



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Aprendizaje profundo y biomarcadores de
imagen en el estudio de enfermedades
metabólicas y hepáticas a partir de resonancia
magnética y tomografía computarizada

*Programa de Doctorado en Tecnologías para la
Salud y el Bienestar*

Presentado por:

Ana Jiménez Pastor

Dirigida por:

Dr. Ignacio Bosch Roig

*Instituto de Telecomunicaciones y Aplicaciones Multimedia.
Dpto. de Comunicaciones. Universitat Politècnica de València.*

Dr. Ángel Alberich Bayarri

Quibim SL.

Septiembre 2023

Agradecimientos

Quería reservar este espacio para dedicar mi agradecimiento a todas las personas que han hecho posible esta tesis doctoral.

A Ángel Alberich Bayarri por su constante apoyo y confianza en mí durante estos siete años desde que me inicié en esta “aventura”. Sin su orientación y guía, no habría sido posible llegar hasta aquí.

A Ignacio Bosch Roig cuyos conocimientos y experiencias han sido fundamentales para el éxito de este proyecto.

A Luís Martí Bonmatí, por ir siempre un paso por delante, sin esa inquietud y visión no estaríamos donde estamos.

A cada uno de los miembros de la familia Quibimer, quienes han sido una parte fundamental de mi vida en los último siete años. Cada una de vuestras aportaciones han sido clave para alcanzar los resultados obtenidos.

No puedo dejar de mencionar a mi familia, quienes han sido mi pilar fundamental en todo momento. A mis padres, Ana y Jesús, y a mi hermano, Jesús, les debo todo lo que soy y siempre estaré agradecido por su amor, paciencia y apoyo incondicional.

Por último, quiero agradecer a Joaquín, quien ha estado presente en todo momento y ha sido un gran apoyo emocional para mí. Gracias por estar ahí y creer en mí.

Resumen

El síndrome metabólico se define como un conjunto de trastornos (e.g., niveles elevados de presión arterial, niveles elevados de glucosa en sangre, exceso de grasa abdominal o niveles elevados de colesterol o triglicéridos) que afectan a un individuo al mismo tiempo. La presencia de uno de estos factores no implica un riesgo elevado para la salud, sin embargo, presentar varios de ellos aumenta la probabilidad de sufrir enfermedades secundarias como la enfermedad cardiovascular o la diabetes tipo II. Las enfermedades difusas hepáticas son todas aquellas enfermedades que afectan a las células funcionales del hígado, los hepatocitos, alterando, de este modo, la función hepática. En estos procesos, los hepatocitos se ven sustituidos por adipocitos y tejido fibroso. La enfermedad de hígado graso no alcohólico es una afección reversible originada por la acumulación de triglicéridos en los hepatocitos. El alcoholismo, la obesidad, y la diabetes son las causas más comunes de esta enfermedad. Este estado del hígado es reversible si se cambia la dieta del paciente, sin embargo, si este no se cuida, la enfermedad puede ir avanzando hacia estadios más severos, desencadenando fibrosis, cirrosis e incluso carcinoma hepatocelular (CHC).

La temprana detección de todos estos procesos es de gran importancia en la mejora del pronóstico de los pacientes. Así, las técnicas de imagen en combinación con modelos computacionales permiten caracterizar el tejido mediante la extracción de parámetros objetivos, conocidos como biomarcadores de imagen, relacionados con estos procesos fisiológicos y patológicos, permitiendo una estadificación más precisa de las enfermedades. Además, gracias a las técnicas de inteligencia artificial, se pueden desarrollar algoritmos de segmentación automática que permitan realizar dicha caracterización de manera completamente automática y acelerar, de este modo, el flujo radiológico.

Por todo esto, en la presente tesis doctoral, se presenta una metodología para el desarrollo de modelos de segmentación y cuantificación automática, siendo aplicada a tres casos de uso. Para el estudio del síndrome metabólico se propone un método de segmentación automática de la grasa visceral y subcutánea en imágenes de tomografía computarizada (TC), para el estudio de la enfermedad hepática difusa se propone un método de segmentación hepática y cuantificación de la grasa y hierro hepáticos en imágenes de resonancia magnética (RM), y, finalmente, para el estudio del CHC, se propone un

método de segmentación hepática y cuantificación de los descriptores de la curva de perfusión en imágenes de RM.

Todo esto se ha integrado en una plataforma que permite su integración en la práctica clínica. Así, se han adaptado los algoritmos desarrollados para ser ejecutados en contenedores Docker de forma que, dada una imagen de entrada, generen los parámetros cuantitativos de salida junto con un informe que resuma dichos resultados; se han implementado herramientas para que los usuarios puedan interactuar con las segmentaciones generadas por los algoritmos de segmentación automática desarrollados; finalmente, éstos se han implementado de forma que generen dichas segmentaciones en formatos estándar como DICOM RT Struct o DICOM Seg, para garantizar la interoperabilidad con el resto de sistemas sanitarios.

Resum

La síndrome metabòlica es defineix com un conjunt de trastorns (e.g., nivells elevats de pressió arterial, nivells elevats de glucosa en sang, excés de greix abdominal o nivells elevats de colesterol o triglicèrids) que afecten un individu al mateix temps. La presència d'un d'aquests factors no implica un risc elevat per a la salut, no obstant això, presentar diversos d'ells augmenta la probabilitat de patir malalties secundàries com la malaltia cardiovascular o la diabetis tipus II. Les malalties difuses hepàtiques són totes aquelles malalties que afecten les cèl·lules funcionals del fetge, els hepatòcits, alterant, d'aquesta manera, la funció hepàtica. En aquests processos, els hepatòcits es veuen substituïts per adipòcits i teixit fibrós. La malaltia de fetge gras no alcohòlic és una afecció reversible originada per l'acumulació de triglicèrids en els hepatòcits. L'alcoholisme, l'obesitat, i la diabetis són les causes més comunes d'aquesta malaltia. Aquest estat del fetge és reversible si es canvia la dieta del pacient, no obstant això, si aquest no es cuida, la malaltia pot anar avançant cap a estadis més severos, desencadenant fibrosis, cirrosis i fins i tot carcinoma hepatocel·lular (CHC).

La primerenca detecció de tots aquests processos és de gran importància en la millora del pronòstic dels pacients. Així, les tècniques d'imatge en combinació amb models computacionals permeten caracteritzar el teixit mitjançant l'extracció paràmetres objectius, coneguts com biomarcadores d'imatge, relacionats amb aquests processos fisiològics i patològics, permetent una estratificació més precisa de les malalties. A més, gràcies a les tècniques d'intel·ligència artificial, es poden desenvolupar algorismes de segmentació automàtica que permeten realitzar aquesta caracterització de manera completament automàtica i accelerar, d'aquesta manera, el flux radiològic.

Per tot això, en la present tesi doctoral, es presenta una metodologia per al desenvolupament de models de segmentació i quantificació automàtica, sent aplicada a tres casos d'ús. Per a l'estudi de la síndrome metabòlica es proposa un mètode de segmentació automàtica del greix visceral i subcutani en imatges de tomografia computada (TC), per a l'estudi de la malaltia hepàtica difusa es proposa un mètode segmentació hepàtica i quantificació del greix i ferro hepàtics en imatges de ressonància magnètica (RM), i, finalment, per a l'estudi del CHC, es proposa un mètode de segmentació hepàtica i quantificació dels descriptors de la corba de perfusió en imatges de RM.

Tot això s'ha integrat en una plataforma que permet la seua integració en la pràctica clínica. Així, s'han adaptat els algorismes desenvolupats per a ser executats en contenidors Docker de manera que, donada una imatge d'entrada, generen els paràmetres quantitatius d'eixida juntament amb un informe que resumisca aquests resultats; s'han implementat eines perquè els usuaris puguin interactuar amb les segmentacions generades pels algorismes de segmentació automàtica desenvolupats; finalment, aquests s'han implementat de manera que generen aquestes segmentacions en formats estàndard com DICOM RT Struct o DICOM Seg, per a garantir la interoperabilitat amb la resta de sistemes sanitaris.

Abstract

Metabolic syndrome is defined as a group of disorders (e.g., high blood pressure, high blood glucose levels, excess abdominal fat, or high cholesterol or triglyceride levels) that affect an individual at the same time. The presence of one of these factors does not imply an elevated health risk; however, having several of them increases the probability of secondary diseases such as cardiovascular disease or type II diabetes. Diffuse liver diseases are all those diseases that affect the functional cells of the liver, the hepatocytes, thus altering liver function. In these processes, the hepatocytes are replaced by adipocytes and fibrous tissue. Non-alcoholic fatty liver disease is a reversible condition caused by the accumulation of triglycerides in hepatocytes. Alcoholism, obesity, and diabetes are the most common causes of this disease. This liver condition is reversible if the patient's diet is changed; however, if the patient is not cared for, the disease can progress to more severe stages, triggering fibrosis, cirrhosis and even hepatocellular carcinoma (HCC).

Early detection of all these processes is of great importance in improving patient prognosis. Thus, imaging techniques in combination with computational models allow tissue characterization by extracting objective parameters, known as imaging biomarkers, related to these physiological and pathological processes, allowing a more accurate stratification of diseases. Moreover, thanks to artificial intelligence techniques, it is possible to develop automatic segmentation algorithms that allow to perform such characterization in a fully automatic way and thus accelerate the radiological workflow.

Therefore, in this PhD, a methodology for the development of automatic segmentation and quantification models is presented and applied to three use cases. For the study of metabolic syndrome, a method of automatic segmentation of visceral and subcutaneous fat in computed tomography (CT) images is proposed; for the study of diffuse liver disease, a method of liver segmentation and quantification of hepatic fat and iron in magnetic resonance imaging (MRI) is proposed; and, finally, for the study of HCC, a method of liver segmentation and quantification of perfusion curve descriptors in MRI is proposed.

All this has been integrated into a platform that allows its integration into clinical practice. Thus, the developed algorithms have been adapted to be executed in Docker containers so that, given an input image, they generate the quantitative output parameters together with a report summarizing these results; tools have been implemented so that users can interact with the segmentations generated by the automatic segmentation algorithms

developed; finally, these have been implemented so that they generate these segmentations in standard formats such as DICOM RT Struct or DICOM Seg, to ensure interoperability with other health systems.

Índice

Motivación	16
Estructura	18
Capítulo 1. Introducción	21
1.1. Bases clínicas	22
1.1.1. Síndrome metabólico y composición corporal.....	22
1.1.2. Enfermedades difusas hepáticas.....	30
1.1.3. Carcinoma hepatocelular (CHC).....	32
1.2. Imagen médica	34
1.2.1. Tomografía computarizada para el estudio de la composición corporal	34
1.2.2. Resonancia magnética para el estudio de la enfermedad difusa hepática	35
1.2.3. Resonancia magnética para el estudio del carcinoma hepatocelular	38
1.3. Biomarcadores de imagen.....	43
1.3.1. ¿Qué es un biomarcador de imagen?	43
1.3.2. Biomarcadores de imagen para la caracterización de la enfermedad difusa hepática	46
1.3.3. Biomarcadores de imagen para la caracterización del carcinoma hepatocelular ...	50
1.3.3.1. Análisis de difusión	50
1.3.3.2. Análisis de perfusión	54
1.4. Aprendizaje profundo	60
1.4.1. Introducción al aprendizaje profundo.....	60
1.4.2. Redes neuronales convolucionales	63
1.4.2.1. Capa de entrada.....	64
1.4.2.2. Capa convolucional	64
1.4.2.3. Capa de agrupación.....	67
1.4.2.4. Capa de submuestreo o convolución transpuesta	68
1.4.2.5. Capa totalmente conectada	68
1.4.3. Optimización del proceso de entrenamiento.....	69
1.4.3.1. Entrenamiento en lotes	69
1.4.3.2. Sobreajuste y regularización.....	69
1.4.4. Estado del arte en el uso de CNN para resolver problemas de clasificación.....	74
1.4.5. Estado del arte en el uso de CNN para resolver problemas de segmentación.....	77
Capítulo 2. Hipótesis y objetivos	82
Capítulo 3. Materiales y métodos	85
3.1. Bases de datos	88
3.1.1. Segmentación automática hepática en secuencias MECSE.....	89
3.1.1.1. Cuantificación y validación clínica	91
3.1.2. Segmentación automática de grasa abdominal en imágenes de TC	93
3.1.2.1. Clasificación de la región abdominal	94
3.1.2.2. Segmentación de grasa visceral y subcutánea	94

3.1.3. Segmentación automática hepática en secuencias de perfusión.....	94
3.2. Anotación.....	96
3.2.1. Segmentación automática hepática en secuencias MECSE.....	96
3.2.2. Segmentación automática de grasa abdominal en imágenes de TC	97
3.2.3. Segmentación automática hepática en secuencias de perfusión.....	98
3.3. Preprocesado	100
3.3.1. Normalización	100
3.3.2. Re-escalado.....	102
3.4. Aprendizaje profundo	104
3.4.1. Software y hardware.....	104
3.4.2. Aumento de datos	104
3.4.3. Selección de la arquitectura.....	107
3.4.3.1. Segmentación automática hepática en secuencias MECSE.....	109
3.4.3.2. Segmentación automática de grasa abdominal en imágenes de TC	109
3.4.3.3. Segmentación automática hepática en secuencias de perfusión	110
3.4.4. Entrenamiento y validación	111
3.4.4.1. Selección de los hiperparámetros	113
3.4.5. Test 125	
3.4.5.1. Procesamiento de estudios de test.....	125
3.4.5.2. Métricas utilizadas en los modelos de segmentación	128
3.4.5.3. Métricas utilizadas en la clasificación	133
3.5. Cuantificación.....	134
3.5.1. Cuantificación PDFF y R2* en secuencias MECSE de RM	134
3.5.1.1. Paralelización.....	137
3.5.2. Descriptores de la curva concentración en secuencias de perfusión de RM	137
Capítulo 4. Resultados y discusión	139
4.1. Segmentación hepática y cuantificación en secuencias MECSE.....	140
4.1.1. Validación del método de segmentación hepática.....	140
4.1.2. Test del método de segmentación hepática	141
4.1.2.1. Test interno	142
4.1.2.2. Test externo	145
4.1.2.3. Discusión	151
4.1.3. Método de cuantificación	152
4.1.3.1. Paralelización y tiempo de ejecución	154
4.1.4. Validación clínica	155
4.1.4.1. Discusión	157
4.2. Segmentación de grasa abdominal en imágenes de TC	159
4.2.1. Detección de la región abdominal	159
4.2.1.1. Validación del modelo de clasificación.....	159
4.2.1.2. Test del modelo de clasificación.....	163
4.2.2. Segmentación de grasa visceral y subcutánea	165
4.2.2.1. Validación del modelo de segmentación	165
4.2.2.2. Test del modelo de segmentación.....	174
4.2.3. Discusión	177

4.3. Segmentación hepática y cuantificación en secuencias de perfusión	180
4.3.1. Validación del modelo de segmentación hepática.....	180
4.3.1.1. Orientación transversal	180
4.3.1.2. Orientación coronal	188
4.3.1.3. Orientación sagital	196
4.3.1.4. Resumen de los modelos seleccionados	203
4.3.2. Test del modelo de segmentación hepática	204
4.3.3. Discusión	208
4.4. Conclusiones generales	210
Capítulo 5. Implementación en práctica clínica.....	212
5.1. Plataforma de análisis de imagen.....	213
5.1.1. Desarrollo de un entorno de anotación	213
5.1.2. Integración de un módulo de análisis	216
5.1.3. Generación de los archivos segmentación para su visualización	218
5.2. Estándares de segmentación.....	219
5.2.1. DICOM RT Structure	219
5.2.2. DICOM Segmentation.....	221
5.2.3. Otros formatos	222
5.2.3.1. Neuroimaging Informatics Technology Initiative (NIfTI)	222
5.2.4. Servicio API REST.....	223
Capítulo 6. Conclusiones y trabajo futuro	227
Referencias.....	232
Anexo I. Informe módulo de análisis.....	245
Anexo II. Cabecera DICOM: Objetos RT Struct y Seg.....	247
Anexo III. Publicaciones.....	263

Abreviaturas

IA	Inteligencia Artificial
RM	Resonancia Magnética
MECSE	Secuencia multi-eco con desplazamiento químico
DCE	Secuencia dinámica con realce de contraste
TC	Tomografía computarizada
CHC	Carcinoma hepatocelular
OMS	Organización mundial de la salud
RCC	Relación cintura-cadera
IMC	Índice de masa corporal
CC	Circunferencia de cintura
BIA	Análisis por impedancia bioeléctrica
ADP	Pletismografía por desplazamiento de aire
DXA	Absorciometría de Rayos X de doble energía
EHGNA	Enfermedad del hígado graso no alcohólico
UH	Unidad Hounsfield
TE	Tiempo de eco
TR	Tiempo de repetición
DWI	Imagen potenciada en difusión
PDFF	Fracción grasa de densidad protónica
LIC	Concentración de hierro hepática
EIBALL	European Imaging Biomarkers Alliance
QIBA	Quantitative Imaging Biomarkers Alliance
ADC	Coefficiente Aparente de Difusión
IVIM	Movimiento intra-voxel incoherente
VEGF	Factor de crecimiento vascular endotelial
TTP	Tiempo al pico
AUC	Área bajo la curva
EEE	Espacio extravascular-extracelular

CNN	Red Neuronal Convolutacional
RAM	Memoria de acceso aleatorio
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
GPU	Unidad de procesamiento gráfico
TPU	Unidad de procesamiento de tensores
FCN	Red neuronal totalmente convolutacional
MPR	Reconstrucción Multiplanar
DC	Coefficiente Dice
TA	Tasa de aprendizaje
SGD	Descenso de gradiente estocástico
BCE	Entropía cruzada binaria
DL	Pérdidas Dice
VP	Verdaderos positivos
VN	Verdaderos negativos
FP	Falsos positivos
FN	Falsos negativos
IoU	Intersección sobre la unión
TL	Pérdidas Tversky
FTL	Pérdidas Focal Tversky
CL	Pérdidas Combo
BCE_m	Entropía cruzada binaria modificada
RVP	Relación de verdaderos positivos
FVN	Fracción de verdaderos negativos
FDR	Tasa de falsos aciertos
ASSD	Distancia media simétrica superficial
MSD	Máxima distancia superficial
RVD	Diferencia de volumen relativo
r	Coefficiente de correlación
ER	Error relativo
GAN	Modelos generativos adversarios

JSON	JavaScript Object Notation
CI/CD	Integración continua / Despliegue continuo
DICOM	Digital Imaging and Communications in Medicine
NEMA	National Electrical Manufacturers Association
NIfTI	NeuroImaging Informatics Technology Initiative
NRRD	Nearly Raw Raster Data
API	Interfaz de programación de aplicaciones
REST	Transferencia de estado representacional

Motivación

La radiología es uno de los campos de la medicina que ha sufrido el mayor cambio tecnológico en los últimos 20 años gracias a la digitalización de las imágenes adquiridas. Debido a esta digitalización, además, se ha podido pasar de evaluar imágenes bidimensionales reveladas en una lámina en la que se observan diferentes tejidos superpuestos en una misma región, a imágenes tridimensionales y n-dimensionales, con cada vez mayor resolución, que permite evaluar cada tejido de manera independiente. Además, el gran avance tecnológico de los componentes empleados ha permitido una evolución en la tecnología de adquisición de diferentes modalidades como la resonancia magnética (RM) o la tomografía por emisión de positrones (PET), entre otras.

