGRADECIMIENTOS

A mi profesor guía, Julio Sotelo gracias a quien descubrí el área de las imágenes médicas. Agradezco profundamente su confianza y el apoyo brindado. A mis amigos Daniel y Joaquín, por su valioso apoyo en este proyecto.

RESUMEN

Resumen— Se propone un modelo de segmentación automática de hígado en imágenes de resonancia magnética utilizando técnicas de aprendizaje profundo, para luego compararlo con dos modelos del estado del arte. Esto se realiza utilizando el Duke Liver Dataset para entrenar y validar el modelo propuesto, basado en una arquitectura 3D U-Net. El modelo propuesto demostró ser competitivo y robusto, superando en pruebas a modelos disponibles del estado del arte. Los resultados son relevantes porque proporcionan una herramienta estandarizada y accesible que puede mejorar significativamente la eficiencia en el diagnóstico y los preparativos preoperatorios en enfermedades hepáticas crónicas.

Palabras Clave— Segmentación de Hígado, Enfermedades Hepáticas Crónicas, Resonancia Magnética.

ABSTRACT

Abstract— A model for automatic liver segmentation in MR imaging using deep learning techniques is proposed, and it is then compared with two state-of-the-art models. This is achieved by using the Duke Liver Dataset to train and validate the proposed model, which is based on a 3D U-Net architecture. The proposed model has proven to be competitive and robust, outperforming existing state-of-the-art models in testing. The results are significant because they provide a standardized and accessible tool that can greatly enhance the efficiency in diagnosing and preoperative planning in chronic liver diseases.

Keywords— Liver Segmentation, Chronic Liver Diseases, Magnetic Resonance.

GLOSARIO

CLD: *Chronic Liver Diseases*, enfermedades hepáticas crónicas.

RM: Resonancia Magnética.

MRI: *Magnetic Resonance Imaging*, Imagen de resonancia magnética.

CT: *Computed Tomography*, Tomografía computarizada.

DLDS: *Duke Liver Dataset*.

CNN: *Convolutional Neural Network*, Red neuronal convolucional.

NAFLD: *Non-Alcoholic Fatty Liver Disease*.

RF: *Radio Frequency*, Radiofrecuencia.

T: Tesla.

IP: *In Phase*, en fase.

OOP: *Out of phase*, fuera de fase.

GRE: *MRI gradient echo*, eco de gradiente en MRI.

CC: Componentes Conexas.

ÍNDICE DE CONTENIDOS

RESUMEN	IV
ABSTRACT	IV
GLOSARIO	V
ÍNDICE DE FIGURAS	VIII
ÍNDICE DE TABLAS	X
INTRODUCCIÓN	1
CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA	2
1.1 CONTEXTO Y DESCRIPCIÓN DEL PROBLEMA	2
1.2 OBJETIVOS	5
CAPÍTULO 2: MARCO CONCEPTUAL	6
2.1 ENFERMEDADES HEPÁTICAS	6
2.1.1 ETIOLOGÍA DE LAS CLD	7
2.1.2 PROGRESIÓN DE LAS CLD	8
2.1.3 TRATAMIENTOS	8
2.2 RESONANCIAS MAGNÉTICAS	9
2.2.1 SECUENCIA DE PULSOS DE RM	10
2.2.2 MÉTRICAS DE LAS RM	12
2.3 MODELOS DE INTELIGENCIA ARTIFICIAL	13
2.3.1 U-NET	14
2.3.2 3D UNET	17
2.3.3 NNUNET	20
2.4 EVALUACIÓN DE LOS MODELOS	22
2.4.1 DICE SCORE	22
2.4.2 INTERSECTION OVER UNION IOU	23
2.4.3 ESTIMACIÓN DE COMPONENTES CONEXAS	24
CAPÍTULO 3: PROPUESTA DE SOLUCIÓN	25
3.1 DUKE LIVER DATASET	25
3.1.1 ANÁLISIS ESTADÍSTICO DEL DATASET	25
3.1.2 ESTANDARIZACIÓN DATASET	27
3.1.3 ENTRENAMIENTO, VALIDACIÓN Y PRUEBA	28
3.2 MODELOS	29
3.2.1 ENTRENAMIENTO 3D UNET'S	29
3.2.2 MODELOS PRE-ENTRENADOS	30
3.3 COMPARACIÓN DE MODELOS	31

CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN	32
4.1 RESULTADOS SOBRE CONJUNTO DE PRUEBAS	32
4.2 RESULTADOS SOBRE SECUENCIAS NO VISTAS POR EL MODELO	36
4.3 RESULTADOS UTILIZANDO POST PROCESAMIENTO	37
4.4 RESULTADOS POR SECUENCIA	38
4.5 EXPERIMENTACIÓN CON CABECERA DICOM	41
CAPÍTULO 5: CONCLUSIONES	42
5.1 Conclusiones Generales	42
5.2 Cumplimiento de objetivos	42
5.3 Trabajo futuro	43
REFERENCIAS BIBLIOGRÁFICAS	45

ÍNDICE DE FIGURAS

1	Ejemplo de un corte axial de RM abdominal. DLDS_ID: 0001, Serie: 16, Secuencia: H, Fase Axial Opuesta.	4
2	Ejemplos de imágenes de morfología hepática (a) normal, (b) cambios leves de enfermedad hepática crónica con nodularidad superficial leve y (c) cambios avanzados de fibrosis con nodularidad superficial extensa y ensanchamiento de fisuras.	7
3	Causas de cirrosis hepática. Esta imagen fue traducida.	8
4	El empeoramiento de la fibrosis hepática conduce a cirrosis, que puede descompensarse y provocar, a su vez, carcinoma hepatocelular. Una intervención oportuna puede prevenir la progresión a la siguiente etapa de la enfermedad. En particular, puede prevenir el carcinoma hepatocelular. Un tratamiento adecuado puede provocar la regresión de la enfermedad de un estadio peor a uno mejor. Incluso la cirrosis puede retroceder, pero el carcinoma hepatocelular no. Esta imagen fue traducida.	9
5	Secuencias básicas de RM en imagen hepática.	10
6	Ejemplo de espaciado de píxeles, para (0.30,0.25). Imagen traducida.	11
7	Vistas axiales de una MRI abdominal extraídas en el estudio DLDS. Las secuencias visibles son: (a) Fuera de fase, (B) en fase, (C) precontraste T1w, y (D) contraste de fase venosa portal realizado T1w.	13
8	(A) Evaluación volumétrica por RM tras EVP derecha. Volumen del lóbulo hepático izquierdo: 353 ml. Volumen total: 672 ml. Crecimiento del 40,8 %. (B) Representación tridimensional de la medición volumétrica por RM. El hígado total y el lóbulo hepático izquierdo están representados a la izquierda y a la derecha, respectivamente. EVP, embolización de la vena porta.	14
9	Arquitectura UNet (ejemplo para 32x32 px en la menor resolución). Cada caja azul corresponde a un mapa de características de múltiples canales. El número de canales está denotado encima de la caja. El tamaño x-y se puede ver en el lado izquierdo inferior de la caja. Las cajas blancas representan mapas de características copiados. Las flechas denotan diferentes operaciones. Fuente: UNet: Convolutional Networks for Biomedical Image Segmentation [Ronneberger <i>et al.</i> , 2015]	15
10	Estrategia de superposición de mosaicos para la segmentación sin fisuras de imágenes arbitrarias de gran tamaño (aquí, segmentación de estructuras neuronales en <i>EM stacks</i> , una competencia iniciada por ISBI en 2012). La predicción de la segmentación en el área amarilla requiere datos de imagen dentro del área azul como entrada. Los datos de entrada faltantes se extrapolan mediante duplicación.	16

11	Paso a paso sobre la generación de deformación elástica en múltiples dimensiones. A) Volumen original de forma esférica. B) Un corte de la esfera para visualizar en dos dimensiones. C) Plano de distorsiones randomizadas. D) Deformaciones suavizadas con un filtro gaussiano. E) Corte del volumen original donde se puede ver el desplazamiento del volumen respecto al centro de la imagen. F) Comparación entre la esfera original y la deformada.	18
12	La arquitectura 3D UNet. Cajas azules representan mapas de características. El número de canales es denotado encima de cada mapa de características. .	19
13	Aprendizaje residual: un bloque de construcción. Fuente: Deep Residual Learning for Image Recognition [He <i>et al.</i> , 2016]	19
14	Bloque residual con una convolución en 3D, <i>batch normalization</i> y ReLU. . .	20
15	Especificación de componentes de nnUNet	21
16	Se ilustra el proceso de identificación de componentes conexas en una matriz 2D. A) Se revisan los valores de todos los vecinos del píxel a etiquetar. B) Ejemplo de un componente conexo de tamaño 4. C) Proceso de etiquetado en una matriz con más de un componente conexo.	24
17	Distribuciones demográficas y antropométricas de los pacientes: Comparación de sexo, edad, peso y tamaño en la población del estudio.	27
18	Transformaciones realizadas a las imágenes pertenecientes al DLDS.	28
19	Arquitectura de 3D UNet.	29
20	Arquitectura de la red 3D Res UNetA. Se pueden ver los procesos de <i>encoder</i> y <i>decoder</i> , con operaciones convolucionales residuales durante la codificación y la reconstrucción de imágenes segmentadas de múltiples objetivos mediante mapas de características deconvolucionales durante la decodificación. También se puede ver el Perceptrón Multicapa utilizado para codificar los Metadatos. Las conexiones salteadas se muestran en cada nivel de resolución.	30
21	Distribución de los puntajes Dice en formato de <i>boxplot</i> para los modelos evaluados.	33
22	Resultados cualitativos de las predicciones con mejor desempeño para cada modelo, en el conjunto de pruebas	35
23	Resultados cualitativos de las predicciones con peor desempeño para cada modelo, en el conjunto de pruebas	36
24	Puntaje Dice por Etiqueta en conjunto de prueba. Predicciones realizadas con 3D UNet. Fuente: Elaboración Propia	39
25	Puntaje Dice por Etiqueta en conjunto de prueba. Predicciones realizadas con 3D Res UNet. Fuente: Elaboración Propia	39
26	Puntaje Dice por Etiqueta en conjunto de prueba. Predicciones realizadas con Volumetry Net. Fuente: Elaboración Propia	40

27	Puntaje Dice por Etiqueta en conjunto de prueba. Predicciones realizadas con nnUNet. Fuente: Elaboración Propia	40
----	--	----

ÍNDICE DE TABLAS

1	Parámetros de las secuencias presentes en el dataset a utilizar.	12
2	Secuencias presentes en el DLDS, la cantidad de apariciones y cuantas de ellas se encuentran etiquetadas. Resaltadas las secuencias que se utilizarán para experimentación.	26
3	Promedio $\pm$ desviación estándar de las métricas relevantes, donde CC significa Componentes Conexas.	32
4	Ejemplos de las mejores predicciones obtenidas por cada uno de los modelos evaluados.	34
5	Ejemplos de predicciones con puntajes en la mediana para cada modelo, representando el desempeño promedio dentro de las muestras evaluadas. . .	34
6	Ejemplos de las peores predicciones realizadas por cada modelo, evidenciando los valores más bajos de puntaje Dice e IoU.	35
7	Parámetros de las secuencias menos presentes en el dataset a utilizar.	37
8	Promedio $\pm$ desviación estándar de las métricas relevantes para secuencias fuera del conjunto de datos principal.	37
9	Resultados utilizando únicamente el volumen más grande de la predicción realizada.	38
10	Dice promedio + desviación estándar por Secuencia según modelo evaluado en conjunto de prueba	38
11	Resultados de entrenamiento utilizando metadatos de la cabecera DICOM. 3D Res UNetA CC muestra los resultados cuando se considera únicamente la CC más grande de la predicción. Se incluyen los resultados ya presentados de 3D Res UNet como referencia	41

INTRODUCCIÓN

La segmentación automática de órganos en imágenes médicas propone un desafío en el campo de la salud, especialmente para el análisis y diagnóstico de enfermedades hepáticas crónicas [Lenchik *et al.*, 2019a]. La creciente prevalencia de estas enfermedades ha resaltado la necesidad de herramientas más precisas y eficientes para la evaluación del hígado, donde las imágenes de resonancia magnética (RM) juegan un papel fundamental debido a su alta resolución de contraste. Sin embargo, la segmentación manual sigue siendo el estándar de oro, presentando desafíos significativos en términos de consistencia y tiempo [Gotra *et al.*, 2017].

Aunque con la RM se puede realizar una detección temprana de anomalías hepáticas, los métodos automáticos para esta modalidad son limitados. Esta memoria busca contribuir con una solución estandarizada que mejore la precisión y eficiencia en el análisis volumétrico del hígado.

El objetivo principal de esta memoria es desarrollar y validar un modelo de segmentación automática de hígado utilizando técnicas avanzadas de aprendizaje profundo, este modelo se comparará con otros modelos del estado del arte, utilizando el *Duke Liver Dataset* (DLDS) para asegurar resultados reproducibles y generalizables. Para realizar esto, en primer lugar se presenta un análisis de los datos, para definir estrictamente los conjuntos de entrenamiento, validación y prueba. Luego se experimenta con diferentes configuraciones para el modelo de inteligencia artificial de segmentación de hígado en 3D. Por último, se comparan las predicciones realizadas por el algoritmo propuesto sobre el conjunto de prueba con dos modelos disponibles ya entrenados del estado del arte.

La estructura de este trabajo está organizada en cinco capítulos. En el Capítulo 1, se define el problema y se establecen los objetivos de la investigación. El Capítulo 2 presenta el marco conceptual, revisando el estado del arte respecto a segmentación automática y la relevancia de las RM en el diagnóstico hepático. En el Capítulo 3, se describe la metodología utilizada y se desarrolla la propuesta de solución. En el Capítulo 4 se valida la eficacia del modelo propuesto mediante experimentos y comparaciones. Finalmente, en el Capítulo 5, se exponen las conclusiones del estudio y se discuten posibles líneas de trabajo futuro.

CAPÍTULO 1

DEFINICIÓN DEL PROBLEMA

1.1. CONTEXTO Y DESCRIPCIÓN DEL PROBLEMA

La creciente incidencia de enfermedades hepáticas crónicas (CLD) y sus complicaciones, como la cirrosis, se ha convertido en un problema de salud global significativo, contribuyendo a un aumento en la mortalidad, morbilidad y carga económica [Asrani *et al.*, 2019]. La prevalencia de CLD, exacerbada por la enfermedad hepática no alcohólica (NAFLD) y la enfermedad hepática relacionada con el alcohol (ALD), ha llevado a un aumento del 8.1 % en las muertes relacionadas con la cirrosis en 2019, en comparación con 2017 [Liu y Chen, 2022]. Este incremento alarmante en la incidencia de CLD y sus complicaciones ha generado una creciente preocupación en el ámbito de la salud pública, ya que estas enfermedades no solo afectan la calidad de vida de los pacientes, sino que también implican una carga significativa para los sistemas de salud a nivel mundial.

De forma más específica, las enfermedades hepáticas causan 2 millones de muertes anualmente y son responsables de un 4 % de todas las muertes en el mundo. Una gran cantidad de esas muertes son atribuidas a complicaciones de la cirrosis y el carcinoma hepatocelular, y la hepatitis aguda representa una proporción menor de las muertes [Devarbhavi *et al.*, 2023]. Esto genera un aumento en el flujo de pacientes con necesidades especiales en los sistemas de salud, y en consecuencia la cantidad de exámenes hepáticos que deben ser revisados por el personal especializado aumenta exponencialmente.

En el contexto mencionado, la segmentación de hígado, un proceso crítico para el análisis volumétrico y morfológico del órgano, se ha convertido en un foco de atención. Aunque la segmentación manual sigue siendo el “estándar de oro”, esta enfrenta desafíos significativos, como la variabilidad intra e interobservador [Gotra *et al.*, 2017]. Este método depende en gran medida de la interacción del usuario, lo que puede llevar a resultados inconsistentes. Además, en casos de imágenes de baja calidad o contraste, los métodos semiautomáticos pueden tener dificultades para separar con precisión el hígado de los órganos adyacentes [Oliva y Saini, 2004]. Estos desafíos inherentes a la segmentación manual y semiautomática han impulsado la necesidad de desarrollar métodos más eficientes y precisos para el análisis de imágenes hepáticas.

Al existir una gran variedad de maquinarias para la toma de las imágenes con RM, diferentes configuraciones para estas y variaciones en la forma de los órganos en diferentes personas, la necesidad de métodos automáticos, eficientes y precisos se vuelve evidente. Estos métodos, impulsados por el aprendizaje profundo y las redes neuronales convolucionales, prometen mejorar la precisión y la velocidad en el diagnóstico y tratamiento de las CLD [Liu *et al.*, 2021]. Lenchik *et al.* afirman que la segmentación automática ha facilitado el desarrollo de marcadores de imágenes cuantitativas más rápidos y reproducibles, aunque la transición a la práctica

clínica sigue siendo un desafío [Lenchik *et al.*, 2019b]. En este caso también se puede notar la implementación directa para la volumetría de hígado, procedimiento crucial para trasplantar el órgano. La segmentación automática del hígado basada en técnicas de aprendizaje profundo ha demostrado ser una herramienta prometedora para superar las limitaciones de los métodos manuales y semiautomáticos, ofreciendo resultados más precisos y reproducibles en un menor tiempo [Gul *et al.*, 2022].

A pesar de que la tomografía computarizada (TC), es una de las técnicas de imágenes pre-dilectas para segmentar el hígado, Sharma *et al.* subraya la sensibilidad superior de la RM en la detección de anomalías, especialmente en las etapas tempranas de la enfermedad [Sharma y Aggarwal, 2010]. Esta observación resalta la necesidad de optimizar los métodos automáticos para la RM, ya que se ha convertido en una herramienta fundamental para el diagnóstico y seguimiento de las CLD. La RM ofrece una mayor resolución de contraste y la capacidad de visualizar estructuras anatómicas en múltiples planos, lo que la convierte en una modalidad de imagen ideal para la evaluación detallada del hígado y la detección temprana de cambios patológicos [Lenchik *et al.*, 2019b].

Pese a la prevalente necesidad, la accesibilidad a algoritmos avanzados de segmentación sigue siendo un obstáculo significativo en la práctica clínica. Romero R. *et al.* (2022) señalan que muchos algoritmos prometedores no están disponibles en *frameworks* de código abierto o para uso libre, lo que limita la investigación reproducible. Los *softwares* comerciales, por otro lado, son a menudo inasequibles y específicos para ciertas marcas de equipos de RM, lo que restringe su aplicabilidad en diversos contextos [Romero R *et al.*, 2022]. Esta falta de accesibilidad a herramientas de segmentación automática eficientes y asequibles ha dificultado la adopción generalizada de estos métodos en la práctica clínica.

Finnegan *et al.* enfatizan la importancia de las soluciones de código abierto en este escenario. No solo facilitan la implementación y validación colaborativa de algoritmos de segmentación, sino que también promueven la atención médica de calidad en diversos entornos [Finnegan *et al.*, 2023]. En el contexto de la creciente prevalencia de CLD, estas soluciones podrían ser cruciales para el análisis de grandes conjuntos de datos y la mejora de la modelación del riesgo para los pacientes, abordando así la urgente necesidad de estrategias efectivas para el diagnóstico y tratamiento de CLD. Las soluciones de código abierto permiten a los investigadores y profesionales de la salud acceder, adaptar y mejorar los algoritmos de segmentación según sus necesidades específicas, fomentando la innovación y la colaboración en el campo de la imagen médica.