Las enfermedades y los procesos biológicos producen pequeños cambios en el organismo que, en ocasiones, son imperceptibles para el ojo humano. Sin embargo, mediante modelos computacionales, se puede lograr la extracción de parámetros objetivos, conocidos como biomarcadores de imagen, que permiten pasar de valoraciones cualitativas de la imagen a valoraciones cuantitativas, permitiendo una estadificación más precisa de las enfermedades.

Para poder realizar la cuantificación de un tejido de interés, es necesaria su previa segmentación, es decir, seleccionar todos los vóxeles que componen el tejido. En la actualidad, este proceso de segmentación se realiza, generalmente, de manera manual o semiautomática, requiriendo del conocimiento experto. Sin embargo, esta interacción requiere tiempo del radiólogo siendo, en algunas ocasiones, una labor tediosa y costosa temporalmente, retrasando, por lo tanto, el flujo radiológico. La inteligencia artificial y los métodos de segmentación automática mediante redes neuronales convolucionales, han demostrado un gran rendimiento en la segmentación de órganos y lesiones en imagen médica radiológica, permitiendo, de este modo, la automatización completa de estos procesos.

La RM, por su naturaleza multivariante al adquirirse diferentes secuencias en una única exploración, se está convirtiendo en una de las modalidades principales en la caracterización de tejidos, así, en combinación con los procesos de cuantificación, permite generar biopsias virtuales ofreciendo, de una manera no invasiva, una visión global del estado del tejido bajo estudio.

Todo esto, permite avanzar en el diagnóstico y seguimiento de pacientes con enfermedades metabólicas y hepáticas mediante la creación de métodos automáticos no invasivos que permitan realizar una caracterización del tejido graso y hepático, para la prevención de enfermedades de peor pronóstico como la cirrosis o el hepatocarcinoma.

Estructura

Este documento se encuentra estructurado en los siguientes apartados:

El Capítulo 1 consta de una recopilación de los conceptos necesarios para el correcto entendimiento de la tesis doctoral. En primer lugar, se exponen las bases clínicas de cada una de las patologías que se van a tratar. A continuación, se realiza una introducción a la imagen médica, detallando las secuencias de RM que se van a emplear para el desarrollo de los diferentes algoritmos. Seguidamente, se introduce el concepto de biomarcador de imagen y se resumen aquellos que se van a emplear para el estudio de las patologías de interés. Finalmente, se exponen los conceptos introductorios al aprendizaje profundo, las redes neuronales convolucionales y el estado del arte de éstas en la resolución de problemas de clasificación y segmentación.

En el Capítulo 2 se expone la hipótesis que desencadena la presente tesis doctoral, así como los objetivos que se plantean para la correcta consecución de la misma.

En el Capítulo 3 se presentan y describen las bases de datos empleadas para el desarrollo de los diferentes algoritmos y se detalla la metodología desarrollada, desde los criterios seguidos para la anotación manual de los casos, pasando por el proceso de definición de arquitectura y el entrenamiento, validación y testeo de los diferentes algoritmos de inteligencia artificial desarrollados, hasta la implementación del algoritmo de cuantificación para la caracterización del parénquima hepático.

En el Capítulo 4 se exponen y discuten los resultados obtenidos en cada uno de los modelos desarrollados. Para cada uno de ellos se muestran los resultados obtenidos durante el proceso de entrenamiento/validación y la capacidad de generalización de dicho modelo mediante la extracción de métricas sobre los subconjuntos de test.

En el Capítulo 5 se presenta cómo se ha realizado la integración de los métodos desarrollados en la práctica clínica. Se muestra la plataforma de análisis de imagen que se ha empleado para su integración y se exponen los diferentes formatos empleados para la exportación de los resultados en formatos estándar.

Finalmente, en el Capítulo 6 se exponen las conclusiones extraídas del trabajo realizado y se proponen las posibles líneas de trabajo futuro que se han abierto con los desarrollos realizados en la tesis doctoral.

Capítulo 1. Introducción

Esta sección introduce todos los conceptos necesarios para el correcto entendimiento de la presente tesis doctoral. Se realiza una breve introducción a las patologías bajo estudio, pasando por las modalidades de imagen que se emplean para su caracterización y con las que se va a trabajar para, finalmente, realizar una introducción a los biomarcadores de imagen, profundizando en aquellos que se van a emplear en los distintos casos de uso. Finalmente se introducen conceptos relacionados con el aprendizaje profundo, y en su aplicación para la resolución de problemas de clasificación y segmentación de imágenes.

1.1. Bases clínicas

En esta sección se profundiza en las bases clínicas de las patologías estudiadas en la presente tesis doctoral. Por ello, se realiza una síntesis del síndrome metabólico, de la enfermedad difusa hepática y de las lesiones focales del hígado, concretamente el tumor hepático o carcinoma hepatocelular (CHC).

1.1.1. Síndrome metabólico y composición corporal

El síndrome metabólico se define como un conjunto de trastornos fisiológicos, bioquímicos y clínicos que afectan a un individuo al mismo tiempo. Entre estos trastornos se incluyen los siguientes:

- Niveles elevados de presión arterial.
- Niveles elevados de glucosa en sangre.
- Exceso de grasa abdominal (obesidad abdominal).
- Niveles elevados de colesterol o triglicéridos.

La presencia de uno de estos factores no implica un riesgo elevado para la salud, sin embargo, presentar varios de ellos aumenta la probabilidad de sufrir enfermedades secundarias como la enfermedad cardiovascular o la diabetes tipo II.

Según la Organización Mundial de la Salud (OMS), desde 1975 la obesidad mundial se ha triplicado. En 2018, más de 1900 millones de adultos tenían sobrepeso, mientras que alrededor de 650 millones padecían de obesidad; 41 millones de niños menores de 5 años padecían sobrepeso u obesidad, ascendiendo a 340 millones entre los 5 y los 19 años.

Capítulo 1. Introducción

Existen diversas guías a la hora de realizar el diagnóstico del síndrome metabólico entre las que encontramos la OMS de 1999, *The National Cholesterol Education Program Adult Treatment Panel III* (NCEP-ATP III) de 2001, la NCEP revisada (NCEP-R) de 2004, la *International Diabetes Federation* (IDF) en 2006, y la *American Association of Clinical Endocrinologists* (AACE) en 2003. Cada una de ellas establece los umbrales de cada uno de los trastornos que comprenden el síndrome metabólico, así como que combinación de ellos hace que un paciente sea diagnosticado con este síndrome (Tabla 1).

Tabla 1. Criterios para establecer el diagnóstico de síndrome metabólico. RCC: Relación Cintura-Cadera; IMC: Índice de Masa Corporal; CC: Circunferencia de cintura; H: hombres; M: mujeres. [1]

	OMS (1999)	NCEP-ATP III (2001)	NCEP-R (2004)	IDF (2006)	AACE (2003)
Obesidad	RCC: >0.90 (H) >0.85 (M) o IMC: > 30kg/m ²	CC: ≥102cm (H) ≥ 88cm (F)	CC: ≥102cm (H) ≥ 88cm (F)	CC: ≥ 94cm (H) ≥ 80cm (F)	IMC: ≥ 25kg/m ²
Triglicéridos	≥ 150mg/dl	≥ 150mg/dl	≥ 150mg/dl o medicación	≥ 150mg/dl o medicación	≥ 150mg/dl
Colesterol HDL	< 35 mg/dl (H) < 39 mg/dl (M)	< 40 mg/dl (H) < 50 mg/dl (M)	< 40 mg/dl (H) < 50 mg/dl (M) o medicación	< 40 mg/dl (H) < 50 mg/dl (M) o medicación	< 40 mg/dl (H) < 50 mg/dl (M)
Presión sanguínea	≥ 140/90 mmHg	≥ 130/85 mmHg o medicación	≥ 130/85 mmHg o medicación	≥ 130/85 mmHg o medicación	≥ 130/85 mmHg o medicación
Glucosa en sangre	≥110mg/dl	≥110mg/dl	≥110mg/dl	≥110mg/dl o diagnóstico	110-126 mg/dl

Capítulo 1. Introducción

	OMS (1999)	NCEP-ATP III (2001)	NCEP-R (2004)	IDF (2006)	AAACE (2003)
				previo de diabetes tipo II	
Otros riesgos	Tasa de excreción de albúmina urinaria $\geq 20 \mu\text{g}/\text{min}$ o relación albúmina/creatinina $\geq 30 \text{mg}/\text{g}$				Antecedentes familiares de diabetes tipo II, hipertensión o accidente cerebrovascular. Síndrome de ovario poliquístico, estilo de vida sedentario, edad avanzada o riesgo de sufrir diabetes o enfermedad cardiovascular.
Diagnóstico	Glucosa en sangre + 2 otros criterios	3 criterios	3 criterios	CC + 2 otros criterios	

Parámetros como el índice de masa corporal (IMC), la circunferencia de cintura (CC) o la relación cintura-cadera (RCC) son parámetros orientativos para definir si un sujeto sufre sobrepeso u obesidad ya que, por ejemplo, para un mismo valor de IMC, la composición corporal entre dos sujetos puede ser distinta. Por ejemplo, dos sujetos pueden tener el mismo peso y altura, sin embargo, uno de ellos poseer un alto porcentaje de grasa corporal, mientras que el otro puede tener un elevado porcentaje de masa muscular. El IMC para ambos sujetos será el mismo, pero no el riesgo de sufrir otras enfermedades como la diabetes tipo II o la enfermedad cardiovascular. Además, no es solo importante conocer el porcentaje de grasa corporal, sino además su distribución. Podemos encontrar grasa intramuscular, grasa subcutánea, grasa visceral o grasa dentro

Capítulo 1. Introducción

de órganos como el hígado. Dependiendo de la distribución de la grasa, el riesgo de sufrir las citadas enfermedades variará. Se ha visto que tener un mayor porcentaje de grasa visceral frente a grasa subcutánea aumenta la probabilidad de sufrir enfermedades secundarias [2]. Esto es debido a que la grasa visceral tiene una mayor cantidad de adipocitos, glucocorticoides y receptores andrógenos que la grasa subcutánea, y, además, es capaz de generar más ácidos grasos. A su vez, la grasa visceral secreta más moléculas bioactivas y está asociada con menor sensibilidad a la insulina y mayor cantidad de triglicéridos circulantes [3, 4].

Son distintas las enfermedades metabólicas en las que el estudio de la composición corporal es importante:

- **Obesidad visceral:** Se mide con la ratio entre la grasa visceral y la subcutánea. Cuanto mayor es la ratio, mayor es la probabilidad de sufrir obesidad visceral y, por tanto, aumento de la resistencia a la insulina.
- **Sarcopenia:** Se define como una disminución en la masa muscular. La sarcopenia está relacionada con la edad principalmente, pero también con una actividad física baja, desnutrición, mala absorción, enfermedad endocrina, enfermedades neurodegenerativas o caquexia, común en determinadas enfermedades crónicas.
- **Obesidad sarcopénica:** Es una combinación de las dos anteriores. Estilos de vida sedentarios, cambios hormonales o malnutrición favorece el desarrollo de esta enfermedad. Además, el tejido adiposo secreta citoquinas pro-inflamatorias que contribuyen a la pérdida de masa muscular.

La composición corporal se define como la cantidad y distribución de los distintos compartimentos corporales. Distintos son los modelos que se han ido empleando hasta la actualidad para definir la composición corporal diferenciados en el número de compartimentos empleados:

- 2 compartimentos: masa grasa y masa libre de grasa.
- 3 compartimentos: masa grasa, agua corporal, masa magra.
- 4 compartimentos: masa grasa, agua corporal, proteínas y hueso.
- 5 compartimentos: subdivide el estudio de la composición corporal en cinco niveles de estudio de distinta complejidad estructural:
 - Nivel I (nivel atómico o elemental): oxígeno, carbono, hidrógeno, calcio, nitrógeno y otros.

Capítulo 1. Introducción

- Nivel II (nivel molecular o químico): agua, lípidos, proteínas, minerales y otros.
- Nivel III (nivel celular): masa celular, fluido extracelular, sólidos extracelulares y células adiposas
- Nivel IV (nivel histológico o tisular): músculo, tejido adiposo, sangre, hueso y otros.
- Nivel V (nivel corporal): incluye todo el cuerpo humano.

Distintas son las técnicas que se emplean para la medida de la composición corporal de manera no invasiva, a continuación, se van a presentar las más empleadas.

Análisis por impedancia bioeléctrica (BIA)

Método para la medición del porcentaje de grasa corporal de un individuo. Para ello, se sitúan dos electrodos en la mano y pie derechos del sujeto. A continuación, una pequeña corriente eléctrica, imperceptible por el sujeto, se envía a través del cuerpo de éste. El BIA mide la restricción al paso de dicha corriente a través del cuerpo debido a los distintos tejidos que encuentra a su paso. Tejidos con grandes cantidades de fluido y electrolitos, como la sangre, presentan elevada conductividad, sin embargo, otros tejidos como la grasa y el hueso presentan mayor resistencia al paso de la corriente. Por lo tanto, el BIA permite realizar una estimación de la cantidad de líquido en el cuerpo, a partir de la cual, a través de ecuaciones matemáticas, se calcula el porcentaje de grasa corporal.

Las principales ventajas de este método es la facilidad de uso, no requiere de formación previa para su correcta utilización y la rapidez en la medida. Sin embargo, la principal desventaja es la cantidad de sesgos que pueden aparecer si la medida no se realiza en las condiciones adecuadas. Para comparar dos medidas hay que asegurarse que el/los sujetos se encuentran en las mismas condiciones de hidratación, con la vejiga vacía y con similar temperatura corporal ya que la vasodilatación cutánea produce una disminución de la resistencia, mientras que una vasoconstricción produce un aumento. En las mujeres también es importante tener en cuenta el ciclo menstrual ya que éste produce variaciones en los fluidos corporales. Finalmente, hay que tener en cuenta que las dos asunciones principales del BIA (el cuerpo – extremidades y tronco – se puede considerar como un único cilindro conductor y la relación entre las principales áreas transversales se mantiene constante) se ven alteradas cuando tratamos con sujetos de edad más avanzada ya que la disminución en la cantidad de tejido muscular esquelético y la redistribución del tejido

Capítulo 1. Introducción

adiposo de las extremidades al tronco dan lugar a diámetros más estrechos para los volúmenes conductores de las extremidades. Por lo tanto, para el cálculo correcto de la composición corporal en sujetos de edad avanzada hay que emplear ecuaciones extraídas con sujetos dentro de este rango de edad en lugar de emplear ecuaciones extraídas de población general.

Peso hidrostático

Mide la densidad corporal de una persona empleando el principio de Arquímedes de desplazamiento del agua, según el cual, la fuerza de flotación de un objeto sumergido es igual al peso del fluido que es desplazado por el objeto. Por lo tanto, dado que el tejido libre de grasa, como el hueso o el músculo, es más denso que el agua, y el tejido graso es menos denso que el agua, una persona con mayor proporción de grasa corporal pesará menos bajo el agua y tendrá una mayor flotación que otra con una mayor proporción de músculo, que pesará más debajo del agua.

Para medir la densidad corporal, en primer lugar, se pesa al sujeto con una báscula fuera del tanque de agua. A continuación, éste se sumerge en el tanque de agua sentado en una silla, se le pide que expulse todo el aire posible de los pulmones y se mide su peso varias veces (para minimizar errores). Todos los valores obtenidos se promedian para obtener un valor final. Finalmente, a través de la fórmula matemática definida según la Ecuación 1, se calcula la densidad corporal del sujeto:

$$D = \frac{M_a}{\frac{M_a - M_w}{D_w} - RV} \quad (1)$$

Donde M_a es la masa del individuo fuera del tanque, M_w es la masa del individuo en el agua, D_w hace referencia a la densidad del agua, y RV se corresponde con el volumen residual (volumen de aire residual en los pulmones + volumen de aire residual en las vías respiratorias tras una exhalación completa).

Una vez se ha calculado la densidad corporal, se puede calcular el porcentaje de grasa corporal siguiendo la ecuación de Brozek et al. [5], definida por la Ecuación 2.

$$Grasa (\%) = \left(\frac{4,570}{D} - 4,141 \right) \cdot 100 \quad (2)$$

Capítulo 1. Introducción

Esta técnica se ha visto que presenta más precisión que otras técnicas como el análisis por bioimpedancia, la evaluación del pliegue cutáneo u otras medidas antropométricas (circunferencia de cintura, pliegues cutáneos, etc.). Sin embargo, tiene una serie de inconvenientes como son la infraestructura que conlleva, determinados sujetos no se ven capaces de exhalar todo el aire mientras se encuentran sumergidos en la piscina o, dado que el tanque debe limpiarse tras cada sujeto, la cantidad de mediciones diarias frente a otras técnicas se ve disminuido.

Pletismografía por desplazamiento de aire (ADP)

Al igual que el peso hidrostático, la ADP permite medir la densidad corporal de un sujeto. En este caso, se mide el desplazamiento de aire en lugar del desplazamiento de agua, basándose en la Ley de Boyle, según la cual, a temperatura constante, el volumen y la presión están inversamente relacionados (Ecuación 3).

$$\frac{P_1}{P_2} = \frac{V_2}{V_1} \quad (3)$$

V_1 y P_1 hacen referencia al volumen y la presión, respectivamente, de la cámara pletismográfica cuando se encuentra vacía, mientras que V_2 y P_2 representan el volumen y la presión de la cámara cuando el sujeto se encuentra dentro de ella.

La cámara pletismográfica más conocida es la *Bod Pod*. Para la correcta medición de la densidad corporal, hay que minimizar los efectos isotérmicos de la ropa, cabello, volumen de aire en los pulmones y vías aéreas y superficie corporal. Por esto, se introduce al sujeto con ropa de baño y con un gorro en la cabeza para comprimir el cabello. A partir de la altura y peso del sujeto se hace una estimación de su superficie corporal. Finalmente, se realiza una estimación del volumen de aire en las vías aéreas y pulmones del sujeto. Toda esta información se emplea por la cámara pletismográfica para minimizar los sesgos.

Este método, frente a la técnica del peso hidrostático, ofrece mayor comodidad para el sujeto.

Absorciometría de Rayos X de doble energía (DXA)

La DXA es una modalidad espacial de imagen que requiere de equipamiento específico ya que necesita un filtrado especial del haz de rayos X y un registro espacial preciso entre

Capítulo 1. Introducción

las dos atenuaciones de energía empleadas, por lo tanto, un sistema de rayos X convencional no permite hacer este tipo de estudios.

Esta técnica, empleando los rayos X, permite medir la composición corporal. Se trata de una exposición mínima a esta radiación ($<10\mu\text{Si}$), siendo 10 veces inferior a la recibida con un estudio de rayos X convencional.

La DXA permite calcular la composición corporal de un individuo extrayendo información relacionada con la cantidad de grasa, hueso y masa magra corporal. Además, divide el cuerpo en brazos, piernas, tronco y cabeza, ofreciendo información de cada componente en cada región individualizada.

Los sistemas DXA, desde que aparecieron en los años 90, han ido evolucionando para adaptarse a todo tipo de sujetos independientemente de su altura, anchura y peso.

Hasta 2009, los datos de referencia empleados para situar al sujeto en un rango o en otro, eran propiedad de cada fabricante y diferentes entre sí. Sin embargo, en 2009 la *National Health and Nutrition Examination Survey* (NHANES) publicó datos de referencia obtenidos de la población estadounidense, siendo posteriormente integrados por los sistemas comerciales para generar los distintos rangos z-score, normalizando, de este modo, el resultado ofrecido entre los distintos fabricantes.

Además, algunos sistemas DXA ofrecen una estimación de la cantidad de grasa visceral, permitiendo tener información, no solo de la cantidad de grasa corporal, sino también de su distribución.

Por lo tanto, es una técnica que ofrece elevada precisión en la medida de la composición corporal, sin embargo, se ha visto que, pese a presentar una elevada correlación con los resultados obtenidos con otras técnicas como la TC, la DXA tiende a subestimar la masa total de grasa corporal [6]. Además de para cuantificar la composición corporal, la DXA es ampliamente utilizada para realizar un análisis de la densidad mineral ósea, especialmente útil en pacientes con osteoporosis.

Técnicas de imagen volumétricas

Tanto la TC como la RM permite realizar un estudio volumétrico preciso de la composición corporal de un individuo. Al tratarse de técnicas tomográficas, es decir, adquieren un conjunto de cortes transversales del cuerpo humano, permiten realizar un estudio tridimensional del mismo. Ya no se trata únicamente de una imagen 2D, como el caso de la DXA, donde en una única imagen se observan distintos tejidos superpuestos,

sino que se trata de una imagen 3D, donde en un vóxel (píxel 3D) tenemos información de una región muy concreta del cuerpo.

Por lo tanto, gracias a este tipo de técnicas podemos, no solo medir la cantidad de grasa corporal, sino además su distribución, pudiendo cuantificar la cantidad de grasa visceral, subcutánea e intramuscular e incluso medir la fracción de grasa en determinados órganos como el hígado.

Estas técnicas las detallaremos más adelante y serán las que emplearemos para realizar el estudio de la composición corporal, siendo además objeto de esta tesis doctoral.

1.1.2. Enfermedades difusas hepáticas

Comprenden todas aquellas enfermedades que afectan a las células funcionales del hígado, los hepatocitos, alterando, de este modo, la función hepática. Estas incluyen infecciones, infiltración de grasa y fibrosis. En estos procesos, los hepatocitos se ven sustituidos por adipocitos y tejido fibroso.

La enfermedad de hígado graso no alcohólico (EHGNA, en inglés NAFLD) es una afección reversible originada por la acumulación de triglicéridos en los hepatocitos. El alcoholismo, la obesidad, y la diabetes son las causas más comunes de esta enfermedad. Este estado del hígado es reversible si se cambia la dieta del paciente, sin embargo, si este no se cuida, la enfermedad puede ir avanzando hacia estadios más severos, desencadenando fibrosis, cirrosis e incluso carcinoma hepatocelular (CHC).

La EHGNA afecta al 25% de la población global adulta, variando desde el 13,5% en África hasta el 31,8% en Oriente Medio [7]. Esta enfermedad se caracteriza por la presencia de grasa en las células hepáticas por causa diferente al consumo de alcohol, drogas, hepatitis u otras enfermedades hepáticas crónicas. La presencia de grasa induce estrés oxidativo de las células hepáticas provocando su daño y muerte. Esta enfermedad varía desde estadios más leves donde únicamente hay presencia de esteatosis hepática hasta estadios más avanzados como la esteatohepatitis no alcohólica, donde además de esteatosis hepática hay daño de los hepatocitos.

La sobrecarga de hierro hepática se encuentra comúnmente en enfermedades hepáticas como la EHGNA o la esteatohepatitis no alcohólica. Esta aparición de hierro en enfermedades hepáticas no alcohólicas es debida a la inducción de la hepcidina por la inflamación del hígado y el estrés oxidativo [8], tanto en los hepatocitos como en las células de Kupffer.

Capítulo 1. Introducción

Cuando hay coexistencia de hierro y grasa en el hígado, el estrés oxidativo producido se multiplica, aumentando la probabilidad de desencadenar fibrosis y aumentando el riesgo de desarrollo de CHC [9,10].

Actualmente, el *gold standard* en el diagnóstico de las enfermedades difusas hepáticas es la biopsia, sin embargo, este método presenta una serie de inconvenientes, entre los que se encuentran:

- Es un método invasivo.
- Puede conllevar complicaciones como sangrado, infección y muerte [11].
- La muestra se encuentra sesgada. Se toma una pequeña muestra del lóbulo hepático VI, sin embargo, la distribución de la grasa y el hierro en el hígado es heterogénea [12-14], por lo tanto, puede que la muestra tomada no sea representativa de la cantidad de grasa y hierro reales.
- El diagnóstico se basa en un sistema de gradación histológica, establecido a través de la percepción visual del hepatólogo. Por lo tanto, está sujeto a una elevada variabilidad intra- e inter- observador [15,16].
- Debido al carácter invasivo de la prueba, se dificulta el seguimiento de los pacientes y su aplicación en el entorno de investigación.

Por lo tanto, se buscan métodos diagnósticos no invasivos que reduzcan todos estos sesgos. Por ello, cada vez más se emplean métodos basados en RM para el diagnóstico y seguimiento de pacientes con enfermedad difusa hepática.

Dentro de la RM, la espectroscopia es la técnica *gold standard* en la cuantificación de grasa, ya que es precisa, reproducible y es muy sensible a pequeñas variaciones en los depósitos de grasa. Sin embargo, presenta una serie de limitaciones como es su elevado coste, se requieren secuencias de pulso dedicadas y, por tanto, no es accesible para cualquier centro, requiriéndose además de programas específicos para el análisis de la señal adquirida, por lo tanto, únicamente se realiza en hospitales especializados.

Estas limitaciones hacen que se recurra a técnicas de RM basadas en imagen que permiten realizar una biopsia virtual hepática, de forma que sea posible la cuantificación simultánea del hierro y la grasa en cada región del hígado. Las secuencias de RM empleadas para ello se comentarán en detalle en el apartado 1.2.2. *Resonancia magnética para el estudio de la enfermedad difusa hepática.*

1.1.3. Carcinoma hepatocelular (CHC)

El CHC es un tumor maligno cuyo origen se encuentra en el hígado. Se trata del 6º tipo de cáncer más diagnosticado y el 4º que más mortalidad produce, después del cáncer pulmonar, colorrectal y de estómago [17]. Los principales factores de riesgo es la presencia de cirrosis, enfermedades hepáticas víricas (hepatitis B y C), enfermedades metabólicas y/o consumo de alcohol. Se estima que un tercio de los pacientes con cirrosis desarrollarán CHC [18].

La incidencia del CHC es heterogénea, el 72% de los casos se dan Asia, 10% en Europa, 7,8% en África, 5,1% en Norteamérica, 4,6% en Latinoamérica y el 0,5% en Oceanía [17].

En un hígado con daño crónico se produce una muerte elevada de hepatocitos que estimula la regeneración de las células dando lugar a tejido fibrótico, que en sus casos más severos produce cirrosis, estado de daño hepático irreversible. El aumento de la actividad celular en un proceso fibrótico aumenta la probabilidad de que se produzca una mutación y la transformación de lesiones hepáticas pretumorales en CHC. Conforme avanza la enfermedad tumoral, el CHC metastatiza a otras regiones del hígado y a otros órganos entre los que encontramos los pulmones (55%), nódulos linfáticos (53%), hueso (28%), glándulas adrenales (11%), peritoneo y/o omento (11%) y cerebro (2%) [19].

La principal técnica diagnóstica empleada hoy día para el diagnóstico de CHC son las técnicas de imagen, siendo la ecografía y la RM, por su carácter no ionizante, las más empleadas. La RM tiene una gran sensibilidad y especificidad en el diagnóstico de CHC, sin embargo, por su elevado coste y baja disponibilidad, se emplea para confirmar el diagnóstico y para la estadificación de la lesión una vez se ha detectado esta por ultrasonidos [20].

Los bordes de la lesión definen la agresividad del tumor, de esta forma, en sus fases iniciales el tumor se presentará como una agrupación de células sin cápsula mientras que en fases más avanzadas veremos tumores más grandes con un halo bien delimitado o con bordes imprecisos. La presencia de bordes imprecisos está relacionada con tumores más agresivos ya que suelen ser infiltrativos. Los tumores con los bordes bien definidos suelen tener mejor pronóstico.

Dado que se trata de un tumor muy vascularizado, el estudio de RM más empleado para el diagnóstico es la técnica de perfusión con contraste intravenoso, que veremos más en detalle en el apartado *1.2.3. Resonancia magnética para el estudio del carcinoma*

Capítulo 1. Introducción

hepatocelular. El comportamiento típico del CHC es un realce de contraste en la imagen correspondiente a la fase arterial, presentando una hiperintensidad en la lesión con respecto al resto del parénquima hepático, y un lavado rápido del contraste en la fase venosa, mostrando una señal más hipointensa (Figura 1).

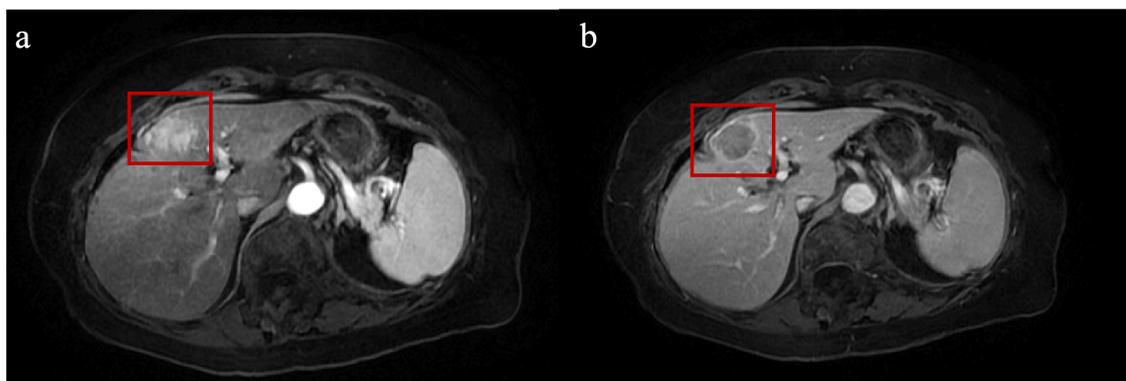


Figura 1. Estudio de RM de perfusión en un paciente con CHC. En rojo se resalta la región tumoral. a) Fase arterial: se puede observar hiperintensidad en la lesión. b) Fase venosa: se puede observar hipointensidad en la lesión.

1.2. Imagen médica

En esta sección se detallan los protocolos de adquisición necesarios para poder llevar a cabo los desarrollos y casos de uso concretos presentados en este trabajo (estudio de la composición corporal, estudio de enfermedades difusas hepáticas y estudio de lesiones focales en el hígado).

1.2.1. Tomografía computarizada para el estudio de la composición corporal

La TC se puede emplear para medir la cantidad de músculo, grasa intramuscular, grasa visceral y grasa subcutánea de un paciente (Figura 2). Generalmente se suele emplear un único corte transversal al nivel de la tercera vértebra lumbar (L3) para realizar estas mediciones ya que, a este nivel, se ha demostrado que existe una elevada correlación con las proporciones reales de grasa y músculo del cuerpo completo [21]. A nivel de L3 encontramos el psoas, los músculos paraespinales y los músculos de la pared abdominal, por lo tanto, se convierte en un corte apropiado para el estudio de la masa muscular. Sin embargo, en el estudio de la grasa puede haber diferencias debidas al sexo, edad o nivel corporal, por lo tanto, hay estudios que sugieren analizar la grasa en distintos niveles anatómicos [22]. Además, dado que el contenido del tracto gastrointestinal está en continuo movimiento, no hay garantía de que dos cortes a nivel de L3, adquiridos en dos instantes diferentes, incluyan la misma anatomía, por lo tanto, resulta de interés realizar un estudio tridimensional de la composición corporal en lugar de analizar un único corte bidimensional.

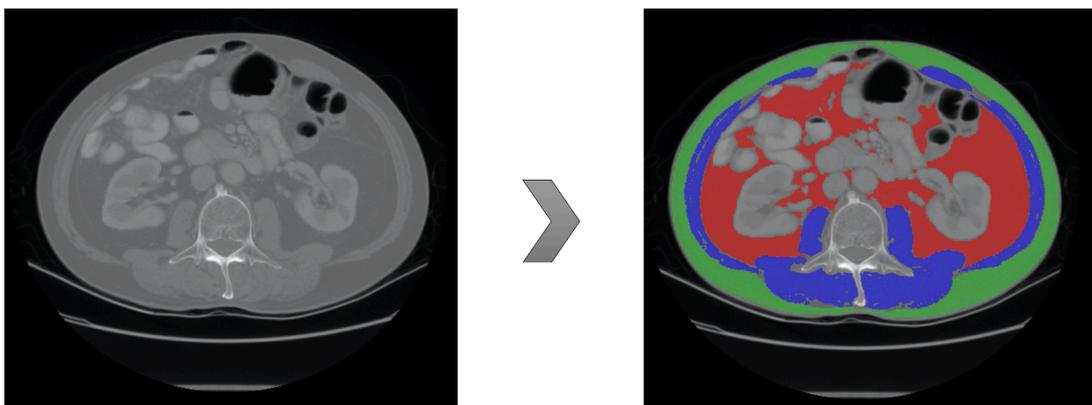


Figura 2. Segmentación grasa visceral (rojo), subcutánea (verde) y músculo (azul) en un corte transversal de TC a nivel de la vértebra L3.