Por lo anterior, resulta de gran importancia contar con una solución que integre herramientas de segmentación automática, de código abierto y optimizadas para la RM. Esto no solo abordará los desafíos de precisión y accesibilidad, sino que también potenciaría la eficacia en el diagnóstico y tratamiento de las CLD. Al disponer de métodos de segmentación automática precisos y accesibles, los profesionales de la salud podrán realizar análisis volumétricos y morfológicos del hígado de manera más eficiente, lo que facilitará la detección temprana de anomalías y la toma de decisiones clínicas informadas. Además, la disponibilidad de estas herramientas en plataformas de código abierto fomentará la colaboración entre investigado-

res y la mejora continua de los algoritmos, acelerando el progreso en el campo de la imagen médica aplicada a las enfermedades hepáticas crónicas.

Un tema relevante a tener en cuenta en el desarrollo de esta solución, es la ética en el trato de los datos. Ya que la recolección de estos fue realizada por el equipo de la universidad de Duke, para la generación del *Duke Liver Dataset*, ellos anonimizaron todos los datos entregados para que realizar cualquier relación entre el dato y el paciente original sea imposible [Macdonald *et al.*, 2023]. En el desarrollo de esta memoria, estoy consciente de los temas éticos que orientan la elaboración de un documento en Ingeniería Informática. Estos temas incluyen la protección de la privacidad y la seguridad de los datos, la transparencia en la presentación de los resultados y la consideración de los impactos sociales y ambientales de las soluciones propuestas.

Volviendo a los datos a tratar para esta problemática, la RM (figura 1), es una técnica de imagen utilizada para obtener información respecto a la estructura y composición de diferentes partes del cuerpo humano. Históricamente, la revisión de estas imágenes se ha realizado de forma cualitativa por radiólogos especializados. Esto puede llevar a problemas de eficiencia y consistencia, ya que los resultados de un examen están sujetos a la opinión del radiólogo en el momento en el que revisó la imagen.

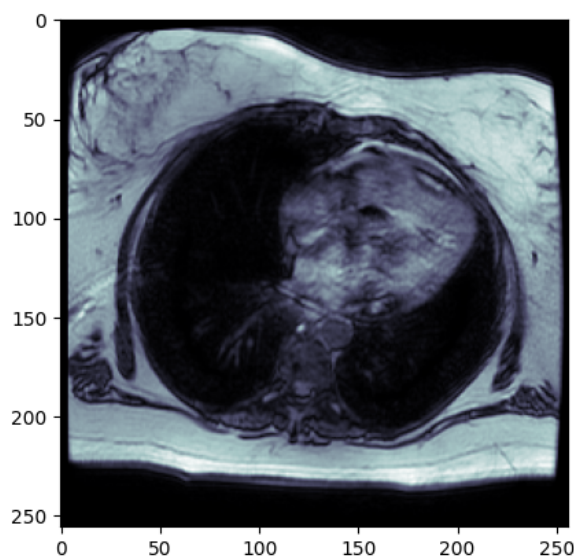


Figura 1: Ejemplo de un corte axial de RM abdominal. DLDS_ID: 0001, Serie: 16, Secuencia: H, Fase Axial Opuesta.

Fuente:Elaboración Propia.

En los últimos años, se ha demostrado crecientemente la relevancia de las técnicas de *Quantitative imaging* [Cashmore *et al.*, 2021], y se han desarrollado soluciones de segmentación automática de diferentes órganos [Lenchik *et al.*, 2019b] para diferentes tipos de imágenes médicas, especialmente para Cerebros con RM [Singh y Singh, 2021]. Se ve una falta de esta implementación para Hígados en imágenes tomadas por RM y es lo que este trabajo es-

para solucionar. Existen implementaciones previas [Huynh *et al.*, 2014] [Gotra *et al.*, 2017] [Jansen *et al.*, 2019], sin embargo, estas utilizan pocos datos de referencia, lo que genera soluciones poco generalizables.

La segmentación de Hígado en imágenes médicas es de especial importancia y utilidad, ya que con esta información se podría realizar volumetría del hígado, cuantificación automatizada de fracción de grasa de densidad de protones (o *proton density fat fraction*, PDFF) hepático, entre otros [Wang *et al.*, 2019]. Este trabajo se encuentra específicamente enfocado en la volumetría del hígado, una métrica de vital importancia para el trasplante de hígado.

1.2. OBJETIVOS

El objetivo general de este trabajo es implementar un algoritmo de segmentación automática de hígado en imágenes de RM.

En cuanto a los objetivos específicos, estos son:

- Definir un algoritmo de segmentación automática de hígado a partir del estudio del estado del arte, en segmentación de órganos en imágenes tomadas con RM.
- Validar el modelo confirmando la efectividad y precisión del algoritmo propuesto.
- Comparar el modelo propuesto con dos modelos del estado del arte en un conjunto de pruebas.

CAPÍTULO 2

MARCO CONCEPTUAL

Dada la naturaleza del proyecto a realizar, se iniciará con una explicación en profundidad respecto a las enfermedades hepáticas 2.1. Luego es necesario dar una explicación respecto a como funciona la RM (lo suficiente para entender las imágenes a trabajar) 2.2. Por último, se realizará una explicación detallada de los modelos del estado del arte a probar y el modelo propuesto en la sección 2.3, al igual que el análisis estadístico de estos en la sección 2.4.

2.1. ENFERMEDADES HEPÁTICAS

Las enfermedades hepáticas crónicas se desarrollan a partir de un daño hepático prolongado que dura meses o años y la cirrosis es la etapa terminal de la enfermedad hepática crónica (CLD). Las causas primarias de cirrosis mundialmente están relacionadas con la hepatitis viral, alcohol, y enfermedad del hígado graso no alcohólico. Históricamente, los virus hepatotrópicos han sido las causas principales de los casos de hepatitis aguda, pero la lesión hepática inducida por fármacos representa cada vez una proporción más importante de los casos [Devarbhavi *et al.*, 2023].

Debido a que la naturaleza de las enfermedades hepáticas implican un cambio en la morfología del hígado, la tarea de segmentación automática y estandarización de diagnósticos aumenta en dificultad. Esto se ve ilustrado en la Figura 2 donde se ve como cambia la morfología del hígado a medida que la enfermedad avanza. Se puede notar la clasificación cualitativa utilizada, donde se considera principalmente la Nodularidad de la Superficie Hepática (LSN). Inicialmente, la LSN en pacientes con vCLD (*viral CLD*) es causada por fibrosis y la formación de nódulos regenerativos, los cuales aumentan en tamaño y cantidad a medida que la enfermedad viral progresa. Estos nódulos y cicatrices suelen ser visibles en imágenes médicas como ecografías, tomografías computarizadas, entre otros. La evaluación cualitativa de LSN junto con varios signos radiológicos adicionales, como la redistribución lobar, la expansión de la fosa vesicular, el estrechamiento de las venas hepáticas, la relación aumentada del lóbulo caudado al derecho y el agrandamiento del espacio periportal hilar, se utilizan para distinguir entre hígados cirróticos y no cirróticos durante el monitoreo por imágenes. En este estudio, se revisan los exámenes de 309 pacientes, comparando sus diagnósticos cualitativos basándose en imágenes de RM con los resultados de histopatología¹. Se encontró que solo un 10 % de los resultados eran concordantes entre sí, sin embargo, de los reportes discordantes, un 90 % identificaron correctamente la CLD, pero subestimaron la gravedad de esta [Nandwana *et al.*, 2021]. Esto nos lleva nuevamente a confirmar la necesidad de medidas cuantitativas confiables en el área.

¹Resultados cuantitativos de exámenes de tejidos como biopsias.

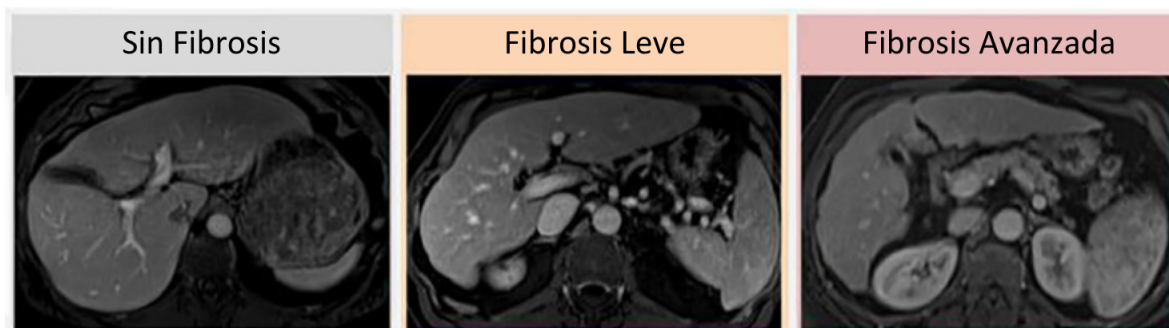


Figura 2: Ejemplos de imágenes de morfología hepática (a) normal, (b) cambios leves de enfermedad hepática crónica con nodularidad superficial leve y (c) cambios avanzados de fibrosis con nodularidad superficial extensa y ensanchamiento de fisuras.

Fuente: [Nandwana *et al.*, 2021]

2.1.1. ETIOLOGÍA DE LAS CLD

Como se mencionó previamente, las enfermedades hepáticas crónicas (CLD) implican la deterioración de la función hepática, caracterizada por inflamación sostenida, fibrosis y daño de las células hepáticas, eventualmente desembocando en cirrosis. La etiología de CLD es multifacética, incluyendo agentes infecciosos y no infecciosos. Causas infecciosas clave incluyen infecciones crónicas con los virus de la Hepatitis B (HBV) y la Hepatitis C (HCV), que inducen constante inflamación del hígado y cambios fibróticos. En Latinoamérica y particularmente en Chile, el HCV es un factor predominante debido a prácticas de transfusión de sangre antiguas y al uso de drogas intravenosas. Otro gran factor es el consumo de alcohol, que causa enfermedad hepática alcohólica (ALD) a través de lesiones hepáticas recurrentes debido al metabolismo del etanol y al estrés oxidativo, lo que lleva a hígado graso, esteatohepatitis y eventualmente cirrosis [Díaz *et al.*, 2024].

La creciente prevalencia de trastornos metabólicos como la obesidad y la diabetes tipo 2 ha traído la enfermedad de hígado graso no alcohólico (NAFLD) a primer plano, y esta progresa en muchos individuos a esteatohepatitis no alcohólica (NASH) [Pouwels *et al.*, 2022]. Esto está impulsado por la resistencia a la insulina, la acumulación lipídica en hepatocitos y las vías inflamatorias. Adicionalmente, la hepatitis autoinmune y los trastornos genéticos como la hemocromatosis y la enfermedad de Wilson contribuyen a la CLD al causar un daño persistente a las células hepáticas. Factores ambientales como la exposición a aflatoxinas y otras hepatotoxinas, y factores de estilo de vida, incluyendo fumar y una mala dieta, también juegan un rol importante en la etiología de la enfermedad. Colectivamente, estas diversas causas subrayan la complejidad de la enfermedad hepática crónica, lo que requiere estrategias multifacéticas de prevención y tratamiento adaptadas a los factores etiológicos específicos que prevalecen en diferentes regiones [Azad *et al.*, 2022, Pouwels *et al.*, 2022].

Otras de las causas más comunes de enfermedades hepáticas crónicas pueden ser vistas en la figura 3, donde cabe destacar la gran variedad de categorías representadas.

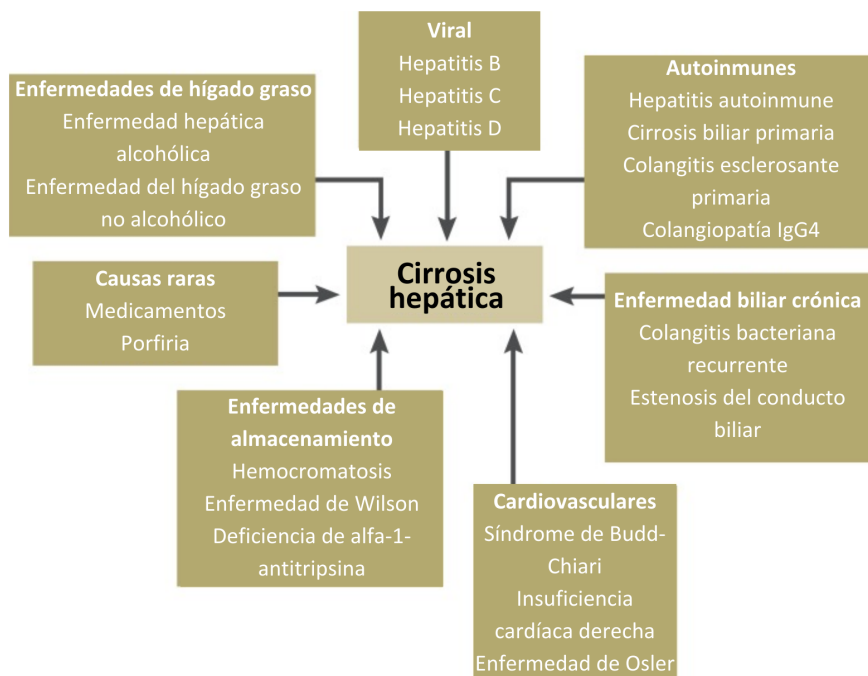


Figura 3: Causas de cirrosis hepática. Esta imagen fue traducida.
Fuente: [Wiegand y Berg, 2013]

2.1.2. PROGRESIÓN DE LAS CLD

El diagnóstico de la cirrosis se realiza a través de hallazgos característicos en el examen clínico, pruebas de laboratorio y estudios complementarios [Wiegand y Berg, 2013]. La progresión de estas es un proceso complejo que involucra múltiples etapas patológicas y mecanismos fisiológicos. Estas tienden a progresar de forma lenta con el paso de los años, por lo que se pueden tomar bastantes medidas de prevención antes de llegar a una cirrosis. En la Figura 4 se puede visualizar un mapa de la progresión de las CLD según el tiempo y efecto del tratamiento efectuado.

2.1.3. TRATAMIENTOS

El manejo de las CLD es multifacético e implica una combinación de intervenciones médicas, cambios en el estilo de vida y en casos avanzados, operaciones quirúrgicas como el trasplante de hígado. La selección del tratamiento adecuado depende de la etapa de la enfermedad, sus causas y la condición general del paciente.

De los diferentes enfoques terapéuticos que se pueden tomar, el más relevante para este trabajo es el trasplante de hígado. Esta es una opción crucial para pacientes en fase terminal o con insuficiencia hepática aguda que no responde a otros tratamientos.

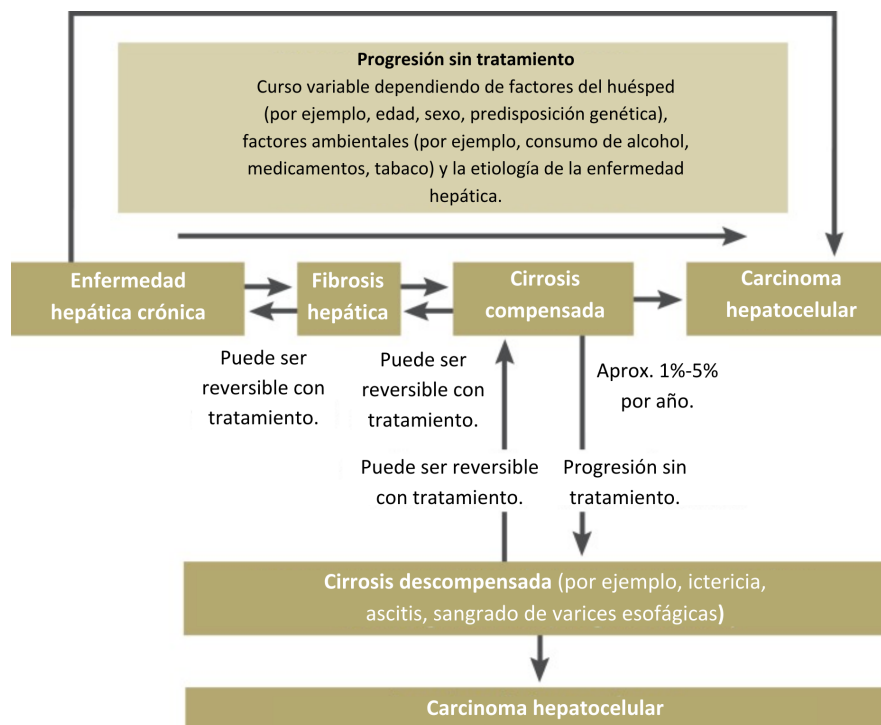


Figura 4: El empeoramiento de la fibrosis hepática conduce a cirrosis, que puede descompensarse y provocar, a su vez, carcinoma hepatocelular. Una intervención oportuna puede prevenir la progresión a la siguiente etapa de la enfermedad. En particular, puede prevenir el carcinoma hepatocelular. Un tratamiento adecuado puede provocar la regresión de la enfermedad de un estadio peor a uno mejor. Incluso la cirrosis puede retroceder, pero el carcinoma hepatocelular no. Esta imagen fue traducida.

Fuente: [Wiegand y Berg, 2013]

La demanda de órganos supera significativamente la oferta, por lo que las innovaciones para aumentar el número de donantes utilizables y asegurar un buen uso de los órganos disponibles son esenciales. La evaluación precisa del volumen hepático, conocida como volumetría hepática, es crucial en este contexto. La volumetría hepática proporciona métricas cuantitativas vitales para evaluar la viabilidad de trasplantes y planificar procedimientos quirúrgicos, asegurando que los injertos sean de tamaño adecuado y funcionalmente eficientes para los receptores. Además, estas métricas permiten una mejor selección y manejo de donantes y receptores, contribuyendo a mejorar los resultados postoperatorios y la supervivencia a largo plazo [Terrault *et al.*, 2023, D'Onofrio *et al.*, 2014].

2.2. RESONANCIAS MAGNÉTICAS

Un punto de suma importancia en la problemática a tratar, es el formato de las imágenes tomadas por RM. Para la toma de estas imágenes, la máquina de RM utiliza bobinas su-

perconductoras, que producen fuertes campos magnéticos que fuerzan a los protones del átomo de hidrógeno del cuerpo humano a alinearse con el campo generado. Luego se envía una señal de radiofrecuencias a través del cuerpo del paciente, los protones se estimulan. Cuando se apaga el campo de radiofrecuencia, las bobinas de recepción de la máquina pueden detectar la energía liberada cuando los protones se realinean con el campo magnético. Tanto el tiempo que tardan los protones en realinearse con el campo magnético como la cantidad de energía liberada cambia dependiendo del entorno y la naturaleza química de las moléculas y los especialistas pueden distinguir entre diferentes tipos de tejidos basados en estas propiedades magnéticas ².

Un examen estándar de RM del hígado, se compone de 7 series principales. Como se puede ver en la figura 5, existen dos secuencias primarias, *T1-weighted* y *T2-weighted* [Vu et al., 2017]. Estas nomenclaturas hacen referencia a la sensibilidad de la imagen en los tiempos de relajación T1 y T2 en el núcleo del átomo de hidrógeno en el cuerpo. Una imagen de RM en T1 tiene un alto contraste entre los tejidos, destacando la anatomía del hígado. Por otro lado, una imagen de RM en T2 se centra en la patología, mostrando los fluidos en tonalidades brillantes, lo que resulta ideal para visualizar inflamaciones, edemas y ciertas lesiones, como tumores y daños por trauma [McRobbie et al., 2017].