Debido a la normalización de los valores de atenuación de cada vóxel en unidades Hounsfield (UH), podemos diferenciar aquellas regiones que se corresponden con tejido adiposo y que tienen valores entre -190 y -50 UH (concretamente entre -150 y -50 UH para la grasa visceral y entre -190 y -50 para la grasa subcutánea e intramuscular) de aquellos que pertenecen a tejido muscular y tiene valores entre -29 y 150 UH.

En el estudio y seguimiento de determinadas patologías, como los procesos oncológicos, la TC se adquiere rutinariamente, por lo tanto, para el estudio de la composición corporal no sería necesaria la adquisición de una nueva imagen y, por tanto, no aumenta la radiación emitida al paciente. Además, procesos metabólicos como la obesidad visceral, la sarcopenia o la obesidad sarcopénica son factores adversos en los pacientes oncológicos ya que afectan a la eficacia y toxicidad de la quimioterapia [23-25] y, por lo tanto, el control de éstos tiene un elevado interés.

1.2.2. Resonancia magnética para el estudio de la enfermedad difusa hepática

En este apartado se propone una de las secuencias de RM necesarias para la caracterización de la enfermedad difusa hepática, especialmente una de las utilizadas para la cuantificación del hierro y la grasa en hígado.

Para la cuantificación simultánea tanto del hierro como de la grasa hepáticas se emplea una secuencia eco de gradiente multi-eco con desplazamiento químico (MECSE-MR) [26].

En esta secuencia aparecen dos términos clave: desplazamiento químico y secuencia multi-eco:

1. **Desplazamiento químico:** los protones de grasa precesan a una frecuencia inferior a la de los protones de agua. Las secuencias de desplazamiento químico adquieren primero una imagen con un tiempo de eco (TE) en la que los protones de agua y grasa se encuentran en fase opuesta, y una segunda imagen con un TE en el que éstos se encuentra en fase. El TE dependerá de la intensidad del campo magnético estático (B_0). Para un B_0 de 1,5T, los protones de agua y grasa están en fase opuesta cada 2,3 ms y en fase cada 4,6 ms. Para un B_0 de 3T, se encuentran en fase opuesta cada 1,15 ms y en fase cada 2,3 ms. La imagen en la que los protones se encuentran en fase es el resultado de la suma de la señal del agua y de la grasa, mientras que la imagen en fase opuesta resulta de la diferencia entre ambas componentes. Por lo tanto, en hígados con presencia de grasa (esteatosis),

como el caso de la Figura 3, la imagen en fase opuesta presentará una caída de la señal (imagen más hipointensa) con respecto a la imagen en fase.

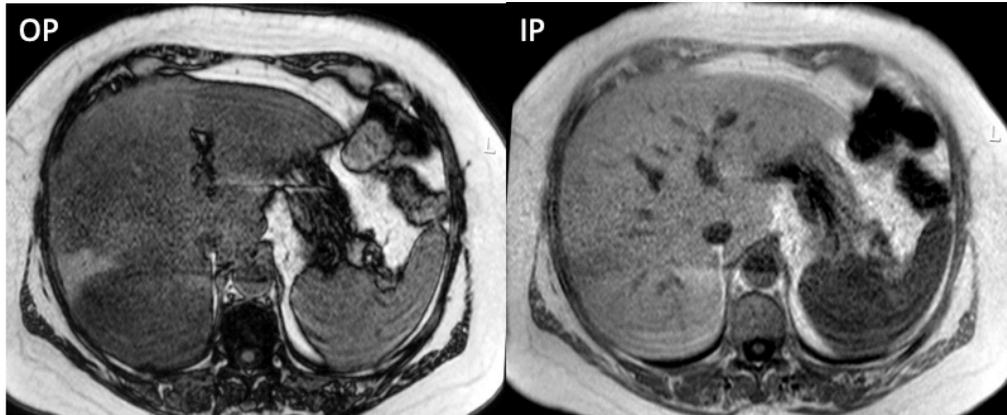


Figura 3. Caso de un paciente con esteatosis. A la izquierda vemos la imagen en fase opuesta (OP, del inglés out-of-phase) donde se observa el hígado más hipointenso que la imagen en fase (IP, del inglés in-phase), a la derecha.

2. **Multi-eco:** la presencia de hierro en el hígado, debido al efecto paramagnético de éste, produce tiempos de relajación $T2^*$ más cortos, lo que se traduce en caída de la señal más rápida. El $T2^*$ o tiempo de relajación transversal, es la constante de tiempo que determina la velocidad a la que los protones excitados alcanzan el equilibrio o se desfasan entre sí debido a las interacciones espín-espín y a la inhomogeneidad del campo. Por lo tanto, la presencia de hierro hepático produce una caída de la señal en la imagen en fase con respecto a la imagen en fase opuesta adquirida anteriormente. Además, si aparte de hierro, hay presencia de grasa en el hígado, la caída de la señal en la imagen en fase opuesta (producida por la grasa) puede pasar desapercibida si también se obtiene caída de la señal en la imagen en fase (debido al hierro), produciéndose, por tanto, una subestimación de la medida de grasa [27], como es el caso de la Figura 4. Por lo tanto, es necesario tener en cuenta la caída del $T2^*$ a través de una secuencia eco de gradiente multi-eco.

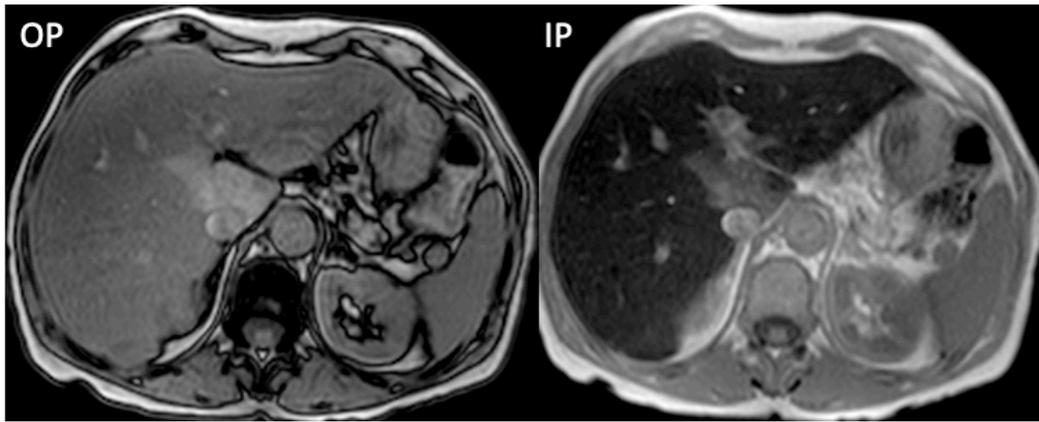


Figura 4. Caso de un paciente con coexistencia de grasa y hierro en el hígado. Se observa un hígado más hipointenso en la imagen en fase (derecha) que en la imagen en fase opuesta (izquierda) debido a la caída de la señal por la presencia de hierro y, por tanto, no se aprecia la existencia de esteatosis.

Una secuencia eco de gradiente multi-eco, permite generar varios ecos en cada intervalo de tiempo de repetición (TR). Esto se consigue activando y desactivando el gradiente de codificación de frecuencia con distintas polaridades (Figura 5), desfasando y refasando los espines. Para el correcto modelado del T2*, en el análisis hepático, se adquieren un mínimo de 3 tiempos de eco (idealmente entre 6 y 12). El primer TE debe tener un valor de aproximadamente 1,15 ms para una RM de 1,5T o de 0,99 ms para una RM de 3T para garantizar que todavía no se ha producido una caída de la señal debido a la presencia de hierro en el hígado, especialmente en hígado con elevada sobrecarga de hierro.

Además del sesgo introducido por el efecto del T2* (caída de la señal en presencia de hierro), que, como hemos comentado, se minimiza gracias al uso de secuencia eco de gradiente multi-eco, hay que tener en cuenta el sesgo que puede introducir en la cuantificación de la grasa el hecho de que el agua y la grasa tengan tiempos de relajación longitudinal T1 diferentes (el T1 de la grasa es menor que el del agua) [28]. Este efecto se puede minimizar, reduciendo al máximo el ángulo de magnetización del pulso de RF. En secuencias 2D, con un TR mayor de 100 ms, el ángulo de magnetización debe de ser menor a 10°. Dado que la relación señal-ruido disminuye considerablemente con ángulos de magnetización bajos [28], el TR debe de ajustarse adecuadamente para compensarlo.

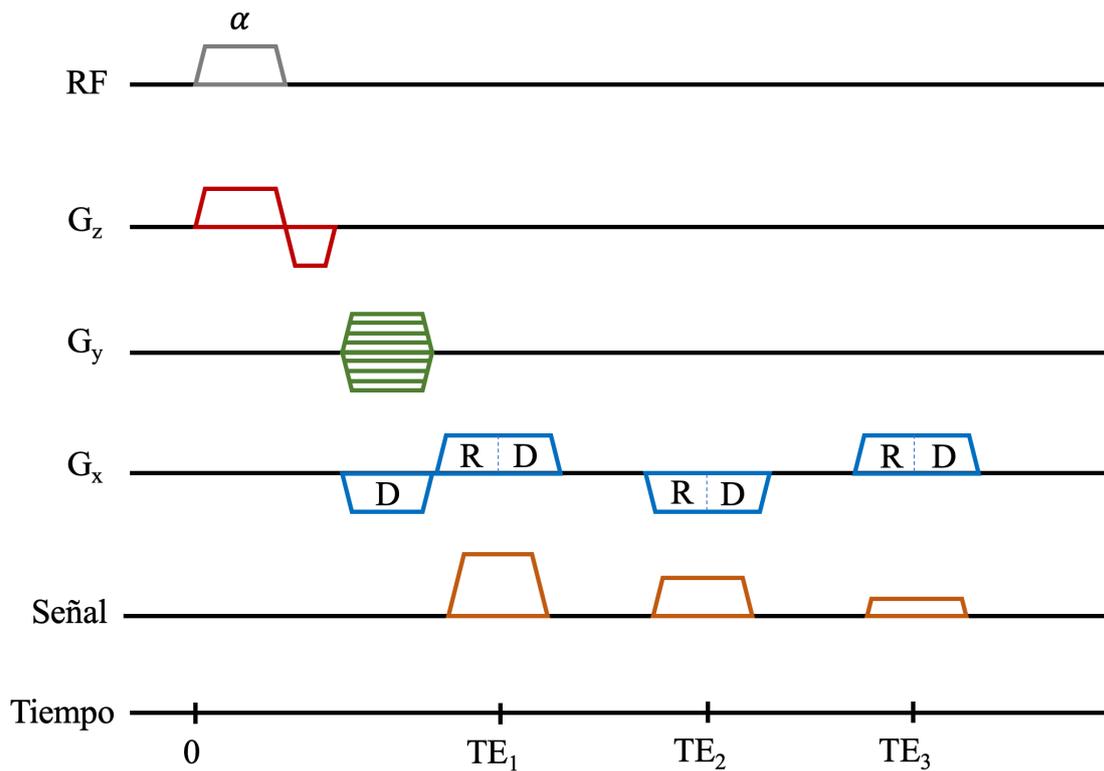


Figura 5. Diagrama de la secuencia eco de gradiente multi-eco. Tras la aplicación del pulso de radiofrecuencia se adquieren distintas señales a distintos tiempos de eco. El gradiente de codificación de frecuencia se activa sucesivas veces con distintas polaridades, desfasando (D) y refasando (R) los espines.

1.2.3. Resonancia magnética para el estudio del carcinoma hepatocelular

Para el estudio y caracterización del CHC con RM, dado que se trata de un tumor con elevada vascularización, la principal secuencia empleada es la perfusión con realce de contraste (DCE, del inglés *Dynamic Contrast Enhanced*). Se trata de una secuencia dinámica eco de gradiente potenciada en T1 adquirida sucesivas veces a lo largo del tiempo, esto es, antes, durante y después del suministro de un agente de contraste (generalmente Gadolinio).

Se suelen adquirir cuatro instantes o fase temporales principales (Figura 6):

- **Fase inicial:** adquisición antes del suministro del agente de contraste que se toma como referencia para el resto de las fases.
- **Fase arterial:** adquisición en el momento del paso del bolo de contraste por el sistema arterial, concretamente por la aorta.
- **Fase venosa:** adquisición en el momento del paso del bolo de contraste por el sistema venoso, concretamente por la vena porta.

- **Fase tardía:** adquisición tras el paso del agente de contraste. Es común la adquisición de varios volúmenes en fase tardía a distintos tiempos para el análisis del lavado del agente de contraste.

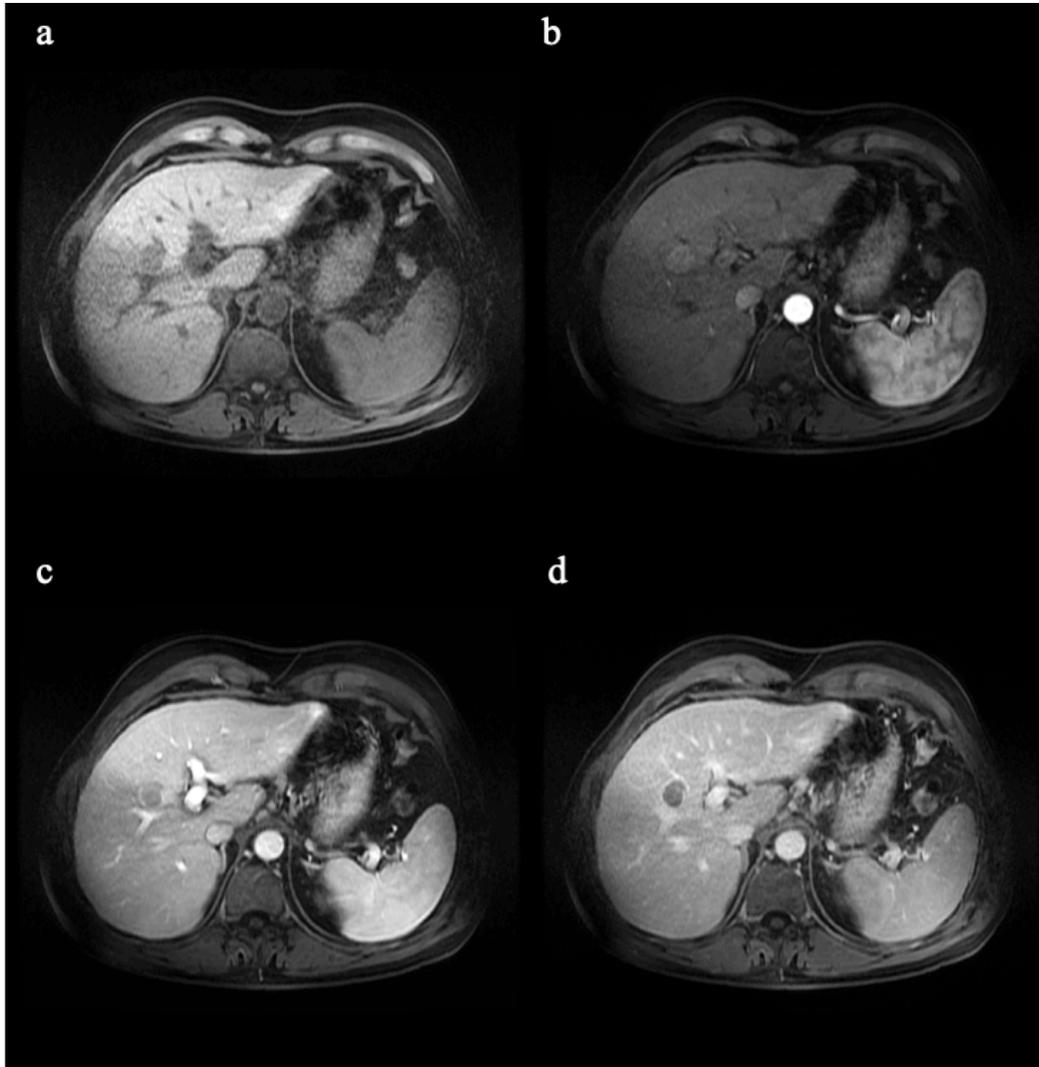


Figura 6. Serie de perfusión hepática DCE de un paciente con CHC. a) Fase temprana. b) Fase arterial. c) Fase venosa. d) Fase tardía.

Además de la secuencia de perfusión, otra secuencia común en el estudio del CHC es la secuencia de difusión (DWI, del inglés *Diffusion Weighted Imaging*). Se trata de una secuencia de RM diseñada para medir el movimiento browniano de las moléculas de agua. En los tejidos biológicos, la movilidad de las moléculas de agua se puede evaluar en el espacio intracelular, extracelular e intravascular. Generalmente, un tejido sano presenta una alta difusividad de las moléculas de agua, éstas pueden moverse libremente, sin embargo, en un tejido tumoral, hay una reducción del espacio intersticial y, por tanto, una disminución de la difusividad de las moléculas de agua. Esta reducción del espacio

Capítulo 1. Introducción

intersticial se produce por diferentes procesos que se producen en los tejidos tumorales: aumento de la celularidad, angiogénesis, inflamación, necrosis o fibrosis.

La secuencia DWI (Figura 7) es una secuencia eco de espín con eco planar (SE-EPI, del inglés, *spin-echo echo-planar imaging*) potenciada en T2 a la que se le han añadido dos gradientes de difusión, aplicados sobre el mismo eje, antes y después de aplicar el pulso de RF de 180° (codificación de difusión Stejskal-Tanner [29]). En moléculas con poco movimiento, estos gradientes de difusión no tendrán efecto, sin embargo, en moléculas en movimiento, el campo magnético que éstas perciben irá variando, dependiendo del lugar en el que se encuentren, afectando a su fase y, por lo tanto, en el momento de capturar la señal, el desfase acumulado generará una disminución en la intensidad de la señal. El grado de difusión está determinado por el valor b, que dependerá de las características del gradiente de difusión (Ecuación 4)

$$b = \gamma^2 \cdot G^2 \cdot \delta^2 \cdot \left(\Delta - \frac{\delta}{3} \right) \quad (4)$$

Donde γ es la constante giromagnética, G es la intensidad del gradiente, δ es la duración del gradiente y Δ es el intervalo de tiempo entre los dos gradientes.

Capítulo 1. Introducción

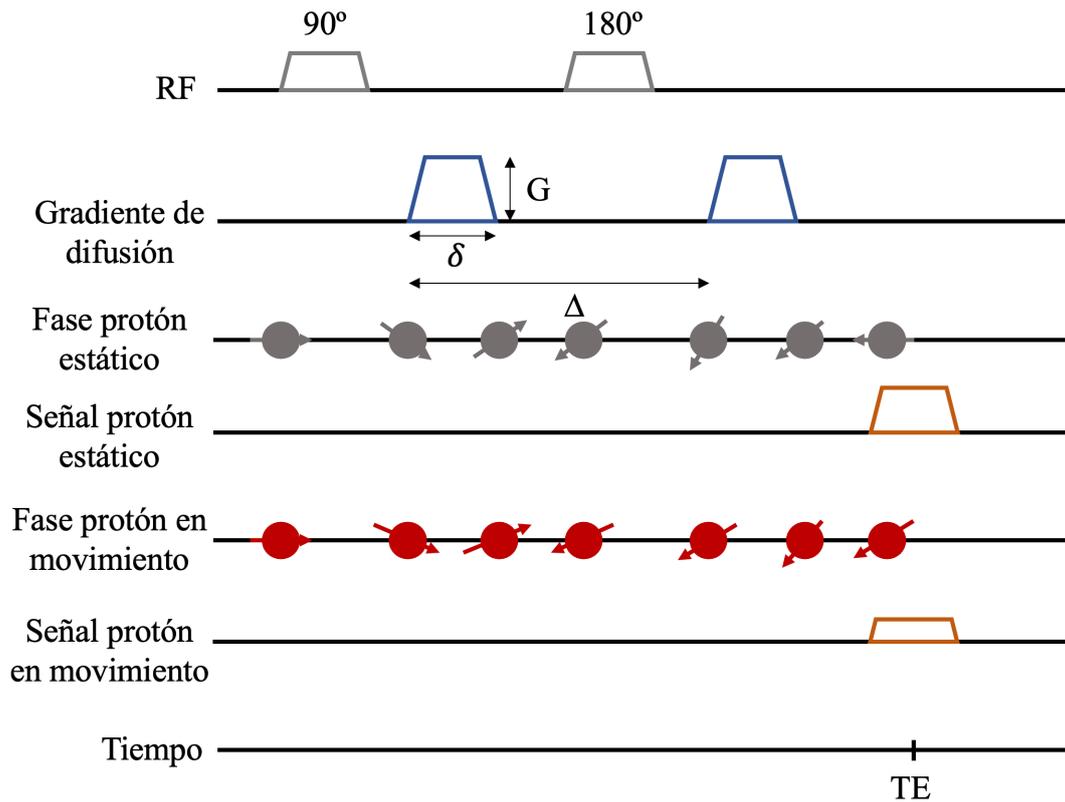


Figura 7. Diagrama de la secuencia de difusión (DWI). Secuencia eco de espín a la que se le añaden dos gradientes de difusión, uno antes del pulso de RF 180° y otro después. Los protones estáticos producen más señal que los protones en movimiento.

Cuanto mayor es el valor b , la señal de RM disminuye, siendo menor esta disminución en las regiones en las que el movimiento de las moléculas de agua está más restringido. Es por esto por lo que, como se observa en la Figura 8, en imágenes adquiridas con valores b altos, se observan regiones hiperintensas en tejidos tumorales como un cáncer de mama.

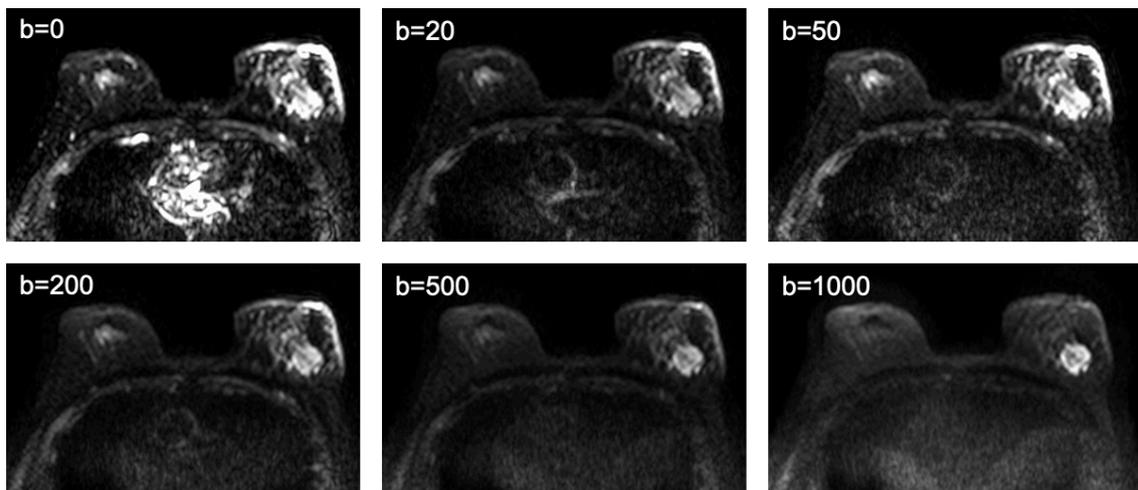


Figura 8. Secuencia de difusión en una paciente con cáncer de mama adquirida con distintos valores b . A valores b bajos vemos diversas zonas hiperintensas, sin embargo, a valores b altos se pueden observar hiperintensidades únicamente en las regiones con restricción del movimiento de las moléculas de agua (tumor).

Capítulo 1. Introducción

Tanto la secuencia de perfusión como la secuencia de difusión requieren la adquisición de distintos volúmenes. En el caso de las secuencias de difusión, el incremento de la resolución espacial implicaría el uso de mayores factores EPI, y por tanto imágenes más sensibles a la distorsión geométrica, limitando así la resolución espacial de este tipo de imágenes. En el caso de las secuencias de perfusión, para poder adquirir varios volúmenes sin alargar excesivamente el tiempo de adquisición se sacrifica la resolución espacial de las imágenes. Por estos motivos, junto a las secuencias de difusión y perfusión se adquieren otras secuencias con mayor resolución espacial, como son las secuencias potenciadas en T1 y T2 (secuencias anatómicas).

1.3. Biomarcadores de imagen

En esta sección se introduce al concepto de biomarcador de imagen y los distintos pasos necesarios para el correcto diseño, implementación y validación de un biomarcador. A continuación, veremos los biomarcadores de imagen principales que se emplean en la presente tesis doctoral, yendo desde la cuantificación de la grasa y el hierro para la caracterización de pacientes con enfermedad difusa hepática, hasta el análisis de difusión y perfusión, de especial interés en pacientes con CHC.

1.3.1. ¿Qué es un biomarcador de imagen?

Los biomarcadores de imagen son parámetros obtenidos objetivamente mediante la aplicación de técnicas de procesamiento de imagen, modelado matemático e inteligencia artificial, que se relacionan con procesos fisiológicos o patológicos que ocurren en el organismo. Estos biomarcadores permiten medir pequeños cambios que ocurren en el cuerpo humano incluso antes de que sean perceptibles por el ojo humano, por lo tanto, poseen un elevado valor diagnóstico, pronóstico y de evaluación del tratamiento [30, 31]. El proceso de desarrollo de un biomarcador de imagen (Figura 9) se puede subdividir en tres fases diferentes [32]:

1. **Validez inicial:** Esta fase se subdivide a su vez en dos pruebas con el objetivo de justificar y demostrar el impacto que puede tener el biomarcador de imagen sobre un determinado proceso fisiológico o patológico. Por lo tanto, en esta fase tenemos:
 - Prueba de concepto: definición del objetivo principal al que se quiere llegar gracias al desarrollo del nuevo biomarcador de imagen. Verificar que la respuesta clínica a responder se puede alcanzar con el uso de técnicas de imagen y postproceso.
 - Prueba de mecanismo: determinación de la relación que existe entre el biomarcador de imagen y el proceso biológico a estudiar. Por ejemplo, el coeficiente aparente de difusión tiene una relación inversamente proporcional al nivel de celularidad tisular.
2. **Validez técnica:** Esta etapa engloba todos los pasos comprendidos entre la adquisición de la imagen hasta la cuantificación y se subdivide en diferentes pasos:

Capítulo 1. Introducción

- Adquisición de imagen: el primer paso es determinar el protocolo de adquisición necesario, es decir, tipo de modalidad (TC, RM, rayos x, ultrasonidos, etc.), tipo de secuencia/serie, si se necesita contraste o no, etc. Además, es importante tener en cuenta una serie de aspectos, como la calidad de la imagen, medida a través de la relación señal-ruido y la relación contraste-ruido; la resolución espacial, mediante el ajuste del tamaño del vóxel; la cobertura anatómica, que hace referencia a la parte del organismo a analizar y que debe incluir toda la región de interés; o la resolución temporal en los estudios dinámicos, para poder modelar adecuadamente los procesos fisiológicos en el tiempo.
- Procesado de las imágenes: previo a la cuantificación puede ser necesario aplicar una serie de transformaciones a la imagen para garantizar la calidad de la medida. Entre las técnicas de procesado tenemos:
 - Filtro de ruido, para maximizar la relación señal-ruido.
 - Filtro de inhomogeneidades, para eliminar artefactos en la imagen.
 - Algoritmos de super-resolución, para aumentar la resolución espacial de la imagen sin alterarla.
 - Algoritmos de corrección, esenciales en los estudios en los que sea necesario emplear imágenes procedentes de distintas modalidades o secuencias/series. Estos algoritmos permiten el alineamiento espacial de distintas imágenes para garantizar la correspondencia vóxel a vóxel.
 - Algoritmos de segmentación, necesarios cuando se desee analizar una región de la imagen en concreto. Estos algoritmos permiten detectar, dentro de una imagen, aquellos vóxeles que pertenecen a una determinada región. Por ejemplo, para la caracterización de la enfermedad difusa hepática será necesaria la segmentación del parénquima hepático.
 - Algoritmos de normalización, necesarios cuando queramos comparar resultados obtenidos de imágenes provenientes de distintas fuentes. Debido a las diferencias entre los algoritmos de reconstrucción de las diferentes máquinas de adquisición existen desviaciones entre las imágenes [33], por lo tanto, es necesario aplicar algoritmos de normalización que permitan minimizar estas desviaciones.

Capítulo 1. Introducción

- Análisis y modelado: esta etapa consiste en el desarrollo de modelos computacionales para extraer la información de interés. Podemos diferenciar las siguientes familias de tipos de variable: volumen y forma, textura, topología o propiedades físicas, biológicas y funcionales.
 - Medidas: dependiendo del tipo de análisis desarrollado se extraerán unas métricas u otras. En determinados análisis, se realiza una cuantificación vóxel a vóxel de la región de interés y, por tanto, se pueden extraer métricas relacionadas con la distribución de los valores en dicha región a través de métricas como la media, desviación estándar, mediana o los percentiles. Otros análisis, en cambio, ofrecen directamente un valor absoluto que concierne a toda la región analizada, como es el caso del volumen o el análisis de textura.
3. **Validez fisiopatológica**: El último paso en el desarrollo de un biomarcador de imagen es su validación, es decir, comprobar que ofrece los resultados esperados en distintas poblaciones. Además, en esta fase, se detectan los sesgos que puedan aparecer, así como variaciones que puedan deberse a factores como la edad o el sexo. Esta etapa se divide a su vez en dos fases:
- Prueba de principio: su principal objetivo es comprobar que se cumplen las pruebas de concepto y mecanismo en un conjunto reducido de casos.
 - Prueba de eficacia y efectividad: una vez se ha pasado satisfactoriamente la prueba de principio, se emplea una población más grande para analizar la eficacia y efectividad del biomarcador de imagen, es decir, si se relaciona correctamente con el objetivo clínico.

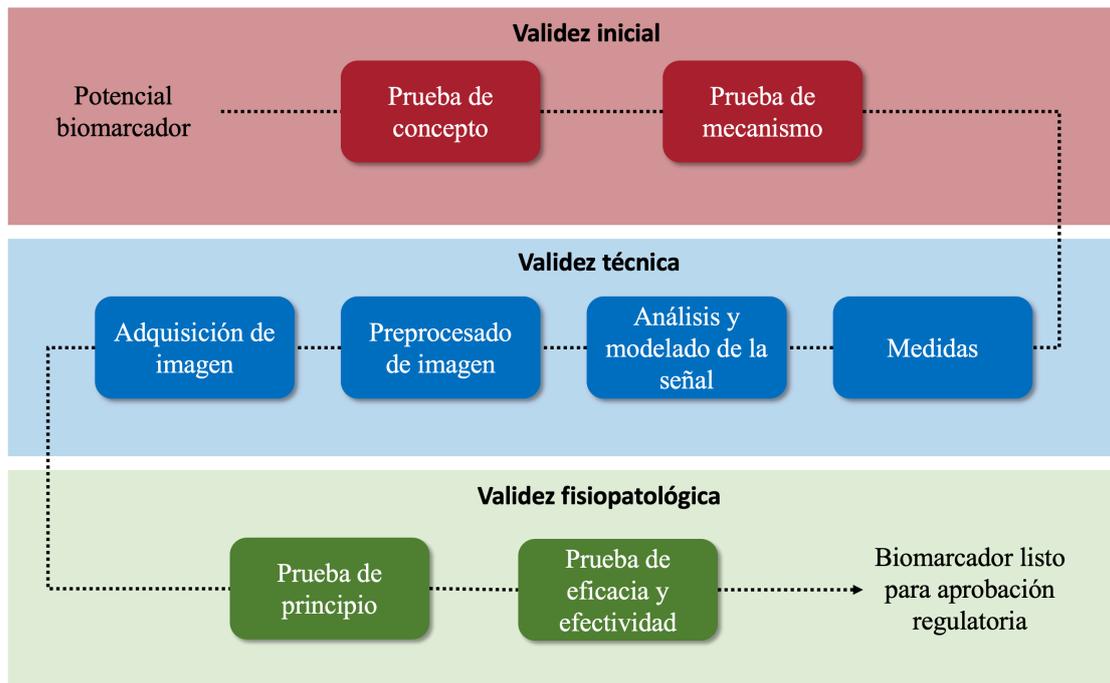


Figura 9. Fases para el desarrollo de un biomarcador de imagen. Adaptado de [32].

1.3.2. Biomarcadores de imagen para la caracterización de la enfermedad difusa hepática

Como hemos visto en secciones anteriores, la enfermedad difusa hepática, entre otras afecciones, comprende la presencia de depósitos de grasa (esteatosis) y/o hierro en el hígado. Para poder realizar la cuantificación de estos componentes se emplea una secuencia de RM multi-eco con desplazamiento químico. Los cambios de señal entre los tiempos de eco en fase y en fase opuesta, junto con las imágenes reales e imaginarias, permiten modelar y diferenciar la señal de grasa y agua en un único vóxel de la imagen. Además, la caída de la señal de agua a lo largo de los tiempos de eco, permite cuantificar el tiempo de relajación $T2^*$, cuya inversa, el $R2^*$, está directamente relacionado con la concentración de hierro [34]. Por lo tanto, esta secuencia permite la cuantificación simultánea de la fracción grasa, a través de la fracción grasa de densidad protónica (PDFF, del inglés, *Proton Density Fat Fraction*) y de la sobrecarga de hierro, a través del coeficiente de relajación $R2^*$. Realizar esta cuantificación de manera simultánea es de especial relevancia para minimizar los sesgos introducidos en la cuantificación grasa ante la presencia de hierro y viceversa [35,36].

El módulo de la señal de RM obtenida en un determinado vóxel sigue la Ecuación 5.

$$|S(TE)| = \sqrt{S_w^2 \cdot e^{-2 \cdot R2_w^* \cdot TE} + S_f^2 \cdot e^{-2 \cdot R2_f^* \cdot TE} + 2 \cdot S_w \cdot S_f \cdot e^{-(R2_w^* + R2_f^*) \cdot TE} \cos(\omega \cdot TE)} \quad (5)$$

De la Ecuación 5 podemos diferenciar tres componentes principales: la primera, correspondiente a la señal del agua (S_w); la segunda, correspondiente a la señal de la grasa (S_f) y; la tercera, que representa el desplazamiento químico entre ambas modelado a través de una función cosenoidal. Se puede observar, además, que en cada componente aparece una exponencial que va a modelar la caída de la señal, con el parámetro $R2^*$, debido a la presencia de hierro. Adicionalmente, en la ecuación, podemos encontrar una serie de constantes y variables conocidas:

- $S(TE) \rightarrow$ Señal de RM adquirida a un determinado tiempo de eco (TE).
- $TE \rightarrow$ Tiempo de eco.
- $\omega = 2\pi f \rightarrow$ Donde f es la frecuencia relativa del pico de grasa con respecto al agua.