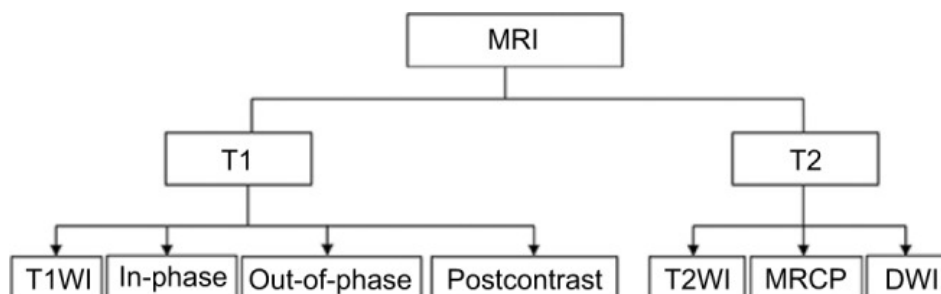


Figura 5: Secuencias básicas de RM en imagen hepática.
Fuente: [Vu et al., 2017]

2.2.1. SECUENCIA DE PULSOS DE RM

En RM, una *sequence* (secuencia) o *pulse sequence* (secuencia de pulsos), se refiere al momento y duración específicos de los pulsos electromagnéticos y las señales de Radio Frecuencia (RF) utilizados para crear una imagen. Cada secuencia está diseñada para producir un tipo específico de contraste entre diferentes tipos de tejido en el cuerpo. Una secuencia está compuesta por 4 etapas principales: excitación, codificación, manipulación y detección. Durante la etapa de excitación, se aplica el campo magnético para alinear los protones. Luego, en la etapa de codificación se aplica un campo magnético de gradiente, donde cada tejido produce una señal ligeramente diferente dependiendo de su ubicación. En la etapa de manipulación

²Información extraída del *National institute of Biomedical Imaging and Bioengineering*, sección *Science education*

se aplica un pulso RF para alterar la alineación de los protones, produciendo el contraste deseado. Por último se realiza la detección, donde se miden las señales RF emitidas por los protones y convirtiéndolas en imágenes digitales [McRobbie *et al.*, 2017, Dale *et al.*, 2015].

Los principales parámetros de una secuencia a tener en cuenta son los siguientes:

- *No. of Sections*: Número de cortes realizados.
- *Slice Thickness* (mm): Grosor del corte tomado, es determinado por el gradiente de selección de corte aplicado.
- *Pixel Spacing* (mm): Espaciado de píxeles, distancia física en el paciente entre centro de cada pixel, especificado por un par de números, para filas y columnas en mm (Figura 6).
- *Matrix Size*, Números de filas y columnas de la imagen digital de la RM.
- *TR* (msec): *Repetition Time*, Tiempo de repetición se refiere al intervalo de tiempo entre excitaciones sucesivas de los tejidos que se están examinando.
- *TE* (msec): *Echo Time*, Tiempo de Eco, se refiere al tiempo entre el pulso de excitación y el máximo de la señal de RM generada por los átomos de hidrógeno en el cuerpo.
- *Flip Angle* (°): Ángulo de giro, determina el ángulo con el cual el campo magnético es aplicado a los átomos de hidrógeno en el cuerpo.

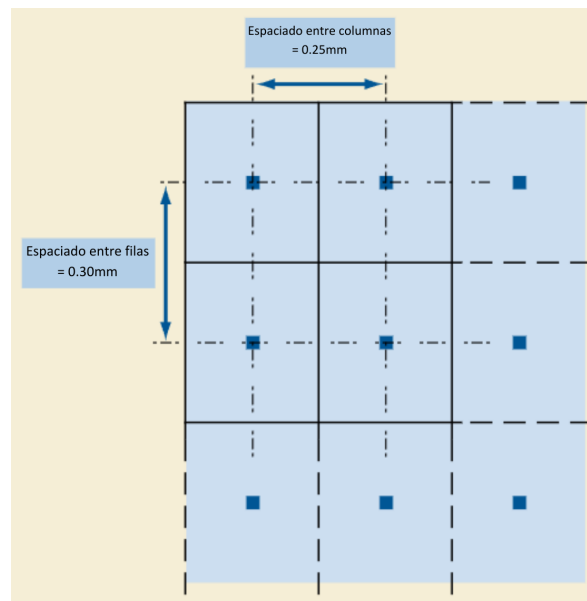


Figura 6: Ejemplo de espaciado de píxeles, para (0.30,0.25). Imagen traducida.
Fuente: DICOM PS3.3 2024b - Information Object Definitions.

En el *dataset* a utilizar, se tienen principalmente 4 tipos de secuencias, las que se muestran en la Tabla 1. Las secuencias con T1w (*T1 weighted*), demuestran las diferencias en los tiempos de relajación T1 de los tejidos y tienden a tener tiempos de TE y TR cortos. Ya que la grasa se realinea rápidamente, se ve más brillante (blanca) en una imagen T1w. De forma contraria, el agua tiene una velocidad de realineación menor, terminando de realinearse después del pulso en TR por lo que se ve oscura [Chen *et al.*, 2018]. Otras, son las secuencias en fase (IP) y fuera de fase (OOP), importantes para evaluar la composición de los tejidos, especialmente en áreas donde la presencia de grasa puede interferir con la interpretación de la imagen. En la secuencia en fase (IP), las moléculas de grasa y agua están alineadas de manera similar, lo que resulta en una señal más fuerte en las zonas donde hay más grasa presente. Por otro lado, en la secuencia fuera de fase (OOP), las moléculas de grasa y agua están desalineadas, lo que lleva a una atenuación de la señal de la grasa, lo que la puede hacer más oscura en comparación a IP [Reeder *et al.*, 2011, Westbrook, 2016].

Tabla 1: Parámetros de las secuencias presentes en el dataset a utilizar.
Fuente: *Duke Liver Dataset* [Macdonald *et al.*, 2023].

Sequence	Scan Count	No. of Sections	Slice Thickness (mm)	Pixel Spacing (mm)	Matrix Size Rows	Matrix Size Columns	TR (msec)	TE (msec)	Flip Angle (°)
T1w fat sat	220	48-100	3-7	0.68-1.76	168-576	256-576	3.46-9.2	1.07-3.13	4-15
T1w Portal venous	106	44-100	3-7	0.68-1.56	208-512	256-512	4-6.78	1.23-3.27	9-12
Axial in phase	115	22-100	4-8	0.69-1.76	168-576	256-576	3.84-175	2.46-7.38	4-90
Axial opposed phase	113	22-100	4-8	0.69-1.76	168-576	256-576	3.84-175	1.23-6.15	4-90

Un ejemplo de las diferentes imágenes generadas por cada tipo de secuencia se pueden ver en la Figura 7, donde también se puede ver la máscara ³ que viene asociada a cada imagen disponible y su secuencia. En esta imagen se puede notar explícitamente cómo la secuencia impacta las características visuales, por ejemplo en la sección D, se ve un hígado en tonos más claros al utilizar la secuencia T1w con contraste en la fase venosa portal realzada.

2.2.2. MÉTRICAS DE LAS RM

Desde la imagen por RM, se pueden extraer diversas métricas, lo cual actualmente se realiza de forma manual por parte de los radiólogos. Estas tienen un impacto muy relevante en el diagnóstico de los pacientes y en la estimación de los avances y gravedad de diversas enfermedades.

Específicamente, en la evaluación preoperatoria del volumen hepático, la imagen por RM es crucial para estimar con precisión el volumen total del hígado, así como sus segmentos individuales. Esta estimación es esencial para la planificación de resecciones hepáticas extensas, como trasplantes o hepatectomías mayores, y para evaluar la regeneración

³Sección de la imagen manualmente etiquetada donde está el hígado.

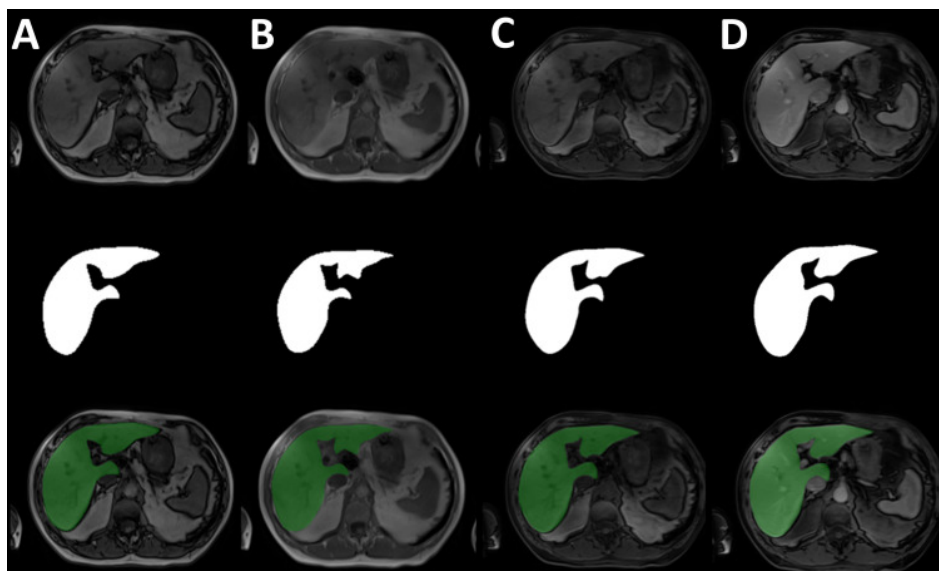


Figura 7: Vistas axiales de una MRI abdominal extraídas en el estudio DLDS. Las secuencias visibles son: (a) Fuera de fase, (B) en fase, (C) precontraste T1w, y (D) contraste de fase venosa portal realizado T1w.

Fuente: *Duke Liver Dataset* [Macdonald et al., 2023].

segmentaría postoperatoria. Además, la MRI permite identificar modificaciones volumétricas del hígado introducidas por procedimientos pre quirúrgicos de hiperatrofia hepática [Le Roy et al., 2018, D'Onofrio et al., 2014].

Un ejemplo de como se extrae la volumetría hepática, se puede visualizar en la figura 8. Se puede notar que la extracción se hace corte a corte y, posteriormente, se puede realizar una proyección en 3D para visualizar el volumen hepático.

2.3. MODELOS DE INTELIGENCIA ARTIFICIAL

El campo de la segmentación automática de órganos en imágenes médicas ha experimentado un avance significativo en la última década, especialmente en lo que respecta a la segmentación de imágenes de tomografía computarizada (CT). Estudios recientes han demostrado que los modelos basados en redes neuronales profundas, con la base de UNets, han alcanzado un rendimiento notable en la segmentación de órganos como el hígado, con una alta precisión y consistencia [Gul et al., 2022]. Es importante notar que los datasets más grandes y disponibles de forma liberada, con etiquetas para el hígado, son de exámenes de CT, por lo que las soluciones son orientadas a este dominio [Bilic et al., 2023, Heimann et al., 2009]. Por esta razón, la investigación en la segmentación de imágenes de RM ha sido menos desarrollada debido a la escasez de datasets públicos disponibles hasta hace poco. Este retraso ha llevado a que muchos estudios iniciales se centraran en la modalidad de CT, dejando la RM

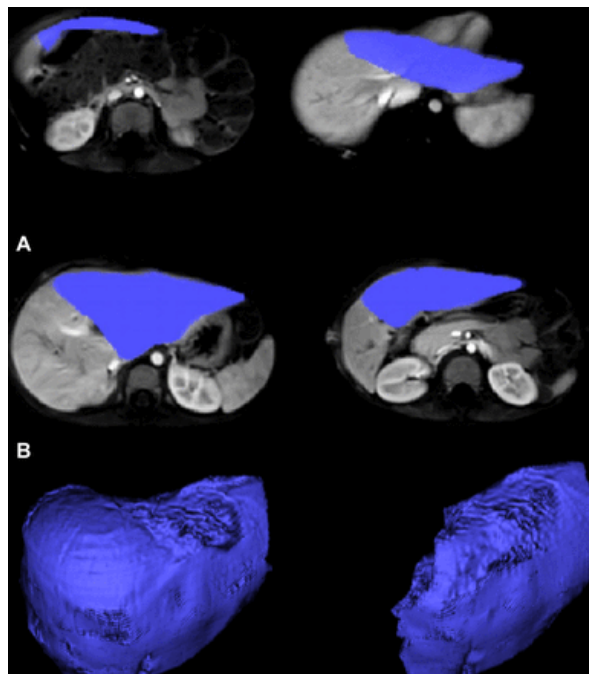


Figura 8: (A) Evaluación volumétrica por RM tras EVP derecha. Volumen del lóbulo hepático izquierdo: 353 ml. Volumen total: 672 ml. Crecimiento del 40,8 %. (B) Representación tridimensional de la medición volumétrica por RM. El hígado total y el lóbulo hepático izquierdo están representados a la izquierda y a la derecha, respectivamente. EVP, embolización de la vena porta.

Fuente: [Glinka *et al.*, 2019].

en un segundo plano hasta la reciente liberación del DLDS en 2023, que proyecta comenzar a llenar esta brecha en la investigación.

2.3.1. U-NET

Las redes UNet son una arquitectura de redes neuronales profundas desarrolladas específicamente para la segmentación de imágenes biomédicas. Introducidas por Ronnenberger *et al.* en 2015, han demostrado ser altamente eficaces en tareas de segmentación debido a su capacidad para capturar tanto el contexto global como detalles locales de las imágenes [Ronneberger *et al.*, 2015].

La UNet incluye un camino de contracción para capturar contexto y un camino de expansión para una localización detallada, facilitando la asignación de una etiqueta de clase a cada píxel de una imagen. Esta estructura garantiza la aplicación de cada operación de agrupamiento máximo a una capa con un tamaño x-y par, resultando en un mosaico de salida sin fisuras y una segmentación altamente precisa. Un ejemplo se puede ver en la Figura 9, donde queda clara la razón de la nomenclatura.

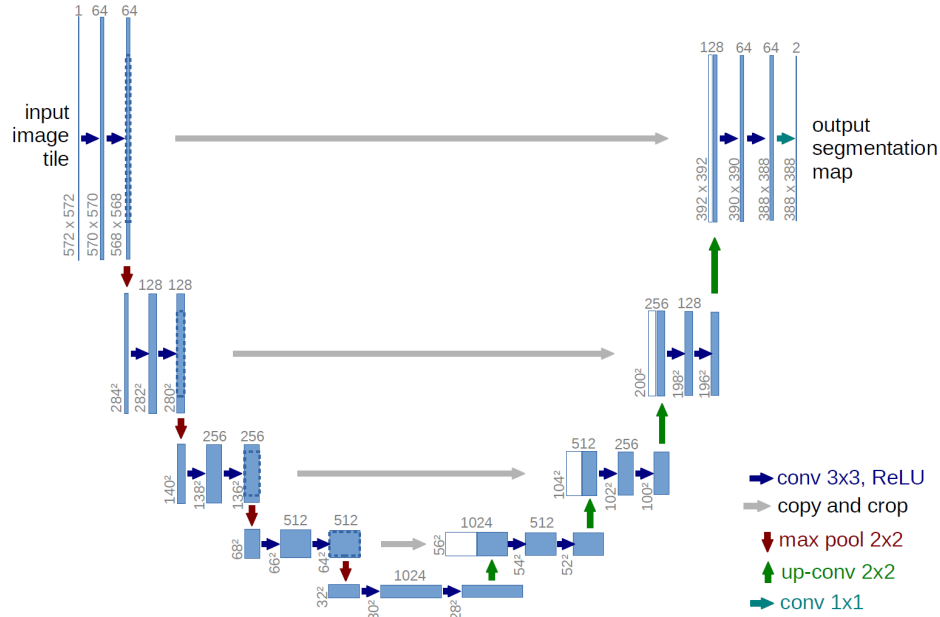


Figura 9: Arquitectura UNet (ejemplo para 32x32 px en la menor resolución). Cada caja azul corresponde a un mapa de características de múltiples canales. El número de canales está denotado encima de la caja. El tamaño x-y se puede ver en el lado izquierdo inferior de la caja. Las cajas blancas representan mapas de características copiados. Las flechas denotan diferentes operaciones. Fuente: UNet: Convolutional Networks for Biomedical Image Segmentation [Ronneberger *et al.*, 2015]

Esta arquitectura ha demostrado ser exitosa en segmentación de diversas imágenes biomédicas, incluyendo tejidos del cerebro, riñón, hígado y pulmón, así como la segmentación de células individuales en microscopía [Azad *et al.*, 2022, Liu *et al.*, 2020].

Es importante notar que la idea de la UNet surge de la clásica *fully convolutional network* (FCN), una red donde solo se utilizan capas conectadas localmente, como convolución, agrupamiento y sobremuestreo [Long *et al.*, 2014]. Evitar el uso de capas densas implica menos parámetros, lo que hace que las redes sean más rápidas de entrenar. Esto también significa que una FCN puede funcionar con tamaños de imagen variables siempre que todas las conexiones sean locales. Las FCN utilizan conexiones de salto (que también se ven en las UNets) o *skip connections* como son conocidas. Estas se utilizan para recuperar información espacial detallada que se puede haber perdido en el proceso de submuestreo.

En cuanto a la arquitectura de la UNet clásica, el camino de contracción sigue la arquitectura típica de una red convolucional. Consiste de la aplicación repetitiva de dos convoluciones de 3x3 (*unpadded*), cada una seguida de una unidad rectificadora lineal (ReLU) y un *max-pooling* de 2x2 con *stride* 2 para submuestrear. En cada paso de submuestreo se duplica la cantidad de canales de características. Cada paso en el camino expansivo, consiste de un sobremuestreo del mapa de características, con una convolución de 2x2 que reduce a la mitad el número de canales de características, una concatenación con el mapa de características

correspondiente del camino de contracción (expresado con las flechas grises en la figura 9), y dos convoluciones de 3x3, cada una seguida de una ReLu. Al realizar las conexiones de salto, es necesario realizar un corte a los mapas de características, ya que gracias a la naturaleza de las operaciones convoluciones, se pierde un borde de píxeles en cada convolución. La capa final es una convolución de 1x1, utilizada para asignar cada vector de características de 64 componentes al número deseado de clases. La red en total cuenta con 23 capas. Debido a la pérdida de borde en las convoluciones, para obtener una segmentación completa, se aumentan los bordes de la imagen original utilizando extrapolación por reflejo.

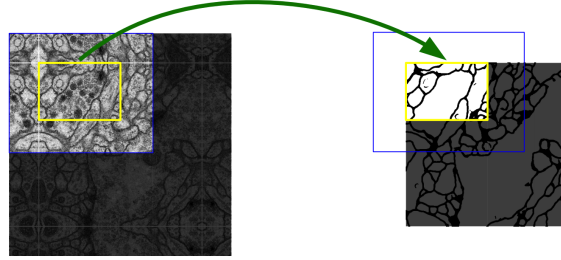


Figura 10: Estrategia de superposición de mosaicos para la segmentación sin fisuras de imágenes arbitrarias de gran tamaño (aquí, segmentación de estructuras neuronales en *EM stacks*, una competencia iniciada por ISBI en 2012). La predicción de la segmentación en el área amarilla requiere datos de imagen dentro del área azul como entrada. Los datos de entrada faltantes se extrapolan mediante duplicación.