La Figura 10 muestra un ejemplo sintético del espectro de la señal, donde tenemos un pico principal, correspondiente al agua, y múltiples picos de grasa. Dependiendo del modelo matemático que se emplee, se cuantificará la señal de grasa empleando unos picos u otros. En el modelo de un pico se emplea el pico principal de grasa, localizado a una frecuencia relativa de 420 Hz (1,46 ppm) al pico del agua. Por otro lado, los modelos multi-pico, tienen en cuenta la señal proveniente de otras frecuencias, ofreciendo una cuantificación más precisa de la fracción de grasa hepática [37-39].

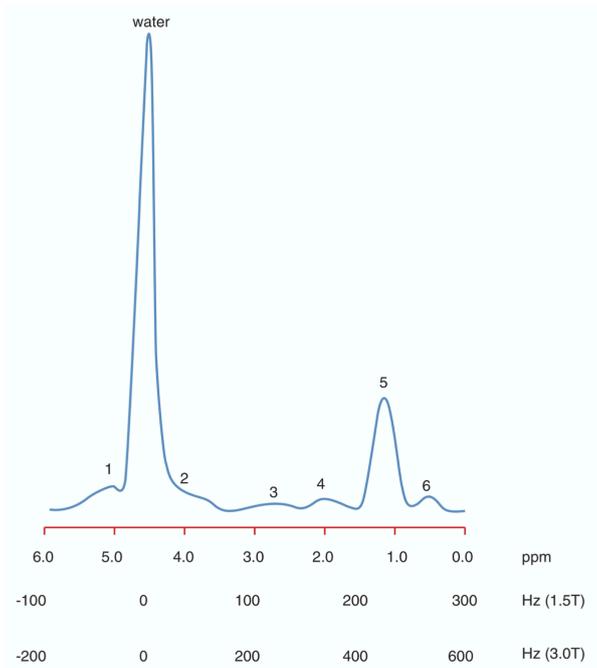


Figura 10. Espectro sintético de la señal. Podemos observar a la izquierda el pico principal del agua y, diferentes picos de grasa a distintas frecuencias [40].

Por otro lado, tenemos parámetros desconocidos y que deberemos ajustar. Entre los biomarcadores de imagen a cuantificar encontramos los siguientes:

- $S_w \rightarrow$ Señal del agua.
- $S_f \rightarrow$ Señal de la grasa.
- $R2_w^* \rightarrow$ Coeficiente de relaxometría $R2^*$ en el agua, relacionado con la concentración de hierro.
- $R2_f^* \rightarrow$ Coeficiente de relaxometría $R2^*$ en la grasa.

De estos se puede medir la PDFF como la proporción de señal grasa con respecto a la señal total en un vóxel:

$$PDFF (\%) = \frac{S_f}{S_f + S_w} * 100$$

Como hemos visto, el coeficiente de relajación $R2^*$, medido como la inversa del tiempo $T2^*$, está directamente relacionado con la presencia de hierro. En la Figura 11 podemos comparar la señal adquirida en un hígado sano con respecto a la que se obtiene en un hígado con esteatosis o uno con sobrecarga de hierro. Podemos observar, que en aquel que sufre de esteatosis hay más variabilidad en la intensidad de señal a lo largo de los tiempos de eco, mientras que en aquel que presenta sobrecarga de hierro, la señal cae más

Capítulo 1. Introducción

rápida ($R2^*$: 35s^{-1} en hígado sano frente a $R2^*$: 185s^{-1} en hígado con sobrecarga de hierro).

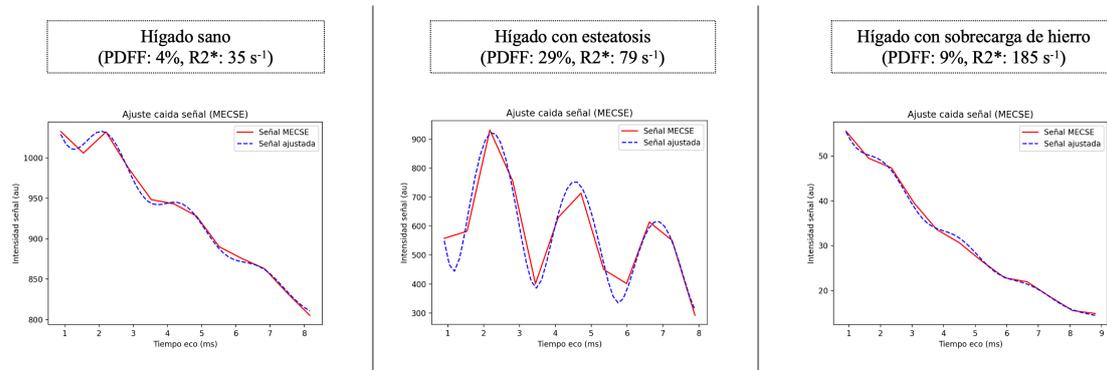


Figura 11. Señal adquirida en un determinado voxel en un hígado sano (izquierda), en un hígado con esteatosis (centro) y en un hígado con sobrecarga de hierro (derecha). Se puede observar que en un hígado con esteatosis hay más variación en la intensidad de señal entre la imagen en fase y la imagen en fase opuesta, mientras que en un hígado con sobrecarga de hierro el efecto que se produce es que la caída de la señal se produce de manera acelerada.

Este método de RM para la cuantificación del hierro mide el efecto de éste en el decaimiento de la señal, no obstante, la espectrofotometría de la muestra hepática mide la cantidad de hierro (mg o μmol) por gramo de tejido hepático seco a través de la concentración de hierro hepática (LIC, del inglés *Liver Iron Concentration*). Es por esto, que diferentes estudios han tratado de desarrollar curvas de calibración que permitan pasar entre una medida y otra [26, 41-43]. Sin embargo, existen diferencias entre dichas curvas de calibración, obteniéndose valores de LIC diferentes para un mismo valor de $R2^*$ dependiendo de la curva empleada. Estas discrepancias se deben a factores como:

- El uso de escáneres comerciales diferentes, con distintos campos magnéticos y distintos protocolos de adquisición (valores de TE, espaciado entre ecos...).
- Diferencias en la población empleada, ya que el valor de $R2^*$ no depende únicamente de la concentración de hierro, sino también de otros factores como el tamaño, forma y distribución de las partículas.
- Diferencias en el método empleado para la cuantificación de $R2^*$, ya que la coexistencia de esteatosis puede alterar esta medida si no se tiene en cuenta en la cuantificación.
- El sesgo inherente de emplear la biopsia como verdad absoluta, ya que la LIC medida a través de la biopsia está sesgada por el tamaño y localización de la muestra tomada [44].

Por todo esto, hoy día no existe un método único para realizar la conversión de $R2^*$ a LIC, aunque, desde alianzas europeas como EIBALL (*European Imaging Biomarker Alliance*) o la ESGAR (*European Society of Gastrointestinal and Abdominal Radiology*) y estadounidenses como QIBA (*Quantitative Imaging Biomarkers Alliance*) o SAR (*Society of Abdominal Radiology*) existen grupos de profesionales que trabajan por estandarizar estos procedimientos [45].

Todos estos análisis permiten realizar la cuantificación vóxel a vóxel tanto de la PDFF como del $R2^*$ o LIC y generar mapas paramétricos, es decir, imágenes en las que en cada vóxel se represente el valor del biomarcador de imagen cuantificado (Figura 12).

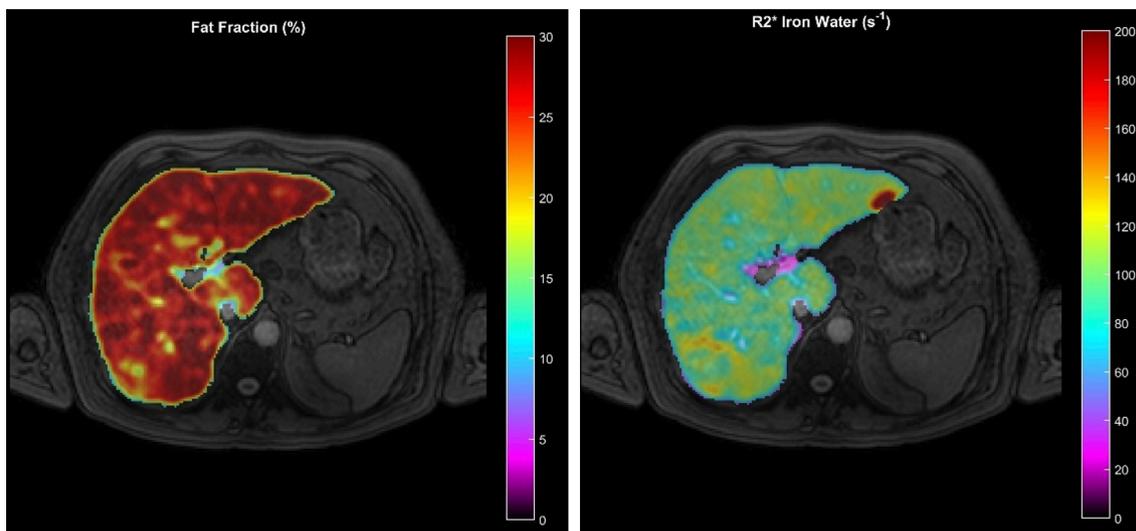


Figura 12. Mapas paramétricos de la PDFF (izquierda) y el $R2^*$ (derecha) del parénquima hepático en un paciente con esteatosis hepática.

1.3.3. Biomarcadores de imagen para la caracterización del carcinoma hepatocelular

Mediante el uso de imágenes de RM, para la caracterización del CHC, principalmente se realiza una evaluación de la difusión del tejido tumoral junto con una evaluación de su perfusión.

1.3.3.1. Análisis de difusión

La difusión mide la distancia que se han movido las moléculas de agua por unidad de tiempo. La difusión en un tejido sano presenta valores elevados, ya que se preserva el espacio intersticial y, por tanto, las moléculas de agua se mueven de forma normal, sin embargo, en un tejido patológico, se produce una reducción del espacio intersticial por

Capítulo 1. Introducción

diferentes procesos (angiogénesis, inflamación, necrosis, fibrosis) y, por tanto, el movimiento de las moléculas de agua está limitado, presentando valores disminuidos de difusión.

El principal biomarcador de imagen empleado en práctica clínica para el análisis de la difusión tisular es el coeficiente de difusión aparente (ADC, del inglés, *Apparent Diffusion Coefficient*). Se ha evidenciado que existe una elevada relación entre la celularidad tisular y el ADC [46, 47]. Hoy día no hay consenso a la hora de establecer los rangos de valores normales de ADC, sin embargo, valores inferiores a $1 \cdot 10^{-3} \text{ mm}^2/\text{s}$ generalmente indican que hay una restricción en el movimiento de las moléculas de agua. Sin embargo, esto dependerá del órgano y patología bajo estudio [48-50].

El ADC asume un movimiento aleatorio de las moléculas de agua en un espacio libre. El término aparente se le atribuye por el hecho de asumir un movimiento libre que sigue una distribución gaussiana.

Para poder cuantificar este biomarcador es necesario adquirir una secuencia de RM potenciada en difusión (DWI) con, por lo menos, dos valores b (uno bajo: 0-50, y otro alto: 800-1000). Con estos se asume una caída exponencial de la señal modelada por el ADC según la Ecuación 6.

$$S(b) = S_0 \cdot e^{-b \cdot ADC} \quad (6)$$

Donde $S(b)$ se refiere a la intensidad de la señal obtenida a un determinado vóxel para un valor b dado; S_0 hace referencia a la intensidad de la señal adquirida con un valor $b=0$; y b es el valor b empleado para adquirir la imagen dada.

Por lo tanto, con el análisis vóxel a vóxel del ADC se puede generar un mapa paramétrico en el que en cada vóxel de la imagen se tenga el valor de ADC correspondiente (Figura 13).

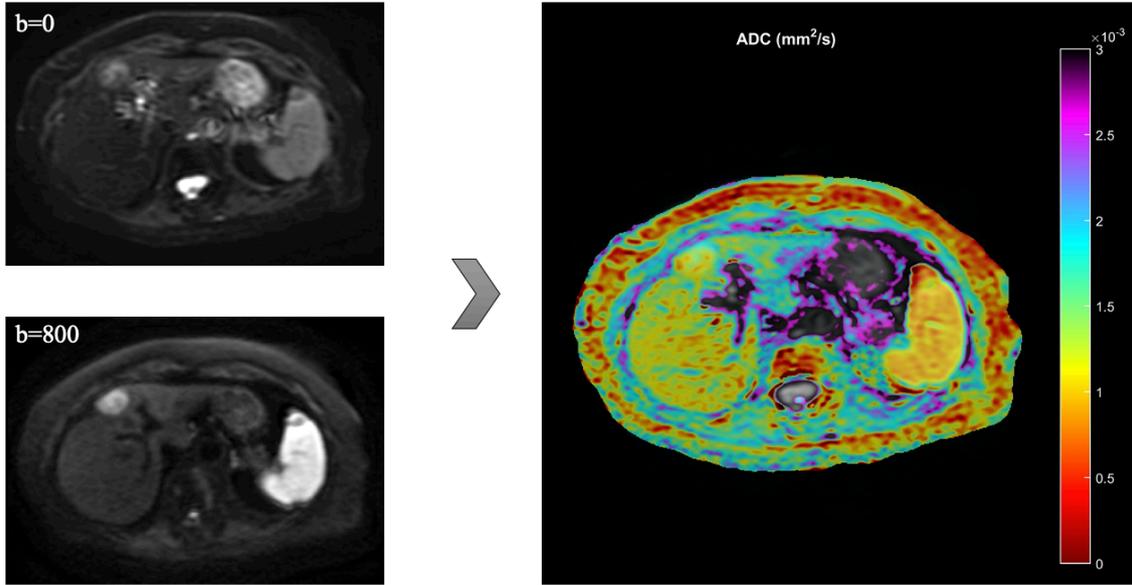


Figura 13. Análisis de difusión en un paciente con CHC. A la izquierda se observan las imágenes de la secuencia DWI empleadas, con valores b igual a 0 y 800. A la derecha observamos el mapa paramétrico de ADC correspondiente.

Sin embargo, la asunción de que las moléculas de agua se mueven libremente no es cierta, ya que éstas presentan restricciones en su movimiento debido a elementos como las membranas celulares. Además, en la cuantificación del ADC se incluye una componente originada por los vasos, generando un efecto de movimiento incoherente intravoxel [51]. De esta última idea surge el modelo IVIM (del inglés, *Intra-Voxel Incoherent Motion*), el cual modela la caída de la señal como una caída bi-exponencial, caracterizada por una primera caída rápida que ocurre a valores b bajos debido al efecto de la perfusión, seguida de una segunda caída, producida por el efecto de la difusión. Este modelo sigue la Ecuación 7.

$$S(b) = S_0 \cdot [f \cdot e^{-b \cdot (D+D^*)} + (1-f) \cdot e^{-b \cdot D}] \quad (7)$$

Donde $S(b)$ es el valor de la señal adquirido para un valor b dado; S_0 es el valor de la señal para un valor $b=0$; b es el valor b empleado para adquirir la imagen dada; f es la fracción vascular, es decir, la fracción de vasos en un determinado vóxel; D^* es la difusión debida a la componente vascular, también conocida como pseudodifusión; y D es el valor de la difusión debido al movimiento de las moléculas de agua.

Para poder llevar a cabo este modelado se requiere una adquisición DWI con mayor número de valores b . Concretamente, se necesitan tres imágenes o más con valores b entre 0 y 200 y, como mínimo, otra imagen adquirida con un valor b entre 800 y 1000 [52].

Capítulo 1. Introducción

Los dos métodos anteriores, asumen una distribución gaussiana en el movimiento de las moléculas de agua en el tejido biológico, es decir, un movimiento libre sin restricciones. Sin embargo, la complejidad intra- y extra- celular hace que la difusión de dichas moléculas se desvíe de esa distribución gaussiana [53]. En estadística, la alteración de un patrón gaussiano recibe el nombre de curtosis, así, la imagen de curtosis de difusión (DKI, del inglés, *Diffusion Kurtosis Imaging*) es un intento de tener en cuenta esta variación para proporcionar un modelo de difusión más preciso y para capturar el comportamiento no gaussiano de la difusión como un marcador de la heterogeneidad tisular [54]. El modelo DKI se rige de acuerdo con la Ecuación 8.

$$\ln \left[\frac{S(b)}{S_0} \right] = -b \cdot D + \frac{1}{6} \cdot b^2 \cdot D^2 \cdot K \quad (8)$$

Donde D es el coeficiente aparente de difusión; y K es la curtosis de la distribución del movimiento de las moléculas de agua, donde $K \geq 0$ [55].

De la Ecuación 8 podemos observar que si $K=0$, la ecuación se corresponde a la vista para el modelo de ADC (Ecuación 6). Este efecto de curtosis se captura empleando valores b altos, ya que la técnica se vuelve más sensible a las distancias moleculares más cortas y a las estructuras heterogéneas. Por lo tanto, para poder aplicar este modelo a una secuencia DWI se requiere que la adquisición conste de, al menos, 2-3 valores b entre 800 y 2000.