Fuente: UNet: Convolutional Networks for Biomedical Image Segmentation
[Ronneberger *et al.*, 2015]

En cuanto al entrenamiento de la red propuesta por Ronnenberger, la función de energía es calculada por una función *soft-max* sobre el mapa de características final combinado con la función de pérdida de entropía cruzada (*cross entropy*). La función *softmax* se define de la siguiente manera,

$$p_k(x) = \frac{\exp(a_k(x))}{\sum_{k'=1}^K \exp(a_{k'}(x))} \quad (1)$$

donde $a_k(x)$ denota la activación en el canal de características k en la posición $x \in \Omega$ con $\Omega \subset \mathbb{Z}^2$. K es el número de clases y $p_k(x)$ es la función de máximo aproximada. Por ejemplo, $p_k(x) \approx 1$ para la k que tiene la activación máxima $a_k(x)$ y $p_k(x) \approx 0$ para todas las otras k . La entropía cruzada penaliza entonces en cada posición la desviación de $p_{l(x)}(x)$ respecto de 1 utilizando

$$E = \sum_{x \in \Omega} \omega(x) \log(p_{l(x)}(x)) \quad (2)$$

donde $l : \Omega \rightarrow \{1, \dots, K\}$ es la etiqueta verdadera de cada pixel. La función $l(x)$ depende de x porque en problemas de segmentación de imágenes, cada píxel tiene una clase verdadera asociada, y esta clase puede variar a lo largo de la imagen. Además, $w : \Omega \rightarrow \mathbb{R}$ es un mapa de pesos que es introducido para darle más importancia a unos píxeles que otros durante el entrenamiento.

El último punto a tratar sobre la red en sí, es la inicialización de sus pesos. Idealmente, estos pesos iniciales deberían estar adaptados de tal forma que cada mapa de características en la red tenga una varianza aproximadamente unitaria, en este caso en particular, esto se puede lograr obteniendo los pesos iniciales desde una distribución Gausiana con una desviación estándar de $\sqrt{2N}$, donde N denota el número de nodos entrantes a una neurona. Por ejemplo, para una convolución de 3x3 y 64 canales de características en la capa anterior, $N=9 \cdot 64=576$.

Lo último que se menciona en el estudio de Ronnenberger et al. es el aumento de datos utilizando diferentes estrategias. El aumento de datos es esencial para enseñarle invariancia y robustez a la red cuando se cuenta con pocos ejemplos para entrenamiento. En el caso de imágenes microscópicas, se presenta la necesidad de invariancia de desplazamiento y rotación, así como robustez a las deformaciones y a las variaciones del valor gris. Especialmente deformaciones elásticas randomizadas de los ejemplos de entrenamiento que parecen ser el concepto clave para entrenar una red de segmentación con muy pocas imágenes anotadas. Generan deformaciones suaves utilizando vectores de desplazamiento aleatorios en una cuadrícula gruesa de 3x3. Los desplazamientos se toman de una distribución gaussiana con una desviación estándar de 10 píxeles. Luego se calculan los desplazamientos por pixel utilizando interpolación bicúbica. Las capas de exclusión (*drop out layers*) al final de la trayectoria de contracción realizan un aumento de datos implícito adicional. [Ronneberger et al., 2015]

En el caso tridimensional, se expande la deformación elástica a la tercera dimensión, logrando un desplazamiento y deformación al mismo tiempo. Este proceso se puede ver ilustrado en la figura 11. El cual es utilizado para el aumento de datos en el área médica, ya que permite deformaciones lo suficientemente sutiles para que sean realistas, y lo suficientemente abruptas para que la imagen resultante sea diferente de la original.

2.3.2. 3D UNET

Considerando la problemática de este proyecto, se busca expandir la arquitectura UNet para ser utilizada con datos tridimensionales para segmentación 3D (figura 12). La expansión fue introducida por Çiçek et al., para la segmentación volumétrica a partir de anotaciones dispersas, considerando 2 casos [Çiçek et al., 2016]. En primer lugar, una configuración semi-automatizada, el usuario anota algunos cortes en el volumen que se va a segmentar. La red aprende de estas anotaciones dispersas y proporciona una segmentación 3D densa. El segundo caso, es una configuración completamente automatizada, donde suponen que existe un conjunto de entrenamiento representativo con anotaciones dispersas. Siendo entrenada

Deformación Elástica en 3 dimensiones

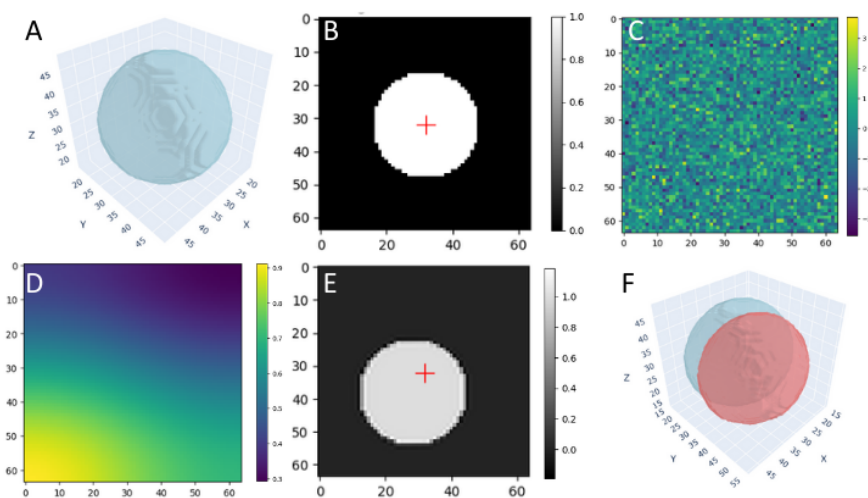


Figura 11: Paso a paso sobre la generación de deformación elástica en múltiples dimensiones. A) Volumen original de forma esférica. B) Un corte de la esfera para visualizar en dos dimensiones. C) Plano de distorsiones randomizadas. D) Deformaciones suavizadas con un filtro gaussiano. E) Corte del volumen original donde se puede ver el desplazamiento del volumen respecto al centro de la imagen. F) Comparación entre la esfera original y la deformada. Fuente: Elaboración Propia.

en este dataset, la red segmenta de forma densa nuevas imágenes volumétricas.

La propuesta de Çiçek et al. reemplaza todas las operaciones 2D de la UNet original por operaciones 3D. La implementación realiza deformaciones elásticas sobre la marcha para aumentación de datos durante el entrenamiento. La red es entrenada de inicio a fin, sin parámetros pre-entrenados. Prueban la propuesta para la segmentación de riñón *Xenopus*, logrando buenos resultados para los dos casos propuestos. Esta arquitectura se ha utilizado como punto de partida para la propuesta de solución en este trabajo, considerando la cercanía a la problemática con la que se está tratando y el uso de imágenes por RM para su implementación.

En las capas convolucionales 3D se utiliza *batch normalization* para normalizar las activaciones de cada *mini-batch* de datos. Esto asegura que cada mini-batch tenga una media de 0 y una varianza de 1, estabilizando el entrenamiento, evitando el problema de desplazamiento interno de covariables y permitiendo el uso de tasas de aprendizaje más altas [Ioffe y Szegedy, 2015].

La misma 3D UNet tiene diferentes expansiones, una propuesta por [Oh et al., 2023], incluye conexiones residuales en el modelo. Estas fueron introducidas en [He et al., 2016] y se puede ver en la Figura 13. Las conexiones residuales permiten que los datos fluyan directamente a través de la red, facilitando el entrenamiento de redes muy profundas, al actuar como un método de regularización, previenen el sobre ajuste de los datos. De forma concreta, la

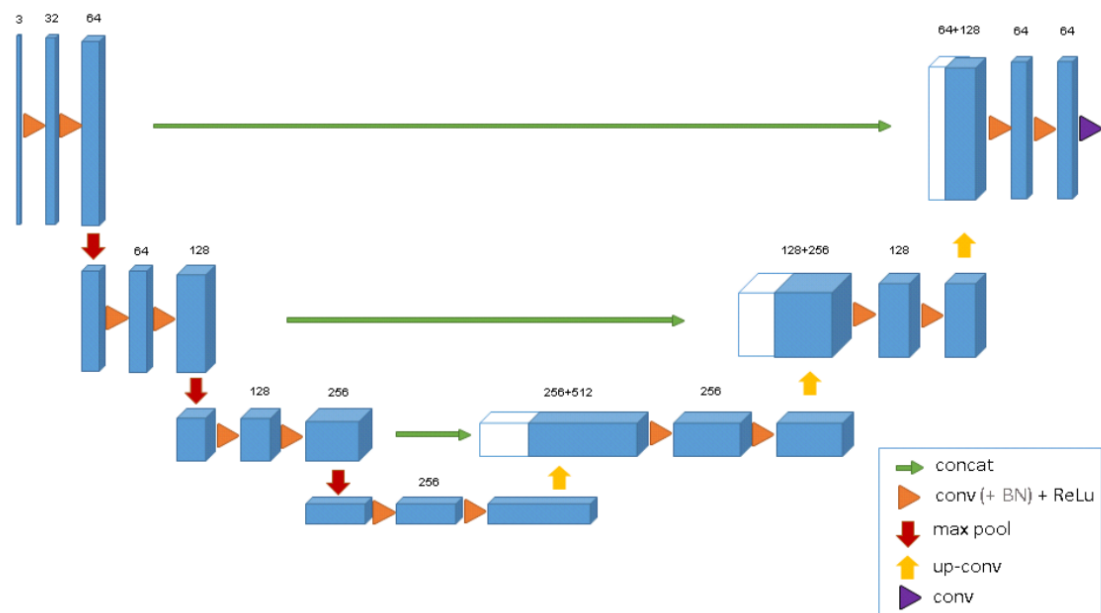


Figura 12: La arquitectura 3D UNet. Cajas azules representan mapas de características. El número de canales es denotado encima de cada mapa de características.

Fuente: [Çiçek et al., 2016]

conexión residual lo único que hace es implementar la función identidad sobre los datos de entrada, para luego concatenarlos con el resultado de una operación intermedia, como se puede ver en la Figura 13.

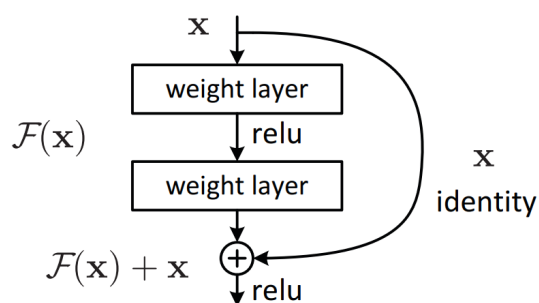


Figura 13: Aprendizaje residual: un bloque de construcción. Fuente: Deep Residual Learning for Image Recognition [He et al., 2016]

La implementación realizada para el modelo 3D UNet, se puede ver en la Figura 14. Donde además se incluyó la mencionada *batch normalization* antes de utilizar la unidad rectificadora lineal.

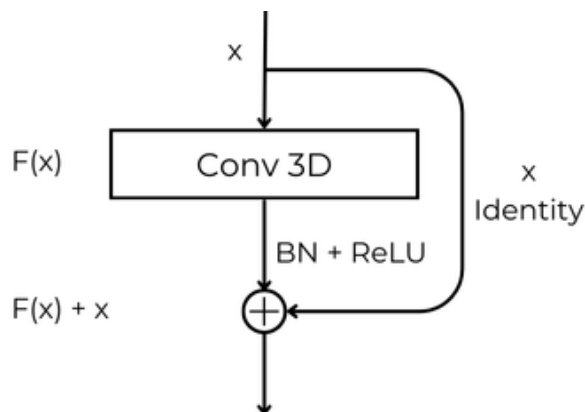


Figura 14: Bloque residual con una convolución en 3D, *batch normalization* y ReLU.

Fuente: Elaboración Propia

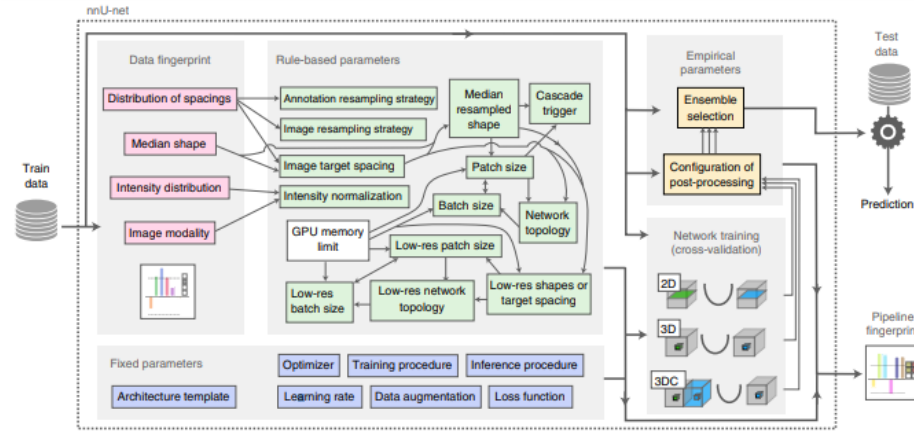
2.3.3. NNUNET

Considerando la cantidad de conocimiento experto requerido para trabajar con imágenes médicas, se propone el *framework* nnUNet [Isensee *et al.*, 2021], donde se deja en código abierto una herramienta de desarrollo de inicio a fin para tareas en el área biomédica. Este utiliza métodos de *Automated Machine Learning* (AutoML), para que no se requiera un extensivo conocimiento técnico para poder generar segmentaciones en imágenes médicas de múltiples fuentes y diversas tareas específicas.

El flujo de procesamiento propuesto por nnUNet se muestra detalladamente en la Figura 15. Este *framework* opera de manera *out-of-the-box*, lo que significa que no es necesario realizar ajustes manuales; el usuario solo debe definir los caminos adecuados a los datos, y asegurarse de que el dataset esté en el formato del *Medical Segmentation Decathlon* (MSD). Si bien este enfoque reduce significativamente la carga de trabajo del usuario, implica una mayor demanda en términos de tiempo y recursos computacionales. Esto se debe a que nnUNet no solo selecciona automáticamente las transformaciones y preprocesamientos más adecuados, sino que también determina la arquitectura de la red, el optimizador, y la función de pérdida a utilizar.

Además, nnUNet evalúa varias configuraciones a través de una validación cruzada quíntuple, permitiendo así una selección empírica del modelo óptimo y del posible posprocesamiento necesario. El sistema se basa en una “huella digital del conjunto de datos” para extraer las propiedades relevantes del mismo (como se muestra en rosa en la figura 15), y aplica un conjunto de reglas heurísticas para deducir los parámetros basados en reglas (en verde) que dependen de las características de los datos. Estos parámetros se complementan con parámetros fijos (en azul), los cuales están predefinidos y no requieren adaptación. Finalmente, se seleccionan y combinan los mejores modelos entrenados, garantizando un rendimiento óptimo en la tarea de segmentación.

MÉTODO DE SEGMENTACIÓN AUTOMÁTICA DE HÍGADO EN IMÁGENES DE RESONANCIA MAGNÉTICA UTILIZANDO UN MODELO DE APRENDIZAJE AUTOMÁTICO



Design choice	Required input	Automated (fixed, rule-based or empirical) configuration derived by distilling expert knowledge (more details in online methods)			
Learning rate	–	Poly learning rate schedule (initial, 0.01)	Image target spacing	Distribution of spacings	If anisotropic, lowest resolution axis tenth percentile, other axes median. Otherwise, median spacing for each axis. (computed based on spacings found in training cases)
Loss function	–	Dice and cross-entropy	Network topology, patch size, batch size	Median resampled shape, target spacing, GPU memory limit	Initialize the patch size to median image shape and iteratively reduce it while adapting the network topology accordingly until the network can be trained with a batch size of at least 2 given GPU memory constraints. For details see online methods.
Architecture template	–	Encoder-decoder with skip-connection (U-Net-like) and instance normalization, leaky ReLU, deep supervision (topology-adapted in inferred parameters)	Trigger of 3D U-Net cascade	Median resampled image size, patch size	Yes, if patch size of the 3D full resolution U-Net covers less than 12.5% of the median resampled image shape
Optimizer	–	SGD with Nesterov momentum ($\mu = 0.99$)	Configuration of low-resolution 3D U-Net	Low-res target spacing or image shapes, GPU memory limit	Iteratively increase target spacing while reconfiguring patch size, network topology and batch size (as described above) until the configured patch size covers 25% of the median image shape. For details, see online methods.
Data augmentation	–	Rotations, scaling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma correction and mirroring	Configuration of post-processing	Full set of training data and annotations	Treating all foreground classes as one; does all-but-largest-component-suppression increase cross-validation performance? Yes, apply; reiterate for individual classes No, do not apply; reiterate for individual foreground classes
Training procedure	–	1,000 epochs $\times$ 250 minibatches, foreground oversampling	Ensemble selection	Full set of training data and annotations	From 2D U-Net, 3D U-Net or 3D cascade, choose the best model (or combination of two) according to cross-validation performance
Inference procedure	–	Sliding window with half-patch size overlap, Gaussian patch center weighting			
Intensity normalization	Modality, intensity distribution	If CT, global dataset percentile clipping & z score with global foreground mean and s.d. Otherwise, z score with per image mean and s.d.			
Image resampling strategy	Distribution of spacings	If anisotropic, in-plane with third-order spline, out-of-plane with nearest neighbor Otherwise, third-order spline			
Annotation resampling strategy	Distribution of spacings	Convert to one-hot encoding $\rightarrow$ If anisotropic, in-plane with linear interpolation, out-of-plane with nearest neighbor Otherwise, linear interpolation			

Figura 15: Método automatizado de configuración para segmentación basada en *deep learning* de imágenes biomédicas. Dada una nueva tarea de segmentación, las propiedades del conjunto de datos se extraen en forma de “huella digital del conjunto de datos” (rosa). Un conjunto de reglas heurísticas modela las interdependencias de los parámetros (mostradas como flechas finas) y opera sobre esta huella digital para deducir los “parámetros basados en reglas” (verde) de la línea de producción que dependen de los datos. Estos se complementan con “parámetros fijos” (azul), que están predefinidos y no requieren adaptación. Se entrenan hasta tres configuraciones en una validación cruzada quíntuple. Por último, nnU-Net realiza automáticamente la selección empírica del conjunto óptimo de estos modelos y determina si es necesario el posprocesamiento (“parámetros empíricos”, amarillo). La tabla de la parte inferior muestra los valores explícitos, así como las formulaciones resumidas de las reglas de todos los parámetros configurados.

Fuente: [Isensee *et al.*, 2021]

nnUNet tiene ciertos parámetros fijos, como la función de pérdida, que utiliza una suma de

Dice loss y cross-entropy. Como optimizador utiliza *Stochastic Gradient Descent* (SGD) con momentum de Nesterov. La tasa de aprendizaje inicial está definida como 0.001 y se ajusta durante el entrenamiento utilizando un esquema de reducción de tasa de aprendizaje basado en la validación.

En el estudio demuestran que nnUNet supera o iguala el rendimiento de otros métodos de segmentación en 33 de 53 conjuntos de datos diferentes, incluidos los del *Medical Segmentation Decathlon*. Los resultados mostraron que nnUNet no solo puede adaptarse a una amplia variedad de tareas de segmentación sin ajustes manuales, sino que también logra un rendimiento de vanguardia, superando incluso a métodos personalizados y optimizados manualmente para tareas específicas. Esto subraya la capacidad del *framework* para ofrecer soluciones robustas y precisas en el campo de la segmentación biomédica.

2.4. EVALUACIÓN DE LOS MODELOS

Ahora se explicará cuáles son las métricas a utilizar para evaluar los modelos probados. En primer lugar, es necesario tener en cuenta que pese a estar trabajando con volúmenes, las métricas a utilizar para segmentación 2D y 3D son las mismas, solo se hacen ajustes a la implementación.

2.4.1. DICE SCORE

A base de una extensa revisión del estado del arte en segmentación para imágenes médicas, la métrica más utilizada es el coeficiente de Dice. Esta métrica, ampliamente utilizada, calcula la relación entre la intersección y la unión de dos conjuntos. La fórmula del Dice Score es la siguiente:

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (3)$$

donde,

- A es el conjunto de voxels predichos como pertenecientes a la estructura de interés.
- B es el conjunto de voxels reales pertenecientes a la estructura de interés.
- $A \cap B$ es el número de voxels correctamente segmentados (intersección entre A y B).
- $|A|$ y $|B|$ son el número de voxels en las segmentaciones predicha y real, respectivamente.

Considerando que un *voxel* o elemento volumétrico es la unidad básica de una imagen tridimensional, similar al *pixel* en una imagen bidimensional. Cada *voxel* representa un valor en el espacio tridimensional, asociado a una coordenada específica en un volumen de datos.

El valor del Dice Score oscila entre 0 y 1, donde 0 indica que no hay superposición entre los conjuntos comparados y 1 indica un calce completo. Esto es especialmente útil en el caso de segmentación con clases desbalanceadas, ya que pondera la coincidencia relativa en lugar de la coincidencia absoluta.

2.4.2. INTERSECTION OVER UNION IOU

Como métrica secundaria se utilizará la métrica IoU, *Intersection over Union*. IoU es una medida ampliamente utilizada en modelos de segmentación [Rezatofighi *et al.*, 2019]. Esta métrica compara el área objetivo con el área estimada por el modelo, utilizando la formula,

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

En adición a la notación presentada anteriormente, $A \cup B$ es el número de total de voxels que están en la segmentación predicha o en la real (unión de A y B). Al igual que en Dice, el valor de IoU varía entre 0 y 1, 0 indicando que no hay superposición y 1 la superposición perfecta. IoU es más estricto que Dice, ya que penaliza más las discrepancias entre las segmentaciones predicha y real.

Ambas métricas son fundamentales para la evaluación y comparación de modelos de segmentación 3D, y se diferencian en,

- *Dice Score*: Es más sensible a pequeñas áreas de coincidencia y es útil para conjuntos de datos desbalanceados.
- *IoU*: Es una medida más estricta y sensible a datos balanceados.

En la práctica, la elección entre *Dice Score* y IoU puede depender del caso de uso específico y de las características del conjunto de datos. Es común utilizar ambas métricas en conjunto para obtener una evaluación más completa del rendimiento del modelo. En este caso se considera como métrica primaria de comparación *Dice score* y de forma secundaria se considera IoU.

2.4.3. ESTIMACIÓN DE COMPONENTES CONEXAS

Para este problema en específico, donde se busca realizar la segmentación de un volumen, se evalúa el desempeño de un modelo revisando cuántas componentes conexas (CC) tiene la predicción. Mientras más cercano a 1 sea el número de componentes, mejor será el desempeño del modelo a evaluar.

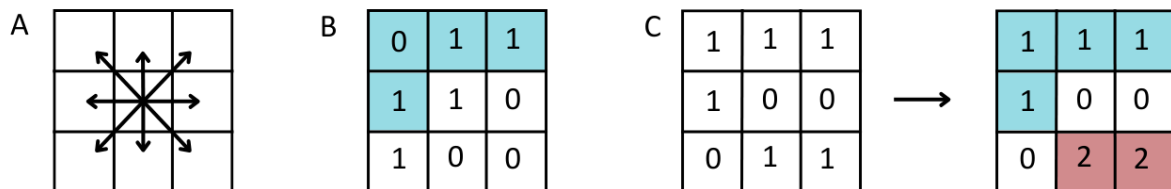


Figura 16: Se ilustra el proceso de identificación de componentes conexas en una matriz 2D. A) Se revisan los valores de todos los vecinos del píxel a etiquetar. B) Ejemplo de un componente conexo de tamaño 4. C) Proceso de etiquetado en una matriz con más de un componente conexo.

Fuente: Elaboración Propia

Para realizar este cálculo, se considera que cada píxel tiene el valor de 0 o 1 (esto se logra mediante un umbral), y si dos píxeles tienen el valor 1 y son vecinos, se consideran conectados (figura 16). Se ilustra en 2D, y la extensión a 3D se realiza considerando ahora los 26 píxeles al rededor del pixel central. Es importante tener en cuenta que las diferentes componentes conexas se diferencian al cambiar el valor de los píxeles, como se muestra en la sección C de la figura 16, donde los componentes del segundo grupo se cambian a un valor de 2 [Fiorio y Gustedt, 1996, Wu *et al.*, 2005].

CAPÍTULO 3

PROPUESTA DE SOLUCIÓN

Para resolver los objetivos de este trabajo, existen dos pilares principales de desarrollo. En primer lugar, el tratamiento de los datos, su forma y transformación. Y en segundo lugar, el desarrollo del algoritmo en sí.

3.1. DUKE LIVER DATASET

Como se mencionó previamente, el *Duke Liver Dataset* (DLDS) fue liberado en 2023 por la Universidad de Duke, con el objetivo de impulsar el desarrollo de soluciones automatizadas en el campo de la imagenología médica. Este conjunto de datos se centra en la segmentación del hígado y contiene imágenes de RM acompañadas de anotaciones detalladas [Macdonald *et al.*, 2023].

El DLDS está diseñado para apoyar la investigación y el desarrollo de algoritmos de segmentación, especialmente en el contexto de enfermedades hepáticas. Con diferentes secuencias de MRI, este conjunto de datos ofrece una base sólida para entrenar y validar modelos de aprendizaje profundo. La precisión y la calidad de las anotaciones proporcionadas permiten mejorar los métodos de diagnóstico y tratamiento de enfermedades hepáticas a través de técnicas avanzadas de procesamiento de imágenes.

3.1.1. ANÁLISIS ESTADÍSTICO DEL DATASET

A continuación, se presenta un análisis estadístico detallado del DLDS, enfocado en la distribución de las imágenes y segmentaciones, así como en las características clave del conjunto de datos. Este análisis permitirá entender mejor la variabilidad y las propiedades intrínsecas de los datos, lo que es esencial para la validación y optimización de modelos de segmentación.

En primer lugar, el dataset se estructura con base en *DLDS_ID*, identificadores únicos y anonimizados que se utilizan para cada paciente. Se cuenta con 105 pacientes y 2146 series de imágenes por RM, de las cuales solo 310 están etiquetadas. Para cada paciente se tienen entre 1 y 5 exámenes disponibles con diferentes secuencias tanto en exámenes etiquetados, como no etiquetados(2). Para el desarrollo de modelos se utilizarán únicamente las secuencias con representatividad considerable en el conjunto de datos etiquetados, dejando de lado para experimentación y prueba de los modelos los 6 exámenes con secuencias con baja representación.

Tabla 2: Secuencias presentes en el DLDS, la cantidad de apariciones y cuantas de ellas se encuentran etiquetadas. Resaltadas las secuencias que se utilizarán para experimentación.

Fuente: Elaboración Propia

Etiqueta	Descripción de la Serie	Cantidad	Cantidad de Datos Etiquetados	% Etiquetados
B	T1w Axial Precontrast Fat Suppressed (dynpre)	220	85	27.4
K	T1w Portal Venous (dynportal)	106	83	26.8
G	Axial In Phase (t1nfs)	115	82	26.5
H	Axial Opposed Phase (opposed)	113	54	17.4
C	T1w Mid Arterial	192	3	1.0
E	T1w Axial Late Dynamic	134	2	0.6
O	T1w Early Arterial	62	1	0.3
M	Axial ADC	99	0	0.0
Q	T1w Late Arterial	41	0	0.0
P	T1w Axial Transitional/Hepatocyte	46	0	0.0
N	T1w Coronal Late Dynamic	87	0	0.0
A	T2w Axial (ssfse)	318	0	0.0
L	Localizers	104	0	0.0
I	Axial DWI	113	0	0.0
F	Anything else	127	0	0.0
D	MRCP	159	0	0.0
J	T2w Coronal	110	0	0.0
Total		2146	310	100.0

PACIENTES

El DLDS incluye a 105 pacientes, de los cuales 76 son hombres y 29 son mujeres, abarcando un rango amplio de edades entre 30 y 80 años. Solo se cuentan con segmentaciones para 95 pacientes, debido a que 10 no tenían exámenes con las secuencias mayoritariamente representadas. Esta diversidad en cuanto a género y edad asegura que los modelos entrenados con este dataset sean más generalizables y puedan aplicarse a una población heterogénea (figura 17).

Además, la calidad del DLDS se ve potenciada por la inclusión de pacientes en distintas etapas de enfermedad hepática. De los 105 pacientes, 87 presentaron hallazgos en imágenes sugestivos de cirrosis, lo que ofrece una gama de morfologías hepáticas que varían desde hígados saludables hasta aquellos significativamente afectados por la enfermedad. Esta variabilidad en las presentaciones patológicas permite que el dataset capture las diferencias sutiles y evidentes en la estructura del hígado, lo que es crucial para el desarrollo y validación de modelos de inteligencia artificial en el campo de la medicina.

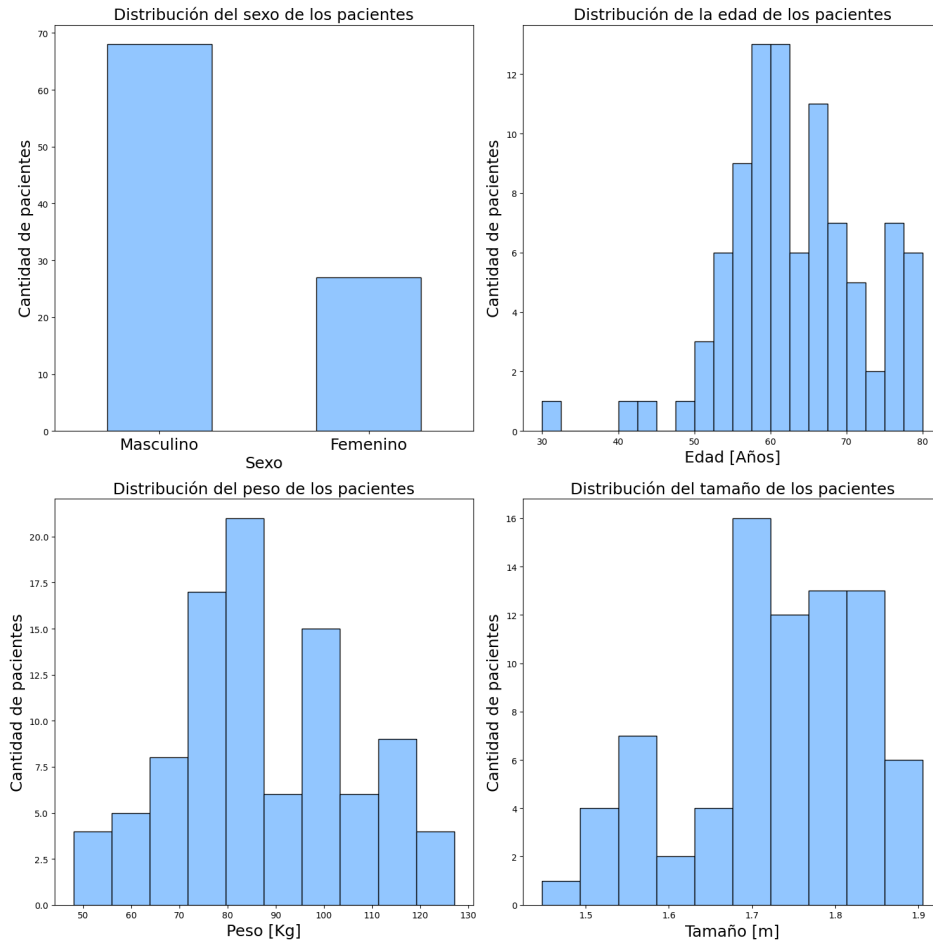


Figura 17: Distribuciones demográficas y antropométricas de los pacientes: Comparación de sexo, edad, peso y tamaño en la población del estudio.

Fuente: Elaboración propia.

3.1.2. ESTANDARIZACIÓN DATASET

Para asegurar la consistencia y comparabilidad de los resultados, se procedió a la estandarización del DLDS. Este proceso incluyó la normalización de las imágenes y la estandarización de la cantidad de cortes por imagen. La estandarización es un paso fundamental para reducir la variabilidad en los datos y mejorar la robustez de los modelos de segmentación que se entrenan y validan con este conjunto de datos.

Debido a que en la base de datos (DLDS) se tienen imágenes de distintos equipos (94 Siemens y 9 GE) a distinto campo magnético (54 a 1.5T y 51 a 3T), es que las imágenes de base contaban con una cantidad variable de cortes, y tamaños de los mismos. Por esta razón se decidió utilizar un método de interpolación lineal para ajustar todas las imágenes al tamaño 64 x 64 x 64, lo que significaría que ahora cada examen está compuesto por 64 cortes de tamaño 64 x 64. Para las máscaras se utilizó el método *Nearest Neighbours* (o vecinos cercanos

en español) para dejarlas en el mismo formato (Figura 18).

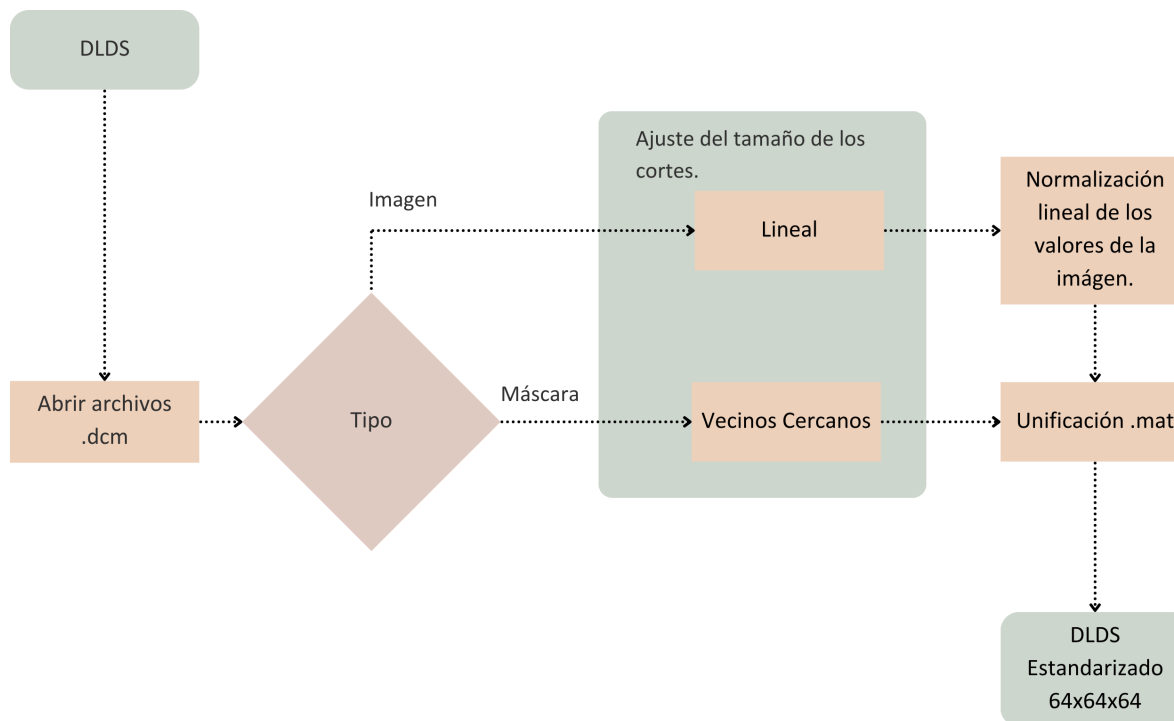


Figura 18: Transformaciones realizadas a las imágenes pertenecientes al DLDS.
Fuente: Elaboración Propia.

3.1.3. ENTRENAMIENTO, VALIDACIÓN Y PRUEBA

La división de los datos se realizó respecto a los pacientes, donde se dejó un 70 % para entrenamiento, 20 % para validación y 10 % para pruebas. Esta distribución se realizó considerando solo las 4 secuencias con mayor representación en el dataset y en total se obtuvieron 207 imágenes de entrenamiento, 62 de validación y 35 de prueba.

Pese a que en el contexto de imágenes médicas, contar con 207 ejemplos de entrenamiento es un número considerable, para modelos de aprendizaje profundo, esto es bastante pequeño. Por esta razón, a las imágenes de entrenamiento se les aplicó aumento de datos durante el entrenamiento (*on-the-fly*) para disminuir las probabilidades de sobre ajuste a los datos. El método que se utilizó fue de deformación elástica, comúnmente utilizado en la literatura, ya que permite cambios sutiles a los volúmenes que aportan nuevas morfologías dentro de los parámetros del dominio (figura 11).

respecto a la toma del examen podría tener en el rendimiento del modelo, comparado con el entrenamiento convencional basado únicamente en imágenes.

La información de las segmentaciones se integró utilizando un Perceptrón Multicapa para procesar los metadatos y luego concatenar la representación de estos con la imagen codificada en el cuello de botella de la 3D Res UNet. Los detalles de la arquitectura se pueden ver ilustrados en la figura 20

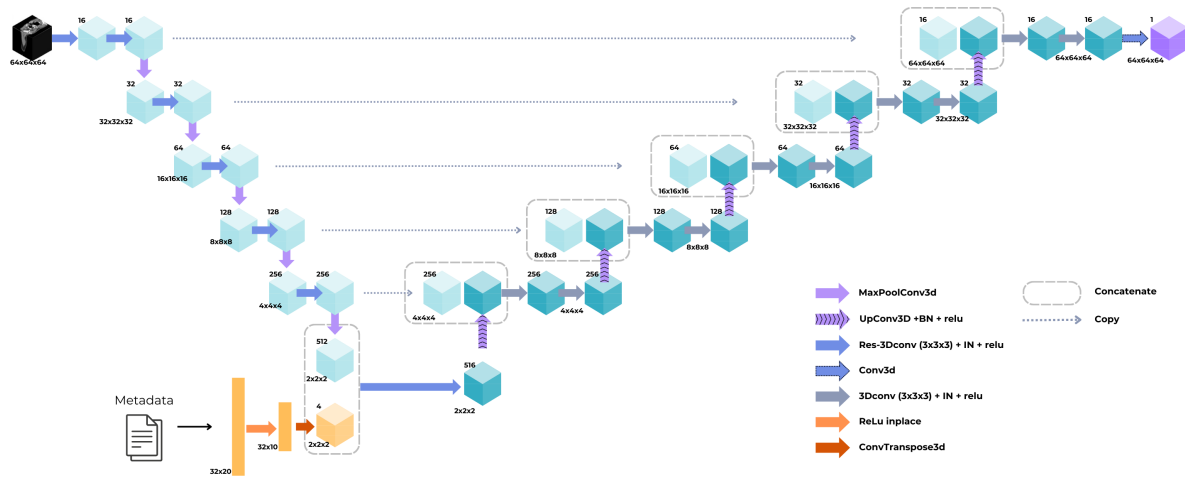


Figura 20: Arquitectura de la red 3D Res UNetA. Se pueden ver los procesos de *encoder* y *decoder*, con operaciones convolucionales residuales durante la codificación y la reconstrucción de imágenes segmentadas de múltiples objetivos mediante mapas de características deconvolucionales durante la decodificación. También se puede ver el Perceptrón Multicapa utilizado para codificar los Metadatos. Las conexiones salteadas se muestran en cada nivel de resolución.