Además del modelo DKI, se propuso también el modelo de difusión estirada (SEM, del inglés *Stretched Exponential Model*) cuando asumimos que el movimiento de las moléculas de agua se aleja de una distribución normal [56]. Este modelo se rige de acuerdo con la Ecuación 9.

$$S(b) = S_0 \cdot e^{(-b \cdot DDC)^\gamma} \quad (9)$$

Donde DDC hace referencia al coeficiente de difusión distribuido (del inglés, *Distributed Diffusion Coefficient*) y γ es el índice de difusión molecular del agua intravoxel y adquiere un valor entre 0 y 1. Al igual que para el modelo DKI, se requieren de valores b elevados para poder realizar un ajuste adecuado de este modelo.

1.3.3.2. Análisis de perfusión

Los procesos oncológicos son conocidos por generar un aumento de los vasos en la región tumoral por los procesos de angiogénesis y neovascularización dirigidos por un aumento de factores pro-angiogénicos como el factor de crecimiento endotelial vascular (VEGF, del inglés, *Vascular Endothelial Growth Factor*). Estos nuevos vasos sanguíneos favorecen el crecimiento tumoral, alimentando las áreas hipóxicas y zonas necróticas del tumor para proporcionarle nutrientes y oxígenos esenciales [57]. Así, las células tumorales liberan moléculas (como el VEGF) que inducen al tejido huésped sano circundante a promover el crecimiento de nuevos vasos sanguíneos, permitiendo que el proceso de angiogénesis comience [58]. Esto además permite que las células cancerígenas se diseminen hacia zonas distantes (metástasis) a través del torrente sanguíneo. Es por esto, que el estudio de la perfusión en estas regiones es de especial interés ya que, en comparación con un tejido sano, un tejido tumoral presentará un aumento de su vascularización y, por tanto, de su perfusión vascular.

Para realizar este tipo de análisis se emplea la secuencia DCE, adquirida tras el suministro de un agente de contraste, generalmente Gadolinio, en el paciente. Por lo tanto, el análisis de perfusión se basa en estudiar el comportamiento de dicho agente de contraste en el organismo, ya que éste se depositará en mayor concentración en aquellas regiones más vascularizadas y, por tanto, en regiones tumorales se observarán una mayor captación de contraste.

La cuantificación de las propiedades de perfusión se puede realizar desde dos perspectivas diferentes: análisis semicuantitativo, cuya base es la extracción de parámetros descriptores de las curvas de captación de contraste, y análisis farmacocinético, que asume dos compartimentos, el vascular y el extracelular-extravascular, y estudia la transferencia del agente de contraste entre ambos compartimentos.

Previo a cualquiera de los análisis anteriores es necesario realizar una transformación de la señal adquirida, convirtiendo las curvas intensidad-tiempo a concentración-tiempo. La alianza norteamericana QIBA (*Quantitative Imaging Biomarkers Alliance*), que vela por conseguir una estandarización en los métodos de cuantificación, en su perfil DCE-MRI (v1.6)¹ recomienda emplear la Ecuación 10 para realizar dicha conversión:

¹ https://qibawiki.rsna.org/images/1/12/DCE-MRI_Quantification_Profile_v1.0.pdf

$$C(t) = \frac{\frac{1}{T1(t)} - \frac{1}{T1(0)}}{R} \quad (10)$$

Donde R es la relajatividad del agente de contraste, obtenida de las especificaciones del fabricante del contraste empleado. T1 es el tiempo de relajación longitudinal, siendo T1(0) el valor de T1 de un tejido antes del suministro del contraste y T1(t), el valor de T1 en un tiempo determinado tras el suministro del contraste.

T1(t) se calcula siguiendo una aproximación no lineal de acuerdo con las Ecuaciones 11-13.

$$\frac{1}{T1(t)} = -\frac{1}{TR} \cdot \ln\left(\frac{1 - A}{1 - \cos(\alpha) \cdot A}\right) \quad (11)$$

$$A = B \cdot \frac{S(t)}{S(0)} \quad (12)$$

$$B = \frac{1 - e^{-TR/T1(0)}}{1 - \cos(\alpha) e^{-TR/T1(0)}} \quad (13)$$

Donde TR es el tiempo de repetición de la adquisición (en segundos), α es el ángulo de magnetización de la adquisición (en radianes), S(0) es el valor de la señal registrado en un instante anterior al suministro de contraste y S(t) es el valor de la señal registrado tras un tiempo t del suministro del agente de contraste.

Los valores de T1(0) se pueden obtener de manera experimental a través de una secuencia de inversión recuperación, saturación recuperación o con ángulos de magnetización variables. El método más preciso es el de inversión recuperación, sin embargo, requiere largos tiempos de adquisición y, por tanto, no asumibles en práctica clínica. La secuencia con ángulos de magnetización variables elimina este problema ya que requiere tiempos de adquisición más cortos, permitiendo su integración en práctica clínica. Sin embargo, esta secuencia introduce errores en el análisis de perfusión, debido principalmente a artefactos de movimiento de la imagen y heterogeneidades de campo magnético [59,60]. Pese a que existen algoritmos de corrección para eliminar el movimiento y filtros de corrección de heterogeneidades, éstos son computacionalmente costosos (alargando el tiempo del análisis) y pueden no eliminar correctamente los artefactos. Por esto, es común realizar el análisis cuantitativo empleando valores fijos obtenidos de la literatura que dependerán del campo magnético empleado y el tejido bajo estudio [61]. Pese a que este

método introduce un sesgo uniforme en la cuantificación, se ha demostrado que ofrece mayor reproducibilidad y valor diagnóstico [62,63].

Análisis semicuantitativo

Análisis descriptivo de la curva de captación de contraste (Figura 14). Los principales parámetros que se cuantifican en cada uno de los vóxeles de la imagen son:

- **Concentración máxima:** pico de concentración de contraste máximo. A mayor vascularización mayor concentración de contraste captará una región.
- **Tiempo al pico (TTP, del inglés, *Time To Peak*):** tiempo que tarda una región en alcanzar el pico máximo de contraste. Cuanto más vascularizado esté el tejido más rápido alcanzará el pico, por tanto, a más vascularización, menor tiempo al pico.
- **Pendiente de captación (*wash-in*):** pendiente de subida de la curva de captación. A mayor pendiente, más vascularización.
- **Área bajo la curva (AUC, del inglés, *Area Under the Curve*):** cantidad de contraste que ha pasado por una región en un determinado tiempo. Generalmente se suele analizar a los 30, 60 y 90 segundos tras el suministro del bolo de contraste. Cuanto mayor sea este valor, más contraste ha captado la región analizada y, por tanto, más vascularizada estará.

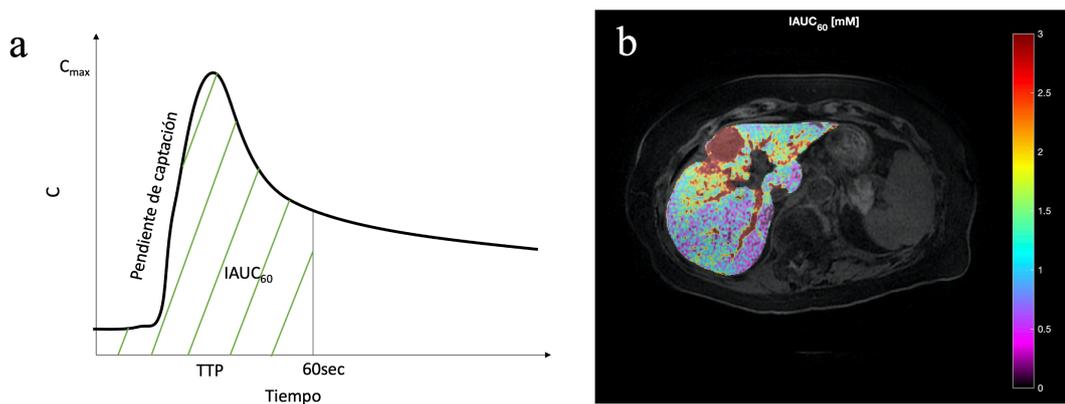


Figura 14. (a) Descriptores semicuantitativos de una curva de captación de contraste en un estudio de perfusión (DCE-RM). (b) Mapa paramétrico del área bajo la curva a los 60 segundos, se puede observar que en la lesión se obtienen valores más elevados.

Análisis farmacocinético

La farmacocinética es una rama de la farmacología que estudia los efectos del suministro de un fármaco en el cuerpo humano. Incluye los mecanismos de absorción, distribución, excreción, el instante en el que el fármaco empieza a hacer efecto y su duración.

El estudio farmacocinético de perfusión basado en RM se basa en el modelo de Tofts [64, 65]. Según éste, cada vóxel tiene influencia de tres componentes distintas (Figura 15): espacio intracelular, vasos sanguíneos y espacio extracelular-extravascular (EEE). Tras el suministro del agente de contraste en el sistema sanguíneo del paciente, éste atraviesa el endotelio vascular hacia el EEE por difusión pasiva debido al gradiente de concentración. La constante de transferencia entre estos dos compartimentos (vascular y EEE) depende del flujo sanguíneo, el área de la superficie vascular y su permeabilidad y se modela por la constante conocida como K^{trans} . A continuación, hay un reflujo desde el EEE a los vasos modelado por la constante k_{ep} . El contraste no llega a entrar a las células, por lo tanto, su concentración va a depender de la fracción de volumen vascular (v_p) y del EEE (v_e).

Existen dos modelos de Tofts, el primero de ellos [64], de 1997, en el que asume que la componente vascular (v_p) es muy inferior a la EE (v_e) y, por tanto, su efecto es muy bajo en la señal de RM. Mientras que el segundo modelo [65], revisión del anterior, de 1999, incluye la contribución de la componente vascular que, pese a añadir complejidad al análisis, es más exacto en tumores muy vascularizados.

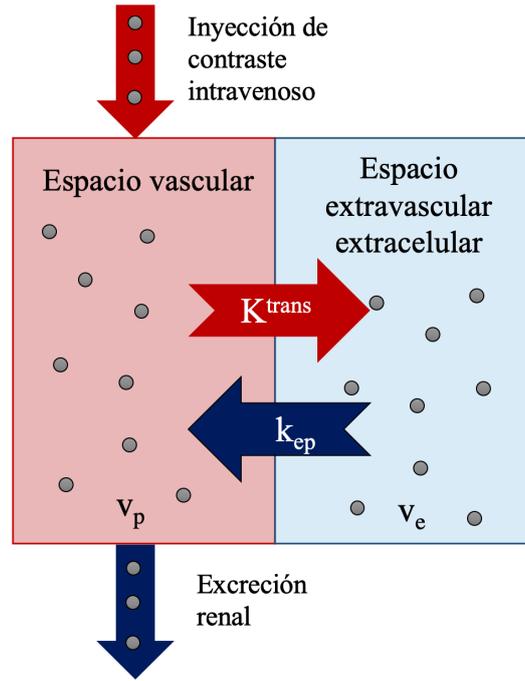


Figura 15. Modelo farmacocinético de transferencia de contraste entre el espacio vascular y el espacio extravascular extracelular.

El modelo de Tofts extendido, se rige por la Ecuación 14.

$$C_t(t) = v_p \cdot C_a(t) + K^{trans} \int_0^t C_a(\tau) \cdot e^{-k_{ep} \cdot (t-\tau)} \cdot d\tau \quad (14)$$

Finalmente v_e se calcula de acuerdo con la Ecuación 15.

$$v_e = \frac{K^{trans}}{k_{ep}} \text{ donde } v_e \in [0,1] \quad (15)$$

Siendo $C_a(t)$ la concentración de la función de entrada arterial en un tiempo determinado. Sin embargo, QIBA, en su perfil DCE-MRI (v1.6) recomienda emplear la versión estándar, es decir, sin tener en cuenta la componente vascular ($v_p C_a(t)$), para minimizar errores en la cuantificación.

Como se ha visto en la ecuación anterior, es necesario definir una función de entrada arterial, para ello se emplean los vasos que irrigan en mayor medida la región donde se encuentra el tumor (i.e., arterias ilíacas para la próstata, arteria cerebral media para el cerebro, etc.). Sin embargo, en el hígado tenemos dos sistemas sanguíneos que interactúan

Capítulo 1. Introducción

en gran medida con este órgano, el sistema venoso, a través de la vena porta, y el sistema arterial, a través de la arteria hepática. Por lo tanto, en un único vóxel la concentración del contraste vendrá determinada por ambas componentes (Ecuación 16).

$$C_t(t) = C_v(t) + C_a(t) \quad (16)$$

Por lo tanto, para poder realizar el estudio de perfusión semicuantitativo es necesario, previo al análisis, delinear mediante regiones de interés, la región sobre la que se va a calcular la función de entrada arterial.

1.4. Aprendizaje profundo

En esta sección se realiza una introducción al aprendizaje profundo (o, del inglés, *Deep Learning*) y nos centramos en su uso concreto en el sector de la imagen médica, concretamente en el uso de redes neuronales convolucionales (CNN, del inglés *Convolutional Neural Networks*) para resolver problemas de clasificación y segmentación. Además, se revisan las arquitecturas más empleadas hoy día para estos fines.

1.4.1. Introducción al aprendizaje profundo

El aprendizaje profundo es un área comprendida dentro del aprendizaje máquina (o del inglés, *Machine Learning*) que busca simular el aprendizaje del cerebro humano. Los algoritmos de aprendizaje profundo se basan en la implementación de una serie de nodos, conocidos como neuronas artificiales, distribuidas en un conjunto de capas sucesivas que se interconectan entre sí, simulando las conexiones sinápticas de las neuronas biológicas (Figura 16).

Podemos diferenciar dos tipos de aprendizajes; el aprendizaje supervisado basado en el uso de datos etiquetados, es decir, para cada dato de entrada hay una salida conocida asociada; y el aprendizaje no supervisado, en el cual disponemos de datos de entrada, pero desconocemos su salida correspondiente. Dentro del aprendizaje no supervisado, los algoritmos más conocidos son los de agrupación o *clustering*, donde se intentan buscar grupos de datos, similares entre sí, sobre los que a priori desconocemos su relación.

En este trabajo, nos centramos en el primer tipo de aprendizaje, el aprendizaje supervisado. Es decir, en redes de aprendizaje profundo basadas en aprendizaje supervisado compuestas: de una capa de entrada, que tendrá tantas neuronas como características tengan los datos; una serie de capas ocultas, cada una de ellas con un número determinado de neuronas; y finalmente una capa de salida, que tendrá tantas neuronas como clases se quieran predecir. Tanto el número de capas ocultas como el número de neuronas por capa oculta se determina de manera experimental por el especialista en ciencia de datos de forma que se alcance una relación de compromiso entre complejidad y rendimiento de la red.

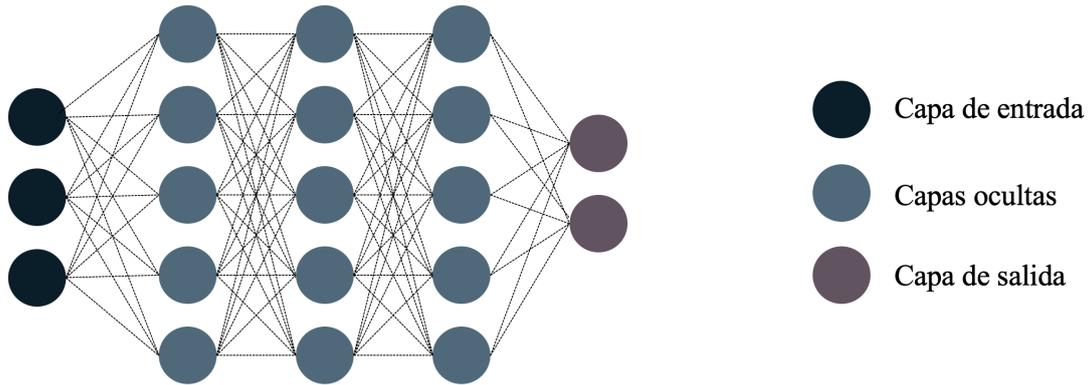


Figura 16. Arquitectura de una red neuronal artificial compuesta por una capa de entrada, una serie de capas ocultas y una capa de salida.

Cada neurona artificial, recibe una entrada que sigue la Ecuación 17.

$$a_i = \sum_{j=1}^n x_j \cdot w_{j,i} + b \quad (17)$$

Siendo n el número de neuronas de la capa anterior, x_j la salida de una de las neuronas de la capa anterior, $w_{j,i}$ el peso de la conexión entre la neurona de la capa anterior y la que se está calculando y b el sesgo. El peso $w_{j,i}$ es lo que se va ajustando durante el proceso de entrenamiento. Finalmente, la salida de la neurona y_i se obtiene aplicando una función de activación $f(a_i)$. Esta transformación se emplea para introducir no linealidades al modelo y, de esta forma, sea capaz de resolver problemas más complejos. En la Figura 17 podemos observar algunas de las funciones de activación más empleadas.