Fuente: Elaboración Propia

3.2.2. MODELOS PRE-ENTRENADOS

Para comparar los modelos entrenados con el DLDS, se obtuvieron modelos disponibles en internet para evaluar con el conjunto de pruebas generado.

NNUNET

En el estudio donde se presenta el *framework* de nnUNet [Isensee *et al.*, 2021], se provee un modelo preentrenado para la detección de hígado y tumores utilizando exámenes abdominales tipo CT. Este conjunto de datos proviene del *Medical Segmentation Decathlon* [Antonelli *et al.*, 2022] y se consideró para la comparación con el DLDS, ya que la mayoría del estado del arte genera soluciones para CT y no para MRI.

La solución propuesta por nnUNet incluye 4 modelos, 2D, 3D *cascade fullres*, 3D *fullres* y 3D

lowres, donde para este proyecto se consideró únicamente la red 3D *fullres* al contar con mejores resultados con su propio *dataset* de entrenamiento.

VOLUMETRY-NET

En el estudio por [Gross *et al.*, 2024] se explora una solución automatizada para la segmentación del hígado en imágenes de RM, destacando que este es uno de los pocos trabajos que cuenta con su código disponible públicamente y que fue aplicado específicamente a un conjunto de datos de RM. El modelo implementado es una UNet 3D de la librería MONAI⁴, diseñada específicamente para trabajo con imágenes médicas. La arquitectura de la UNet utilizada tiene dimensiones 3D con una entrada de un solo canal y produce dos canales de salida, correspondiente a las clases de segmentación. Las capas convolucionales están configuradas con 16, 32, 64 y 128 filtros, y se emplea normalización por lotes (*BatchNorm*) junto con unidades residuales para mejorar la capacidad del modelo para generalizar y adaptarse a las variaciones en los datos. La red también incluye un *dropout* configurable para prevenir el sobreajuste.

3.3. COMPARACIÓN DE MODELOS

Para evaluar el desempeño de los modelos, se llevará a cabo una revisión exhaustiva que incluirá tanto análisis cuantitativos como cualitativos. En primer lugar, se realizará una comparación cuantitativa utilizando las métricas Dice score e IoU. Estas métricas son esenciales para medir la precisión y la superposición de las segmentaciones generadas por los modelos en relación con las segmentaciones verdaderas. Además, se analizará la cantidad de componentes conexos en las predicciones realizadas, dado que, en este contexto, se busca identificar una única región de interés: el hígado. La identificación precisa de esta región es crucial, por lo que evaluar la cohesión y la integridad de la segmentación será fundamental.

En segundo lugar, se llevará a cabo una comparación cualitativa de las segmentaciones obtenidas. Este análisis se centrará en examinar ejemplos representativos de las predicciones, categorizándolas en función de su desempeño en términos de la métrica Dice. Se seleccionarán las predicciones con el mejor y el peor Dice score, así como las predicciones cuyo Dice score sea la mediana del conjunto de pruebas para cada modelo. Esta evaluación permitirá una apreciación más detallada de la calidad visual y la precisión de las segmentaciones en casos específicos, proporcionando una comprensión más completa de cómo los modelos se comportan en situaciones concretas y diversas.

⁴Documentación librería MONAI

CAPÍTULO 4

VALIDACIÓN DE LA SOLUCIÓN

Los modelos implementados y entrenados para esta solución (3D UNet y 3D Res UNet) se validaron mediante su comparación con dos modelos del estado del arte, nnUNet y Volumetry Net, ambos con modelos preentrenados disponibles públicamente. Esta validación permitió evaluar la eficacia de los modelos propuestos en relación con estándares reconocidos en la comunidad.

4.1. RESULTADOS SOBRE CONJUNTO DE PRUEBAS

A continuación, se presentan de manera detallada los resultados obtenidos para los diferentes modelos evaluados en este estudio. En la Tabla 3, se realiza una comparación exhaustiva de las métricas de desempeño de los modelos 3D UNet y 3D Res UNet frente a dos de los modelos considerados como el estado del arte en la segmentación médica: nnUNet y Volumetry Net. Las métricas consideradas en esta comparación incluyen la puntuación Dice, IoU, y la cantidad de componentes conexas, cada una de las cuales proporciona una perspectiva única sobre la calidad de las segmentaciones generadas por los modelos. Estos resultados son fundamentales para analizar en profundidad la efectividad y precisión de los modelos propuestos en comparación con los modelos preentrenados disponibles públicamente.

Tabla 3: Promedio $\pm$ desviación estándar de las métricas relevantes, donde CC significa Componentes Conexas.

Fuente: Elaboración propia.

	Puntaje Dice	IoU	Cantidad CC	Tiempo de Inferencia [s]
3D UNet	0.8822 $\pm$ 0.0673	0.7951 $\pm$ 0.1014	2.2857 $\pm$ 1.4667	0.0069 $\pm$ 0.0042
3D Res UNet	0.8785 $\pm$ 0.0717	0.7900 $\pm$ 0.1073	2.6286 $\pm$ 1.9110	0.0034 $\pm$ 0.0002
Volumetry Net	0.7975 $\pm$ 0.1555	0.6866 $\pm$ 0.1882	12.0286 $\pm$ 7.4300	3.0698 $\pm$ 3.8900
nnUNet	0.2471 $\pm$ 0.1875	0.1545 $\pm$ 0.1305	11.9714 $\pm$ 14.565	78.3825 $\pm$ 9.9043
3D Res UNetA	0.8410 $\pm$ 0.0979	0.7363 $\pm$ 0.1304	2.0000 $\pm$ 0.0000	0.0197 $\pm$ 0.0164

Con base en los resultados obtenidos sobre el conjunto de pruebas (Tabla 3, Figura 21), se destaca que la solución base propuesta en este proyecto, representada por los modelos 3D UNet y 3D Res UNet, genera las mejores predicciones promedio en el conjunto de prueba, lo que refleja un alto nivel de desempeño. Este rendimiento superior también se puede apreciar de forma gráfica en la Figura 21, donde se observa que las medianas de ambos modelos implementados se sitúan por encima de la marca de puntaje Dice de 0.9. Este resultado es particularmente significativo, ya que representa un excelente rendimiento en comparación

con el estado del arte, lo que subraya la robustez y precisión de las soluciones desarrolladas en este trabajo.

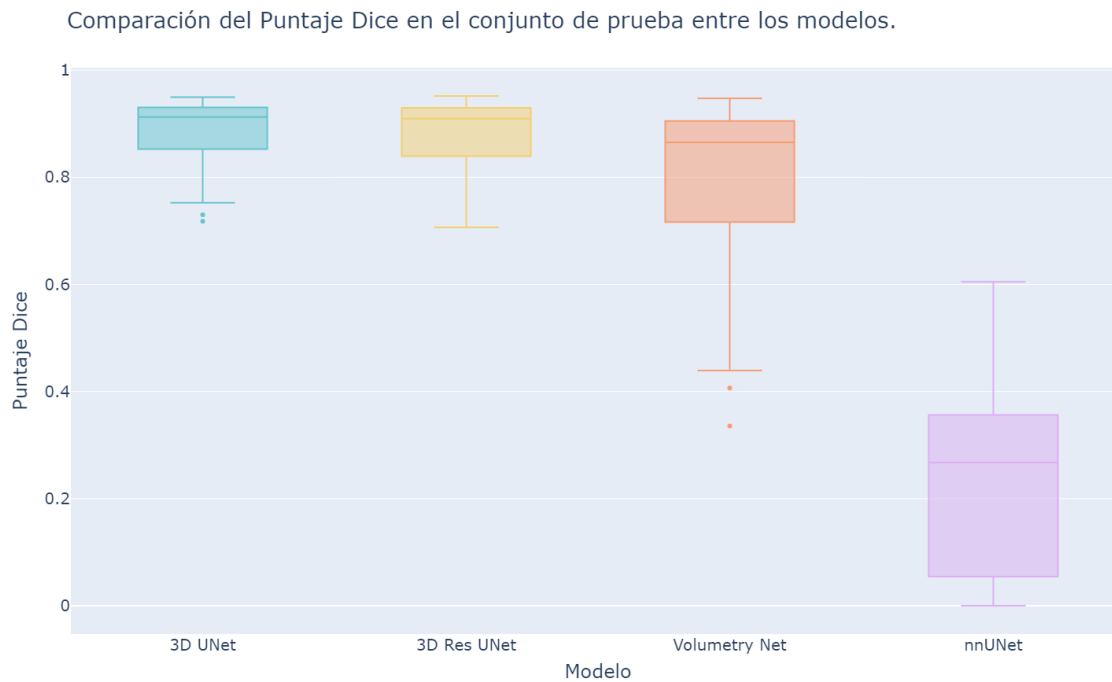


Figura 21: Distribución de los puntajes Dice en formato de *boxplot* para los modelos evaluados.

Fuente: Elaboración propia.

El desempeño de nnUNet fue significativamente inferior en comparación con otros modelos como 3D UNet y 3D Res UNet. Esto se reflejó en métricas tanto en el Puntaje Dice como en IoU, donde obtuvo valores promedio de 0.2471 y 0.1545, respectivamente, muy por debajo de los resultados de los otros modelos, esto se evidencia de forma muy clara en la Figura 21. Además, el elevado número de Componentes Conexas generadas por nnUNet indica que las segmentaciones eran fragmentadas y poco coherentes, en contraste con las segmentaciones más precisas y compactas logradas por los otros modelos.

Este bajo rendimiento puede atribuirse a que nnUNet fue entrenado con exámenes de tomografía computarizada (CT) y probado con imágenes de RM. Las diferencias entre estas modalidades, como el contraste y la textura, dificultaron la capacidad del modelo para generalizar correctamente a las imágenes de MRI, lo que resultó en segmentaciones inadecuadas. Para mejorar su desempeño, sería necesario entrenar nnUNet directamente en imágenes de MRI o emplear técnicas que faciliten la adaptación del modelo entre modalidades. Esto no fue posible en este proyecto, ya que es un *framework* bastante extenso y tanto computacionalmente como temporalmente requiere muchos recursos para ser ejecutado de inicio a fin.

Tabla 4: Ejemplos de las mejores predicciones obtenidas por cada uno de los modelos evaluados.

Fuente: Elaboración propia.

	Puntaje Dice	IoU	Cantidad de CC	Tiempo de Inferencia [s]	DLDS_ID	Serie	Secuencia
3D UNet	0.9497	0.9043	1	0.0039	0050	5	B
3D Res UNet	0.9518	0.9081	1	0.0034	0045	14	K
Volumetry Net	0.9473	0.8999	12	0.8129	0050	9	K
nnUnet	0.6049	0.4335	10	-	0026	4	H

En las tablas 4, 5 y 6, se presentan una serie de ejemplos que ilustran de manera clara y precisa los casos de mejor desempeño, peor desempeño, y el desempeño con puntaje Dice en la mediana para cada uno de los modelos evaluados. Al observar estos resultados, es relevante destacar que el modelo 3D UNet sobresale de manera consistente, obteniendo el mejor resultado en cada una de las tablas mencionadas. Esta consistencia en el desempeño subraya la robustez y eficacia del modelo 3D UNet en una variedad de escenarios, consolidándolo como una opción destacada dentro de las soluciones propuestas en este estudio.

Se puede notar que ciertas combinaciones de DLDS_ID-Serie se repiten en las tablas, lo que nos indica que hay ciertos exámenes que tienen mejor predisposición a la segmentación que otros. Por ejemplo, el DLDS 0050 es el examen con mejor desempeño tanto para 3D UNet como para Volumetry Net. De forma contraria, el examen 0094-0094 (DLDS_ID-Serie), es él con peor desempeño en tres de los cuatro modelos evaluados, por lo tanto, existe una alta probabilidad de que este examen presente un mayor desafío para la tarea de segmentación con los modelos seleccionados para probar.

Tabla 5: Ejemplos de predicciones con puntajes en la mediana para cada modelo, representando el desempeño promedio dentro de las muestras evaluadas.

Fuente: Elaboración propia.

	Puntaje Dice	IoU	Cantidad de CC	Tiempo de Inferencia [s]	DLDS_ID	Serie	Secuencia
3D UNet	0.9126	0.8393	1	0.0140	0088	29	B
3D Res UNet	0.9094	0.8339	3	0.0035	0088	29	B
Volumetry Net	0.8652	0.7624	15	1.1225	0004	4	H
nnUnet	0.2674	0.1543	57	-	0010	3	G

Tabla 6: Ejemplos de las peores predicciones realizadas por cada modelo, evidenciando los valores más bajos de puntaje Dice e IoU.

Fuente: Elaboración propia.

	Puntaje Dice	IoU	Cantidad de CC	Tiempo de Inferencia [s]	DLDS_ID	Serie	Secuencia
3D UNet	0.7180	0.5600	3	0.0073	0094	29	G
3D Res UNet	0.7065	0.5461	5	0.0035	0094	29	G
Volumetry Net	0.3356	0.2016	16	12.9767	0094	29	G
nnUNet	0.0000	0.0000	3	-	0010	4	H

En cuanto a los ejemplos con mejor dice, se puede ver el calce casi perfecto entre etiqueta y predicción para 3D Unet, 3D Res Unet y Volumetry Net (figura 22). Mientras que la predicción de nnUNet está en la posición correcta, sin embargo, no reconoce el tamaño adecuado.

De forma cualitativa se puede ver la diferencia drástica en calidad de las predicciones realizadas, especialmente si se comparan los ejemplos con peor puntaje Dice (figura 23). Tanto 3D UNet como 3D Res UNet, mantienen un estándar mínimo de aproximación a la segmentación, sin embargo, tanto Volumetry Net como nnUNet realizan predicciones extremadamente erróneas.

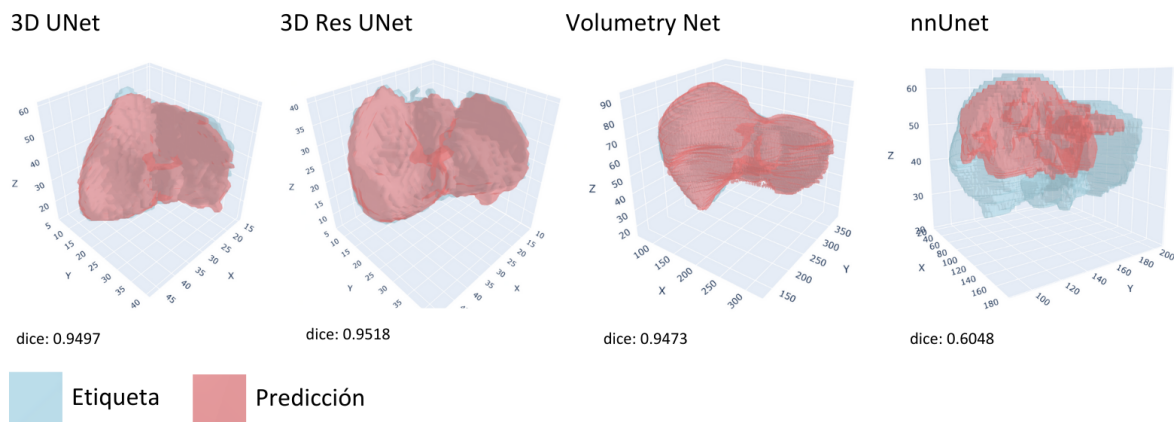


Figura 22: Resultados cualitativos de las predicciones con **mejor** desempeño para cada modelo, en el conjunto de pruebas

Fuente: Elaboración Propia

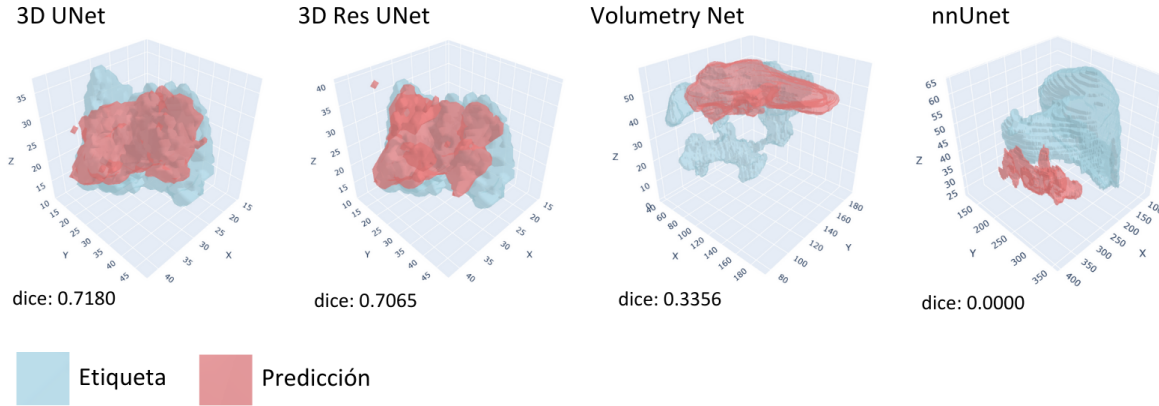


Figura 23: Resultados cualitativos de las predicciones con **peor** desempeño para cada modelo, en el conjunto de pruebas

Fuente: Elaboración Propia

4.2. RESULTADOS SOBRE SECUENCIAS NO VISTAS POR EL MODELO

Se evaluaron ejemplos de prueba que fueron excluidos del conjunto de datos principal debido a su escasa representación. Aunque no formaron parte del análisis principal, estos ejemplos se utilizaron para verificar el rendimiento de los modelos en escenarios menos representativos. Los resultados obtenidos se muestran en la Tabla 8, donde se puede notar que en los resultados son mejores para todos los modelos respecto al conjunto de pruebas inicial.

Los ejemplos eliminados son los siguientes:

- DLDS_ID: 0060 Serie: 904 Secuencia: *T1w Early Arterial*
- DLDS_ID: 0092 Serie: 43 Secuencia: *T1w Axial Late Dynamic*
- DLDS_ID: 0081 Serie: 10 Secuencia: *T1w Mid Arterial*
- DLDS_ID: 0089 Serie: 60 Secuencia: *T1w Axial Late Dynamic*
- DLDS_ID: 0087 Serie: 14 Secuencia: *T1w Mid Arterial*
- DLDS_ID: 0001 Serie: 34 Secuencia: *T1w Mid Arterial*

Las secuencias que no fueron previamente presentadas debido a su escasez se describen de forma detallada en la Tabla 7. A pesar de su baja frecuencia en el conjunto de datos, los resultados obtenidos en los 6 ejemplos presentados en la Tabla 8, para los dos modelos propuestos, son notables. Estos resultados reflejan la capacidad del modelo para generalizar a partir de secuencias menos comunes.

Tabla 7: Parámetros de las secuencias menos presentes en el dataset a utilizar.
Fuente: *Duke Liver Dataset* [Macdonald *et al.*, 2023].

Sequence	Scan Count	No. of Sections	Slice Thickness (mm)	Pixel Spacing (mm)	Matrix Size Rows	Matrix Size Columns	TR (msec)	TE (msec)	Flip Angle (°)
T1w Mid Arterial	192	48-168	3-7	0.68-1.56	208-512	256-512	2.83-6.04	1.28-3.13	9-12
T1w Axial Late Dynamic	134	30-92	3-8	0.70-1.56	256-512	256-512	3.62-6.12	1.07-3.13	9-30
T1w Early Arterial	62	50-88	4-7	0.70-1.56	256-512	256-512	2.83-6.96	1.28-3.13	9-12

Este desempeño excepcional puede explicarse al comparar las secuencias escasas con aquellas mayoritariamente presentes en el conjunto de datos. Cabe destacar que tres de las secuencias (C, E, O) son del tipo T1w, una modalidad que también predomina en el conjunto de entrenamiento. La similitud entre estas secuencias y las predominantes en el entrenamiento parece haber facilitado la adaptación del modelo, permitiéndole lograr resultados sobresalientes incluso en casos menos representados.

Tabla 8: Promedio $\pm$ desviación estándar de las métricas relevantes para secuencias **fuera** del conjunto de datos principal.

Fuente: Elaboración propia.

	Puntaje Dice	IoU	Cantidad de CC	Tiempo de Inferencia [s]
3D UNet	0.9445 $\pm$ 0.0104	0.8950 $\pm$ 0.0186	1.6667 $\pm$ 0.8165	0.0012 $\pm$ 0.0010
3D Res UNet	0.9424 $\pm$ 0.0103	0.8912 $\pm$ 0.0185	2.0000 $\pm$ 0.6325	0.0014 $\pm$ 0.0007
3dResUNetA	0.8410 $\pm$ 0.0979	0.7363 $\pm$ 0.1304	1.8571 $\pm$ 1.1152	0.0116 $\pm$ 0.0048
Volumetry Net	0.9004 $\pm$ 0.0550	0.8224 $\pm$ 0.0866	7.5000 $\pm$ 5.7879	1.0169 $\pm$ 0.4392
nnUNet	0.2471 $\pm$ 0.1875	0.1545 $\pm$ 0.1305	11.9714 $\pm$ 14.5652	64.342 $\pm$ 3.9074

4.3. RESULTADOS UTILIZANDO POST PROCESAMIENTO

Un punto a considerar de la segmentación del hígado, es que este es solo un volumen a predecir, por lo tanto, en la predicción perfecta solo se debería tener una componente conexa (CC). Como se puede ver en las tablas 3 y 8, este no es el caso en los modelos que fueron probados. Para eliminar este error, se experimentó con dejar únicamente la CC más grande en la predicción. Con esto se puede notar que los resultados de las redes 3D UNet y 3D Res UNet se mantienen iguales, lo que quiere decir que las componentes conexas para estas redes, eran muy pequeñas, aportando muy poco al puntaje Dice (tabla 9). En Volumetry Net hay una mejora sustancial en el puntaje Dice promedio, aumentando en 2 décimas a 0.8129 ± 0.1434 . De forma opuesta, en nnUNet se ve un desempeño drásticamente peor, lo que demuestra la gran fragmentación existente en las predicciones realizadas por este modelo.

Se mantiene que la red con mejor desempeño es la propuesta en este proyecto, 3D UNet.

Tabla 9: Resultados utilizando únicamente el volumen más grande de la predicción realizada.
Fuente: Elaboración propia.

	Puntaje Dice	IoU	Cantidad de CC	Tiempo de Inferencia [s]
3D UNet	0.8824 ± 0.0670	0.7955 ± 0.1011	1.0000 ± 0.0000	0.0038 ± 0.0006
3D Res UNet	0.8794 ± 0.0710	0.7914 ± 0.1066	1.0000 ± 0.0000	0.0061 ± 0.0038
Volumetry Net	0.8129 ± 0.1434	0.7059 ± 0.1807	1.0000 ± 0.0000	3.0698 ± 3.8900
nnUNet	0.0463 ± 0.0223	0.0238 ± 0.0117	1.0000 ± 0.0000	-

Tabla 10: Dice promedio + desviación estándar por Secuencia según modelo evaluado en conjunto de prueba

Secuencia	3D UNet	3D Res UNet	Volumetry Net	nnUNet
B	0.8958 ± 0.0566	0.8928 ± 0.0594	0.8607 ± 0.0765	0.2187 ± 0.1713
G	0.8456 ± 0.0758	0.8417 ± 0.0783	0.6065 ± 0.1680	0.3156 ± 0.1396
H	0.8870 ± 0.0789	0.8805 ± 0.0882	0.8504 ± 0.1012	0.4314 ± 0.2517
K	0.8976 ± 0.0619	0.8950 ± 0.0683	0.8734 ± 0.0769	0.1247 ± 0.1166

4.4. RESULTADOS POR SECUENCIA

El último análisis a realizar sobre los resultados es sobre la calidad de las secuencias para realizar las segmentaciones. Los resultados numéricos se pueden ver en la tabla 10 donde se puede notar una alta similitud entre las etiquetas B, H y K en términos de Dice promedio, con valores superiores a 0.88 para todos los modelos evaluados. En particular, la secuencia K presentó los mejores resultados globales, alcanzando un Dice promedio de 0.897 con el modelo 3D UNet, lo que sugiere una mayor consistencia en la segmentación de esta secuencia. Por otro lado, la secuencia G mostró un desempeño significativamente menor, especialmente con el modelo Volumetry Net, que obtuvo un Dice promedio de solo 0.606, lo que indica una menor efectividad de esta secuencia para la segmentación. En el caso de nnUNet, se observó un rendimiento inferior en todas las secuencias, destacando valores especialmente bajos en las secuencias B y K, lo que comprueba la dificultad de transferencia entre dominios desde datos CT a MRI.

De forma más clara, se puede confirmar la superioridad de la secuencia K para realizar segmentaciones de hígado en las Figuras 24, 25, 26 y 27. En los *boxplots* se puede notar el impacto de la desviación estándar a través del tamaño de las cajas, donde Volumetry Net es la más consistente de las 3, en los contrastes B, K y H, mientras que la menos consistente en el contraste G, donde se ve un notorio declive respecto a los otros modelos. Esto nos muestra de forma drástica que la red Volumetry Net no realiza segmentaciones certeras en exámenes tomados con la secuencia G (Axial en Fase). Además, nnUNet presenta una mayor variabilidad en las secuencias evaluadas.

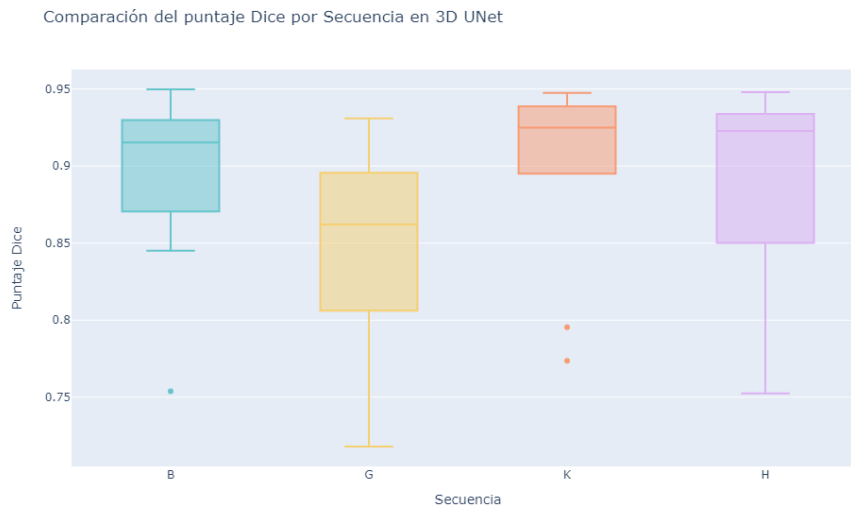


Figura 24: Puntaje Dice por Etiqueta en conjunto de prueba. Predicciones realizadas con 3D UNet.

Fuente: Elaboración Propia

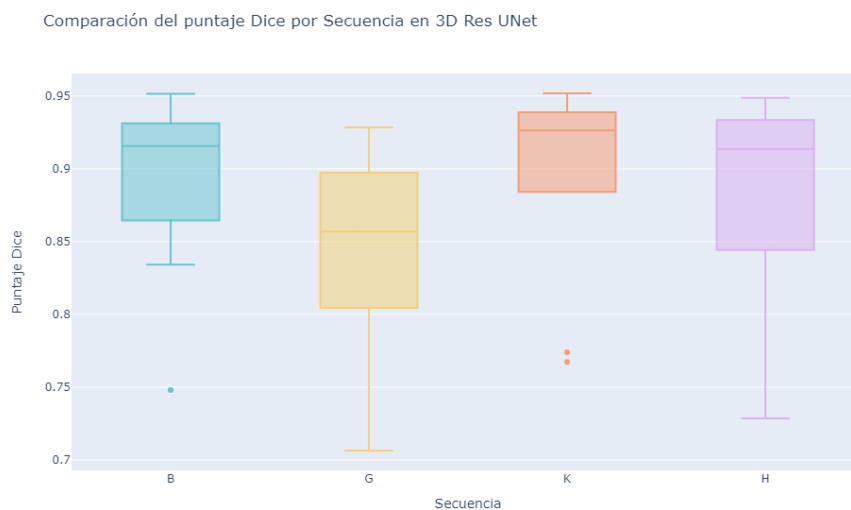


Figura 25: Puntaje Dice por Etiqueta en conjunto de prueba. Predicciones realizadas con 3D Res UNet.

Fuente: Elaboración Propia

Se puede observar que ambas redes propuestas, 3D UNet y 3D Res UNet, muestran un desempeño notable en exámenes tomados con las secuencias B, K y H, alcanzando un puntaje Dice promedio cercano a 0.89. Este rendimiento es una indicación clara de la eficacia de estas arquitecturas para la tarea de segmentación de hígado en imágenes de RM. Comparado

con el estado del arte en este campo, estos resultados representan una mejora significativa, lo que subraya la capacidad de estas redes para proporcionar segmentaciones precisas y fiables en estos tipos de secuencias, demostrando su potencial para segmentación de hígado en MRI.

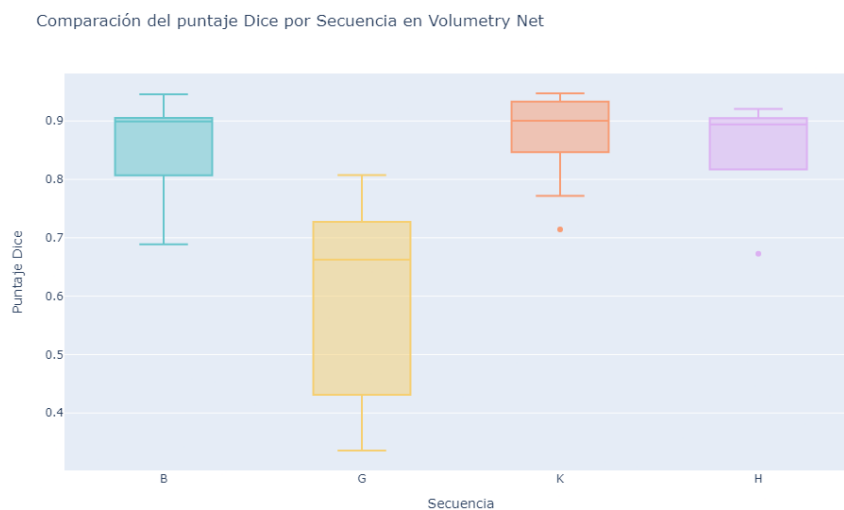


Figura 26: Puntaje Dice por Etiqueta en conjunto de prueba. Predicciones realizadas con Volumetry Net.

Fuente: Elaboración Propia

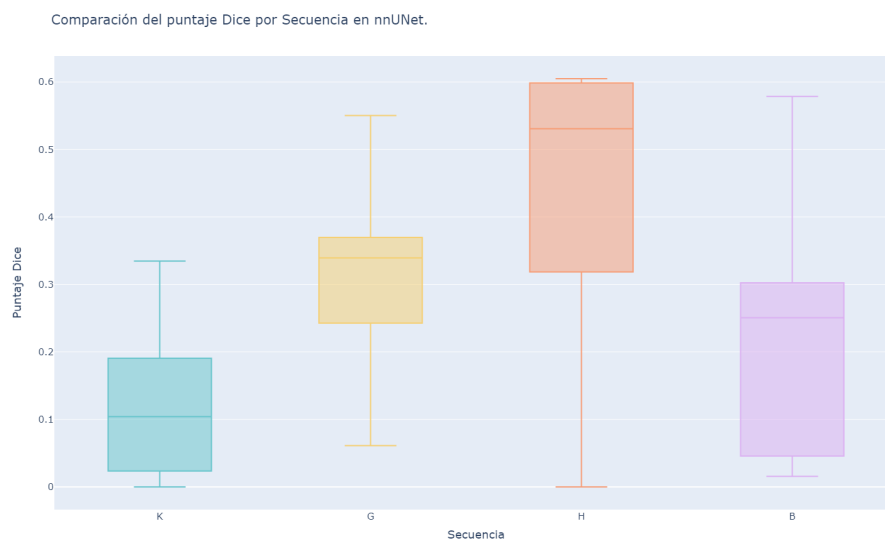


Figura 27: Puntaje Dice por Etiqueta en conjunto de prueba. Predicciones realizadas con nnUNet.

Fuente: Elaboración Propia

Tanto 3D UNet como 3D Res UNet demuestran ser modelos robustos y consistentes, superando a Volumetry Net en la mayoría de las etiquetas, con especial relevancia en la etiqueta G. Estos resultados destacan la superioridad de los primeros dos modelos en términos de precisión y estabilidad, sugiriendo que podrían ser opciones más adecuadas para aplicaciones que requieren un alto nivel de exactitud y consistencia en la segmentación 3D.

4.5. EXPERIMENTACIÓN CON CABECERA DICOM

Los resultados del modelo 3D UNetA se mantuvieron separados del resto, ya que no se observó una mejora en comparación con el modelo base, 3D Res UNet. En la Tabla 11, se puede notar que los resultados sobre el conjunto de pruebas empeoran. Sin embargo, se considera de interés la metodología, que integra una codificación de información textual para ser entregada al modelo. La información de la cabecera DICOM contenía principalmente datos sobre la máquina de RM y algunas descripciones generales del paciente, como sexo, edad, altura y peso. Sin embargo, estos datos no fueron de gran utilidad para la arquitectura implementada. Se puede observar una predicción para el mismo examen con 3D Res UNet y su variación que considera metadatos, 3D Res UNetA, donde la diferencia en la calidad de las predicciones no es notable. Esta es una vía de interés para este conjunto de datos, ya que las imágenes tienen diferentes características según sus secuencias, lo que hace que traspasar esta información al modelo sea de alta relevancia.

Tabla 11: Resultados de entrenamiento utilizando metadatos de la cabecera DICOM. 3D Res UNetA CC muestra los resultados cuando se considera únicamente la CC más grande de la predicción. Se incluyen los resultados ya presentados de 3D Res UNet como referencia

Fuente: Elaboración Propia.

	Puntaje Dice	IoU	Cantidad de CC	Tiempo de Inferencia
3D Res UNetA	0.8410 ± 0.0979	0.7363 ± 0.1304	1.3226 ± 0.6720	0.0197 ± 0.0164
3D Res UNetA CC	0.8413 ± 0.0979	0.7367 ± 0.1305	1.0000 ± 0.0000	0.0090 ± 0.0071
3D Res UNet	0.8785 ± 0.0717	0.7900 ± 0.1073	2.6286 ± 1.9110	0.0034 ± 0.0002

CAPÍTULO 5

CONCLUSIONES

5.1. Conclusiones Generales

En esta memoria se implementó un algoritmo de inteligencia artificial para la segmentación de Hígado en imágenes de RM y se experimentó con diferentes configuraciones, para luego comparar lo propuesto con el estado del arte. Lo anterior se realizó con el objetivo de proveer una solución estandarizada y fiable para el análisis volumétrico del hígado, abordando la necesidad de herramientas precisas en el ámbito médico, particularmente en el diagnóstico y tratamiento de enfermedades hepáticas crónicas.

Se realizó una comparación exhaustiva entre el modelo propuesto y dos modelos del estado del arte, demostrando que la solución presentada no solo es competitiva sino también robusta en diversas condiciones. Los resultados obtenidos resaltan la importancia de contar con algoritmos adaptados para el dominio de imágenes médicas, especialmente RM, dada la superioridad de estas en la detección temprana de anomalías hepáticas en comparación con otras modalidades de imagen. En esta comparación, se subraya la necesidad de contar con datasets amplios y bien anotados, como el DLDS, para el entrenamiento y validación de modelos de segmentación. La integración de técnicas de aprendizaje profundo ha mostrado ser efectiva para superar las limitaciones de los métodos tradicionales, ofreciendo una herramienta que puede ser aplicada de manera más eficiente y precisa en entornos clínicos.

Este trabajo contribuye al avance en el campo de la segmentación automática de órganos al demostrar la calidad de las soluciones con base en un buen conjunto de datos. Con la continua evolución de los algoritmos de inteligencia artificial y la disponibilidad creciente de datos, se espera que las soluciones como la presentada en esta memoria sigan mejorando y se integren cada vez más en la práctica clínica diaria.

5.2. Cumplimiento de objetivos

El resultado final de esta memoria se puede medir en función del cumplimiento de cada uno de los objetivos generales y específicos descritos en la sección 1.2, en primer lugar se presentan los resultados de los objetivos específicos.

- **Definir un algoritmo de segmentación automática de hígado a partir del estudio del estado del arte, en segmentación de órganos en imágenes tomadas con RM.:** Fue logrado exitosamente mediante el desarrollo de un algoritmo de segmentación automática que incorpora técnicas avanzadas de aprendizaje profundo. A partir de un análisis exhaustivo de las metodologías existentes en la literatura (sección 2.3), se definió

una arquitectura adecuada para las particularidades de las imágenes de RM, asegurando que el algoritmo diseñado sea capaz de manejar las complejidades inherentes a estas imágenes, tales como variaciones en la textura y morfología hepática (sección 2.4.3). Este enfoque permitió construir una solución que se adapta eficazmente a las necesidades del diagnóstico hepático.

- **Validar el modelo confirmando la efectividad y precisión del algoritmo propuesto.:** El modelo propuesto fue validado rigurosamente utilizando un conjunto de datos diverso y bien anotado, el DLDS (sección 3.3). Los resultados de las pruebas confirmaron la efectividad y precisión del algoritmo en la segmentación del hígado, superando los desafíos asociados con la variabilidad en las imágenes médicas. El desempeño del modelo se evaluó utilizando métricas estandarizadas, como el Dice Score y el Intersection over Union (IoU), demostrando una alta capacidad de generalización y fiabilidad en las predicciones, lo que lo convierte en una herramienta prometedora para su aplicación clínica.
- **Comparar el modelo propuesto con dos modelos del estado del arte en un conjunto de pruebas.:** Este objetivo también se ve evidenciado en la sección 3.3, donde se compara con dos modelos del estado del arte que están disponibles en código abierto. Los resultados obtenidos destacaron que el modelo desarrollado no solo es competitivo sino también robusto en diversas condiciones clínicas. La comparación mostró que, aunque todos los modelos tienen sus fortalezas, el algoritmo diseñado ofrece una mejora significativa en la segmentación del hígado en imágenes de RM

Realizando una composición del trabajo desarrollado para cumplir con los objetivos específicos, el objetivo general de esta memoria, “Implementar un algoritmo de segmentación automática de hígado en imágenes de RM”, se cumplió.

5.3. Trabajo futuro

En futuras investigaciones, se podría explorar el uso de diferentes variaciones de la arquitectura 3D U-Net para mejorar la segmentación del hígado en imágenes de RM. Ejemplos de estas variaciones incluyen UNETR (UNet Transformer), que integra mecanismos de atención para capturar mejor las relaciones espaciales en las imágenes [Hatamizadeh *et al.*, 2021], UNet++, que utiliza caminos de codificación más densos para refinar la segmentación y mejorar la precisión [Zhou *et al.*, 2018], y *Attention* U-Net, que incorpora módulos de atención para enfocarse en regiones de interés específicas dentro de las imágenes médicas [Oktay *et al.*, 2018].

Otra área prometedora de trabajo futuro es el preentrenamiento de la red utilizando las numerosas imágenes no etiquetadas disponibles en el DLDS. Dado que el dataset contiene una gran cantidad de datos sin etiquetas, se podría emplear un enfoque de aprendizaje semi-supervisado o auto-supervisado, donde la red es preentrenada en estas imágenes no eti-

quetadas para aprender representaciones más generales y robustas. Este preentrenamiento podría mejorar significativamente el desempeño del modelo en la tarea de segmentación automática, especialmente cuando se enfrenta a variaciones en las imágenes o a datos limitados en el conjunto de entrenamiento etiquetado.

Respecto al mismo análisis del DLDS, se podría realizar un rebalanceo del dataset para equiparar la cantidad de ejemplos de secuencias T1 y T2. En los resultados, se observó que la insuficiente representación de secuencias T2 en el dataset contribuyó a un desempeño subóptimo del modelo en secuencias T2. Al abordar este desequilibrio, se podría mejorar la capacidad del modelo para generalizar mejor entre diferentes tipos de secuencias de RM, logrando una segmentación más precisa y confiable.

Adicionalmente, sería de valioso contar con segmentaciones de cada subdivisión del hígado. La capacidad de segmentar de manera precisa los diferentes segmentos hepáticos podría ofrecer un análisis mucho más exhaustivo y completo del órgano. Esto no solo enriquecería las aplicaciones en el ámbito clínico, sino que también tendría aplicaciones cruciales en la planificación de tratamientos para enfermedades hepáticas crónicas, donde la precisión y el detalle son fundamentales para el éxito terapéutico [Zhao *et al.*, 2022]. Así, la implementación de técnicas avanzadas de segmentación en los distintos segmentos del hígado podría marcar una diferencia significativa en el manejo clínico de estas patologías.

Explorar estas direcciones futuras no solo podría mejorar la precisión y la generalización del modelo desarrollado, sino que también contribuiría a la evolución de las técnicas de segmentación en imágenes de RM, potenciando su aplicación en el diagnóstico clínico y en la planificación de tratamientos para enfermedades hepáticas crónicas.

REFERENCIAS BIBLIOGRÁFICAS

- [Antonelli et al., 2022] Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., Litjens, G., Menze, B., Ronneberger, O., Summers, R. M., van Ginneken, B., Bilello, M., Bilic, P., Christ, P. F., Do, R. K. G., Gollub, M. J., Heckers, S. H., Huisman, H., Jarnagin, W. R., McHugo, M. K., Napel, S., Pernicka, J. S. G., Rhode, K., Tobon-Gomez, C., Vorontsov, E., Meakin, J. A., Ourselin, S., Wiesenfarth, M., Arbeláez, P., Bae, B., Chen, S., Daza, L., Feng, J., He, B., Isensee, F., Ji, Y., Jia, F., Kim, I., Maier-Hein, K., Merhof, D., Pai, A., Park, B., Perslev, M., Rezaiifar, R., Rippel, O., Sarasua, I., Shen, W., Son, J., Wachinger, C., Wang, L., Wang, Y., Xia, Y., Xu, D., Xu, Z., Zheng, Y., Simpson, A. L., Maier-Hein, L., y Cardoso, M. J. (2022). The medical segmentation decathlon. *Nature Communications*, 13(1):4128.
- [Asrani et al., 2019] Asrani, S. K., Devarbhavi, H., Eaton, J., y Kamath, P. S. (2019). Burden of liver diseases in the world. *Journal of hepatology*, 70(1):151–171.
- [Azad et al., 2022] Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J. P., Adeli, E., y Merhof, D. (2022). Medical image segmentation review: The success of u-net. *arXiv preprint arXiv:2211.14830*.
- [Bilic et al., 2023] Bilic, P., Christ, P., Li, H. B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szekin, A., Jacobs, C., Mamani, G. E. H., Chartrand, G., Lohöfer, F., Holch, J. W., Sommer, W., Hofmann, F., Hostettler, A., Lev-Cohain, N., Drozdal, M., Amitai, M. M., Vivanti, R., Sosna, J., Ezhov, I., Sekuboyina, A., Navarro, F., Kofler, F., Paetzold, J. C., Shit, S., Hu, X., Lipková, J., Rempfler, M., Piraud, M., Kirschke, J., Wiestler, B., Zhang, Z., Hülsemeyer, C., Beetz, M., Ettlinger, F., Antonelli, M., Bae, W., Bellver, M., Bi, L., Chen, H., Chlebus, G., Dam, E. B., Dou, Q., Fu, C.-W., Georgescu, B., i Nieto, X. G., Gruen, F., Han, X., Heng, P.-A., Hesser, J., Moltz, J. H., Igel, C., Isensee, F., Jäger, P., Jia, F., Kaluva, K. C., Khened, M., Kim, I., Kim, J.-H., Kim, S., Kohl, S., Konopczynski, T., Kori, A., Krishnamurthi, G., Li, F., Li, H., Li, J., Li, X., Lowengrub, J., Ma, J., Maier-Hein, K., Maninis, K.-K., Meine, H., Merhof, D., Pai, A., Perslev, M., Petersen, J., Pont-Tuset, J., Qi, J., Qi, X., Rippel, O., Roth, K., Sarasua, I., Schenk, A., Shen, Z., Torres, J., Wachinger, C., Wang, C., Weninger, L., Wu, J., Xu, D., Yang, X., Yu, S. C.-H., Yuan, Y., Yue, M., Zhang, L., Cardoso, J., Bakas, S., Braren, R., Heinemann, V., Pal, C., Tang, A., Kadoury, S., Soler, L., van Ginneken, B., Greenspan, H., Joskowicz, L., y Menze, B. (2023). The liver tumor segmentation benchmark (lits). *Medical Image Analysis*, 84:102680.
- [Cashmore et al., 2021] Cashmore, M. T., McCann, A. J., Wastling, S. J., McGrath, C., Thornton, J., y Hall, M. G. (2021). Clinical quantitative mri and the need for metrology. *The British Journal of Radiology*, 94:20201215.
- [Chen et al., 2018] Chen, Y., Almarzouqi, S. J., Morgan, M. L., y Lee, A. G. (2018). T1-weighted image. *Encyclopedia of Ophthalmology*, pp. 1747–1750.
- [Çiçek et al., 2016] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., y Ronneberger, O. (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation. En

- Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II* 19, pp. 424–432. Springer.
- [Dale et al., 2015] Dale, B. M., Brown, M. A., y Semelka, R. C. (2015). *MRI: basic principles and applications*. John Wiley & Sons.
- [Devarbhavi et al., 2023] Devarbhavi, H., Asrani, S. K., Arab, J. P., Nartey, Y. A., Pose, E., y Kamath, P. S. (2023). Global burden of liver disease: 2023 update. *Journal of hepatology*, 79:516–537.
- [Díaz et al., 2024] Díaz, L. A., Villota-Rivas, M., Barrera, F., Lazarus, J. V., y Arrese, M. (2024). The burden of liver disease in latin america. *Annals of Hepatology*, 29(3):101175.
- [D’Onofrio et al., 2014] D’Onofrio, M., De Robertis, R., Demozzi, E., Crosara, S., Canestrini, S., y Mucelli, R. P. (2014). Liver volumetry: Is imaging reliable? personal experience and review of the literature. *World journal of radiology*, 6(4):62.
- [Finnegan et al., 2023] Finnegan, R. N., Chin, V., Chlap, P., Haidar, A., Otton, J., Dowling, J., Thwaites, D. I., Vinod, S. K., Delaney, G. P., y Holloway, L. (2023). Open-source, fully-automated hybrid cardiac substructure segmentation: development and optimisation. *Physical and Engineering Sciences in Medicine*, 46(1):377–393.
- [Fiorio y Gustedt, 1996] Fiorio, C. y Gustedt, J. (1996). Two linear time union-find strategies for image processing. *Theoretical Computer Science*, 154(2):165–181.
- [Glinka et al., 2019] Glinka, J., Bruballa, R., de Santibañes, M., Clariá, R. S., Ardiles, V., Mazza, O. M., Pekolj, J., y de Santibañes, E. (2019). Preoperative portal vein embolization followed by right hepatectomy to treat a complex common bile duct injury in a 5-year-old child. *World Journal of Pediatric Surgery*, 2(2).
- [Gotra et al., 2017] Gotra, A., Sivakumaran, L., Chartrand, G., Vu, K. N., Vandenbroucke-Menu, F., Kauffmann, C., Kadoury, S., Gallix, B., de Guise, J. A., y Tang, A. (2017). Liver segmentation: indications, techniques and future directions.
- [Gross et al., 2024] Gross, M., Huber, S., Arora, S., Ze’evi, T., Haider, S. P., Kucukkaya, A. S., Iseke, S., Kuhn, T. N., Gebauer, B., Michallek, F., Dewey, M., Vilgrain, V., Sartoris, R., Ronot, M., Jaffe, A., Strazzabosco, M., Chapiro, J., y Onofrey, J. A. (2024). Automated mri liver segmentation for anatomical segmentation, liver volumetry, and the extraction of radiomics. *European Radiology*, 34(8):5056–5065.
- [Gul et al., 2022] Gul, S., Khan, M. S., Bibi, A., Khandakar, A., Ayari, M. A., y Chowdhury, M. E. (2022). Deep learning techniques for liver and liver tumor segmentation: A review. *Computers in Biology and Medicine*, 147:105620.
- [Hatamizadeh et al., 2021] Hatamizadeh, A., Yang, D., Roth, H., y Xu, D. (2021). Unetr: Transformers for 3d medical image segmentation. 2022 ieee. En *CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1748–1758.

- [He et al., 2016] He, K., Zhang, X., Ren, S., y Sun, J. (2016). Deep residual learning for image recognition. En *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.
- [Heimann et al., 2009] Heimann, T., van Ginneken, B., Styner, M. A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., Bello, F., Binnig, G., Bischof, H., Bornik, A., Cashman, P. M. M., Chi, Y., Cordova, A., Dawant, B. M., Fidrich, M., Furst, J. D., Furukawa, D., Grenacher, L., Hornegger, J., Kainmüller, D., Kitney, R. I., Kobatake, H., Lamecker, H., Lange, T., Lee, J., Lennon, B., Li, R., Li, S., Meinzer, H.-P., Nemeth, G., Raicu, D. S., Rau, A.-M., van Rikxoort, E. M., Rousson, M., Rusko, L., Saddi, K. A., Schmidt, G., Seghers, D., Shimizu, A., Slagmolen, P., Sorantin, E., Soza, G., Susomboon, R., Waite, J. M., Wimmer, A., y Wolf, I. (2009). Comparison and evaluation of methods for liver segmentation from ct datasets. *IEEE Transactions on Medical Imaging*, 28(8):1251–1265.
- [Huynh et al., 2014] Huynh, H. T., Karademir, I., Oto, A., y Suzuki, K. (2014). Computerized liver volumetry on mri by using 3d geodesic active contour segmentation. *American Journal of Roentgenology*, 202.
- [Ioffe y Szegedy, 2015] Ioffe, S. y Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. En *International conference on machine learning*, pp. 448–456. pmlr.
- [Isensee et al., 2021] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., y Maier-Hein, K. H. (2021). nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211.
- [Jansen et al., 2019] Jansen, M. J. A., Kuijf, H. J., Niekel, M., Veldhuis, W. B., Wessels, F. J., Viergever, M. A., y Pluim, J. P. W. (2019). Liver segmentation and metastases detection in mr images using convolutional neural networks. *Journal of Medical Imaging*, 6:1.
- [Le Roy et al., 2018] Le Roy, B., Dupré, A., Gallon, A., Chabrot, P., Gagnière, J., y Buc, E. (2018). Liver hypertrophy: underlying mechanisms and promoting procedures before major hepatectomy. *Journal of visceral surgery*, 155(5):393–401.
- [Lenchik et al., 2019a] Lenchik, L., Heacock, L., Weaver, A. A., Boutin, R. D., Cook, T. S., Itri, J., Filippi, C. G., Gullapalli, R. P., Lee, J., Zagurovskaya, M., Retson, T., Godwin, K., Nicholson, J., y Narayana, P. A. (2019a). Automated segmentation of tissues using ct and mri: A systematic review. *Academic Radiology*, 26(12):1695–1706.
- [Lenchik et al., 2019b] Lenchik, L., Heacock, L., Weaver, A. A., Boutin, R. D., Cook, T. S., Itri, J., Filippi, C. G., Gullapalli, R. P., Lee, J., Zagurovskaya, M., Retson, T., Godwin, K., Nicholson, J., y Narayana, P. A. (2019b). Automated segmentation of tissues using ct and mri: A systematic review. *Academic Radiology*, 26:1695–1706.
- [Liu et al., 2020] Liu, L., Cheng, J., Quan, Q., Wu, F.-X., Wang, Y.-P., y Wang, J. (2020). A survey on u-shaped networks in medical image segmentations. *Neurocomputing*, 409:244–258.

- [Liu *et al.*, 2021] Liu, X., Song, L., Liu, S., y Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. *Sustainability*, 13(3):1224.
- [Liu y Chen, 2022] Liu, Y.-B. y Chen, M.-K. (2022). Epidemiology of liver cirrhosis and associated complications: Current knowledge and future directions. *World Journal of Gastroenterology*, 28(41):5910.
- [Long *et al.*, 2014] Long, J., Shelhamer, E., Darrell, T., y Berkeley, U. (2014). Fully convolutional networks for semantic segmentation. *arXiv preprint arXiv:1411.4038*.
- [Macdonald *et al.*, 2023] Macdonald, J. A., Zhu, Z., Konkel, B., Mazurowski, M. A., Wiggins, W. F., y Bashir, M. R. (2023). Duke liver dataset: A publicly available liver mri dataset with liver segmentation masks and series labels. *Radiology: Artificial Intelligence*, 5.
- [McRobbie *et al.*, 2017] McRobbie, D. W., Moore, E. A., Graves, M. J., y Prince, M. R. (2017). *MRI from Picture to Proton*. Cambridge university press.
- [Nandwana *et al.*, 2021] Nandwana, S. B., Finazzo, J., Alexander, L. F., y Cox, K. (2021). Using qualitative descriptors of chronic liver disease on mri: A practice prone to error. *Clinical Imaging*, 74:89–92.
- [Oh *et al.*, 2023] Oh, N., Kim, J.-H., Rhu, J., Jeong, W. K., Choi, G.-s., Kim, J. M., y Joh, J.-W. (2023). Automated 3d liver segmentation from hepatobiliary phase mri for enhanced preoperative planning. *Scientific Reports*, 13(1):17605.
- [Oktay *et al.*, 2018] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M. J., Heinrich, M. P., Misawa, K., Mori, K., McDonagh, S. G., Hammerla, N. Y., Kainz, B., Glocker, B., y Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. *ArXiv*, abs/1804.03999.
- [Oliva y Saini, 2004] Oliva, M. R. y Saini, S. (2004). Liver cancer imaging: role of ct, mri, us and pet. *Cancer imaging*, 4(Spec No A):S42.
- [Pouwels *et al.*, 2022] Pouwels, S., Sakran, N., Graham, Y., Leal, A., Pintar, T., Yang, W., Kasir, R., Singhal, R., Mahawar, K., y Ramnarain, D. (2022). Non-alcoholic fatty liver disease (nafld): a review of pathophysiology, clinical management and effects of weight loss. *BMC endocrine disorders*, 22(1):63.
- [Reeder *et al.*, 2011] Reeder, S. B., Cruite, I., Hamilton, G., y Sirlin, C. B. (2011). Quantitative assessment of liver fat with magnetic resonance imaging and spectroscopy. *Journal of magnetic resonance imaging : JMRI*, 34.
- [Rezatofghi *et al.*, 2019] Rezatofghi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., y Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. En *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 658–666.

- [Romero R et al., 2022] Romero R, W. A., Viallon, M., Spaltenstein, J., Petrusca, L., Bernard, O., Belle, L., Clarysse, P., y Croisille, P. (2022). Cmrsegtools: An open-source software enabling reproducible research in segmentation of acute myocardial infarct in cmr images. *Plos one*, 17(9):e0274491.
- [Ronneberger et al., 2015] Ronneberger, O., Fischer, P., y Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. En *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III* 18, pp. 234–241. Springer.
- [Sharma y Aggarwal, 2010] Sharma, N. y Aggarwal, L. M. (2010). Automated medical image segmentation techniques. *Journal of medical physics*, 35(1):3–14.
- [Singh y Singh, 2021] Singh, M. K. y Singh, K. K. (2021). A review of publicly available automatic brain segmentation methodologies, machine learning models, recent advancements, and their comparison. *Annals of Neurosciences*, 28:82.
- [Terrault et al., 2023] Terrault, N. A., Francoz, C., Berenguer, M., Charlton, M., y Heimbach, J. (2023). Liver transplantation 2023: status report, current and future challenges. *Clinical Gastroenterology and Hepatology*, 21(8):2150–2166.
- [Vu et al., 2017] Vu, L. N., Morelli, J. N., y Szklaruk, J. (2017). Basic mri for the liver oncologists and surgeons. *Journal of Hepatocellular Carcinoma*, 5:37.
- [Wang et al., 2019] Wang, K., Mamidipalli, A., Retson, T., Bahrami, N., Hasenstab, K., Blansit, K., Bass, E., Delgado, T., Cunha, G., Middleton, M. S., Loomba, R., Neuschwander-Tetri, B. A., Sirlin, C. B., y Hsiao, A. (2019). Automated ct and mri liver segmentation and biometry using a generalized convolutional neural network. *Radiology: Artificial Intelligence*, 1.
- [Westbrook, 2016] Westbrook, C. (2016). *MRI at a Glance*. John Wiley & Sons.
- [Wiegand y Berg, 2013] Wiegand, J. y Berg, T. (2013). The etiology, diagnosis and prevention of liver cirrhosis: part 1 of a series on liver cirrhosis. *Deutsches Ärzteblatt International*, 110(6):85.
- [Wu et al., 2005] Wu, K., Otoo, E., y Shoshani, A. (2005). Optimizing connected component labeling algorithms. En *Medical Imaging 2005: Image Processing*, volumen 5747, pp. 1965–1976. SPIE.
- [Zhao et al., 2022] Zhao, D., Zhang, K.-J., Fang, T.-S., Yan, X., Jin, X., Liang, Z.-M., Tang, J.-X., y Xie, L.-J. (2022). Topological approach of liver segmentation based on 3d visualization technology in surgical planning for split liver transplantation. *World Journal of Gastrointestinal Surgery*, 14(10):1141.
- [Zhou et al., 2018] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., y Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. En *Deep Learning*

in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, pp. 3–11. Springer.