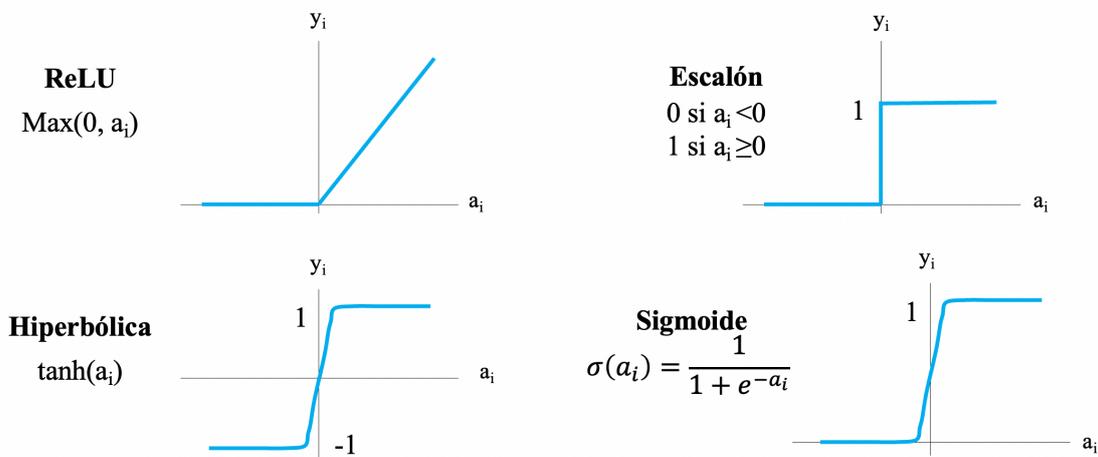


Figura 17. Funciones de activación empleadas en redes neuronales artificiales.

Como hemos comentado previamente, los pesos de las conexiones entre las neuronas se van ajustando durante el proceso de entrenamiento, el cual se basa en el algoritmo de descenso de gradiente [66]. Este algoritmo calcula el gradiente de error en cada iteración del entrenamiento, empleando las muestras de entrenamiento, y, a continuación, actualiza los pesos del modelo empleado el algoritmo de propagación hacia atrás [67]. Cuánto se actualizan estos pesos viene definido por dos conceptos principales: la tasa de aprendizaje y la función de pérdidas. La tasa de aprendizaje controla cuánto cambia el modelo en respuesta al error estimado en cada iteración. Seleccionar una tasa de aprendizaje adecuada es de gran importancia ya que un valor muy bajo puede conllevar periodos de aprendizaje muy largos, mientras que, un valor demasiado alto puede hacer que la red converja muy rápido en un mínimo local y, por tanto, no aprenda todo lo que podría. Por otro lado, el error estimado viene definido por la función de pérdidas. Ésta, mide cómo de buena ha sido la predicción respecto al valor real, conocido, de la salida y , en función de esto se realiza la actualización de los pesos de la arquitectura definida. Por lo tanto, es una función cuyo valor se desea minimizar. La Ecuación 18 define la actualización de los pesos tras una iteración del proceso de aprendizaje.

$$W' := W - \eta \cdot \frac{\partial J(W)}{\partial W} \quad (18)$$

Donde W' representa el valor nuevo de los pesos, W representa el valor de los pesos actuales; η hace referencia a la tasa de aprendizaje y, finalmente, $J(W)$ refleja la función de pérdidas.

Dependiendo del problema que se quiera resolver, la función de pérdidas empleada variará. En problemas de clasificación (salida categórica), la función de pérdidas más extendida es la entropía cruzada (Ecuación 19), que mide cómo de lejos la probabilidad predicha se encuentra de la clase real, mientras que para problemas de regresión (salida continua), se ajusta más una función como el error cuadrático medio (Ecuación 20).

$$J(W) = -\frac{1}{n} \sum_{i=1}^n (y \cdot \log(y') + (1 - y) \cdot \log(1 - y')) \quad (19)$$

$$J(W) = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2 \quad (20)$$

Donde n es el número de muestras; y hace referencia al valor real esperado; e y' se refiere al valor predicho por el modelo.

El proceso de entrenamiento (Figura 18) es un proceso iterativo en el que en cada iteración se realizan predicciones sobre las muestras de entrenamiento, se calcula el error en la predicción, y se ajustan los pesos de la arquitectura. Cada una de estas iteraciones se conoce como época.

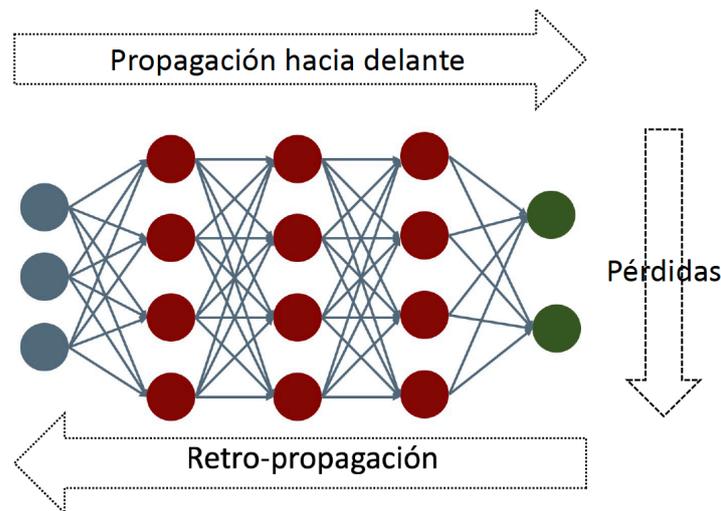


Figura 18. Proceso de entrenamiento. Las conexiones tienen unos pesos determinados, con estos (1) se calcula la salida (propagación hacia delante), a continuación (2) se calcula el error cometido en la predicción haciendo uso de la función de pérdidas, y, finalmente (3) se actualizan los pesos mediante el algoritmo de retro-propagación haciendo uso, además, de la tasa de aprendizaje. Todo este proceso se realiza en una época.

1.4.2. Redes neuronales convolucionales

Las redes neuronales convolucionales (CNN, del inglés *Convolutional Neural Network*) se diseñaron especialmente para trabajar con imágenes. La principal característica de este tipo de redes, y de la cual reciben su nombre, es que están compuestas por una serie de filtros convolucionales dispuestos a lo largo de distintas capas, encargados de extraer características de las imágenes a distintos niveles. Las CNN permiten capturar las dependencias espaciales de las imágenes. Previo al uso de las CNN, la metodología tradicional para trabajar con imágenes se basaba en la implementación de algoritmos compuestos por un conjunto de filtros, diseñados por el propio programador, seguido de un clasificador de aprendizaje máquina (como bosques de decisión, máquinas vectoriales de soporte, k vecinos más cercanos, entre otros). Esto tenía dos problemas principales; uno, se requería de un conocimiento experto en el problema a resolver para diseñar los filtros más apropiados y, dos, era muy complicado implementar filtros que se ajustaran a la amplia heterogeneidad del mundo real (i.e., un mismo objeto puede presentar cambios

de iluminación, escala, visibilidad, perspectiva, etc.). Sin embargo, gracias a la introducción de las CNN, el diseño de los filtros extractores de características se embebe en la propia red, y, por lo tanto, los parámetros de estos también se ajustan durante el proceso de entrenamiento. De este modo, permiten ampliar la complejidad y variabilidad de los filtros empleados para que estos se adapten de la forma más eficiente en la resolución del problema objetivo.

A la hora de diseñar una CNN hay que diferenciar entre dos términos, los parámetros y los hiperparámetros. Los parámetros son aquellos valores que se ajustan y aprenden durante el proceso de entrenamiento, en cambio, los hiperparámetros son aquellos valores que el diseñador de la arquitectura debe seleccionar (e.g., tamaño del filtro, función de pérdidas, tasa de aprendizaje, número de capas, entre otros). Estos últimos se suelen ajustar tras un proceso de experimentación en el que se prueban distintas combinaciones. Además, a la hora del diseño de la arquitectura se emplean diferentes elementos que, el programador, decide cómo emplear y disponer. A continuación, veremos cada uno de ellos.

1.4.2.1. Capa de entrada

A la entrada de la CNN se proporciona la imagen a procesar, por lo que a la hora de diseñar la arquitectura de la CNN hay que definir el tamaño que va a tener la imagen de entrada (número de filas y columnas), así como su tipo, definido por el número de canales (4 canales: CMYK; 3 canales: RGB, HSV; 1 canal: escala de grises, etc.). En imagen médica, los canales también se suelen emplear para introducir diferentes tipos de imagen (e.g., diferentes secuencias de RM del mismo corte, diferentes preprocesados, etc.) [68, 69].

El tamaño de la imagen también se debe escoger teniendo en cuenta las limitaciones del hardware que se vaya a emplear, especialmente en términos de memoria RAM (del inglés, *Random Access Memory*). Si se define un tamaño de imagen muy grande, aparte de que el tiempo de ejecución será mayor, es posible que no se pueda llevar a cabo el entrenamiento por falta de memoria RAM.

1.4.2.2. Capa convolucional

Esta capa es el elemento esencial de las CNN. Generalmente, una CNN se compone de varias capas convolucionales consecutivas. De esta forma, se consigue extraer

Capítulo 1. Introducción

características a distintos niveles. En capas más iniciales se obtienen características de más bajo nivel y menor complejidad, genéricas a cualquier problema, como bordes de la imagen o ejes, mientras que cuando se va avanzando en la profundidad de la red, se van obteniendo características cada vez más complejas, teniendo en las últimas capas características que definen formas completas y son más concretas del problema a resolver. La combinación de estas características a distintos niveles es lo que da gran potencia a las CNN al trabajar con imágenes.

El primer parámetro que se debe definir a la hora de diseñar la capa convolucional es el tamaño de los filtros convolucionales, también conocido como campo receptivo. Este se define como la región de la imagen de entrada por la que se ve afectada una característica en concreto. Sin embargo, no todos los píxeles incluidos dentro del campo receptivo tienen la misma importancia para generar la imagen de salida, los píxeles que están más cerca del centro del campo receptivo contribuyen en mayor medida en la generación de la salida. En la Figura 19 podemos observar un ejemplo de cómo funciona una capa convolucional al aplicar un filtro de 3x3 a una imagen de entrada. El valor de un píxel de la imagen de salida se obtiene al realizar el producto escalar entre los valores del filtro y la región de la imagen y sumando los valores obtenidos. Cada componente se obtiene mediante el desplazamiento del filtro a lo largo de la imagen. De este diagrama se puede intuir que los píxeles internos de la imagen contribuyen en la creación de características en varios movimientos del campo receptivo, sin embargo, los píxeles de los bordes contribuyen una única vez.

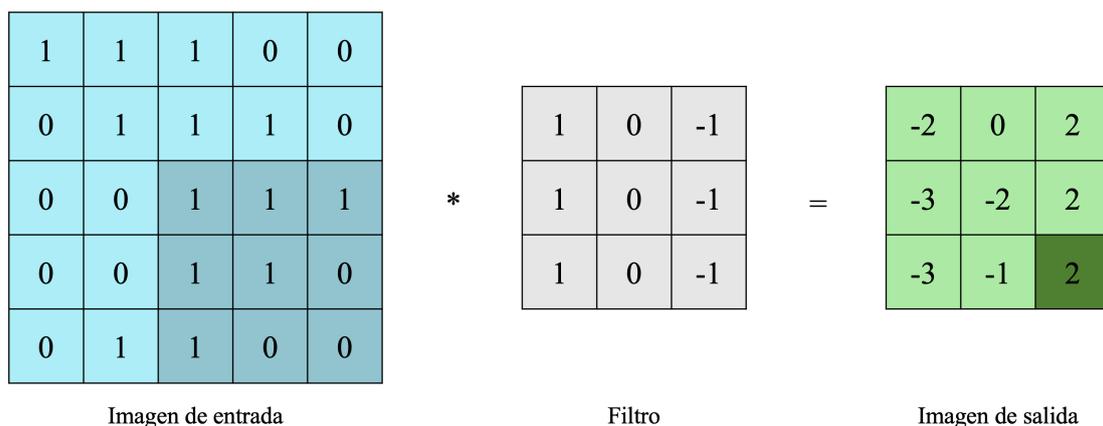


Figura 19. Funcionamiento de una capa convolucional. Ante una imagen de entrada, se aplica un campo receptivo, para generar una imagen de salida, correspondiente a un mapa de características de la imagen de entrada.

Otros dos parámetros de gran importancia son el relleno (más conocido por su traducción del inglés *padding*) y el avance (del inglés, *stride*) (Figura 20). El relleno es el número de filas y columnas con valor cero que se añaden a la imagen de entrada antes de aplicar el filtro convolucional. Este parámetro es importante para evitar perder información de los bordes de la imagen, como hemos visto anteriormente, los bordes de la imagen es la región que menos contribuye en la generación de la imagen de salida al aplicar el filtro convolucional y, por tanto, es información que se va perdiendo conforme se van aplicando capas convolucionales. Además, como se observa en la Figura 19, tras aplicar el filtro convolucional, la imagen reduce sus dimensiones, por lo tanto, si no se aplica relleno a la imagen de entrada, tras una serie de capas convolucionales, se perdería la información de la imagen inicial. Una práctica común es emplear un valor de relleno tal que el tamaño de la imagen de entrada y salida sea el mismo (*padding same*). Por otro lado, el avance es el salto que se aplica entre convolución y convolución.

Con todo esto, el tamaño de la imagen de salida queda definido por la Ecuación 21.

$$n_{salida} = \left\lfloor \frac{n_{entrada} + 2 \cdot p - f}{s} + 1 \right\rfloor \quad (21)$$

Donde $n_{entrada}$ es el tamaño de la imagen de entrada; p el relleno aplicado; f el tamaño del filtro convolucional; s el salto aplicado; y n_{salida} el tamaño del mapa de características generado, siendo el valor final el valor truncado obtenido en la operación descrita.

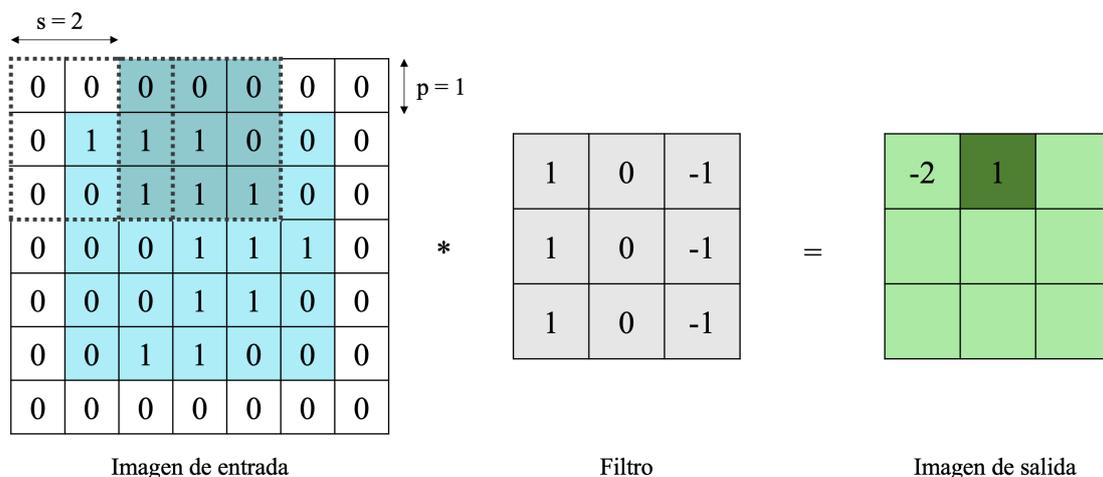


Figura 20. Ejemplo de funcionamiento de una capa convolucional a la que se le ha aplicado un relleno de 1 píxel ($p=1$) y un avance de 2 píxeles ($s=2$).

Tras aplicar el filtro convolucional a la imagen de entrada, al mapa de características resultante se le aplica una función de activación, de las vistas en el apartado anterior, para introducir no linealidades en el modelo.

1.4.2.3. Capa de agrupación

Uno de los inconvenientes de las capas convolucionales es que capturan la posición exacta de las características de la imagen de entrada. Esto quiere decir que, si no se altera el tamaño de la imagen, se obtendrán mapas de características con información similar a la salida de cada capa convolucional. Sin embargo, pequeñas variaciones en la posición de los píxeles de la imagen de entrada, dan lugar a mapas de características diferentes. Una práctica común para minimizar este problema es reducir el tamaño de la imagen, preservando las características esenciales de la misma, eliminando, de este modo, los detalles finos de la imagen que pueden no ser de interés para la tarea a resolver.

Esta reducción del tamaño de la imagen, como hemos visto en la sección anterior, se puede conseguir variando el valor del relleno y salto en la capa convolucional. Sin embargo, una aproximación más común, y que no introduce más carga computacional, es emplear una capa de agrupamiento (o, del inglés, capa de *pooling*). Esta capa se suele aplicar justo después de aplicar la función de activación al mapa de características generado con el filtro convolucional.

En relación con esta capa, se deben de seleccionar tres hiperparámetros, por un lado, el campo receptivo, por otro, el salto y, finalmente, el tipo de agrupación a aplicar. En cuanto al campo receptivo, es práctica común emplear un tamaño de 2×2 con un avance de 2, tras la cual el tamaño de la imagen se reduce a la mitad. Por otro lado, en relación con el tipo de reducción, son dos las operaciones más comunes (Figura 21):

1. Agrupación de media (o *mean pooling*, en inglés): se calcula el valor medio de cada sección.
2. Agrupación de máxima (o *max pooling*, en inglés): se calcula el valor máximo de cada sección.

Siendo el segundo el más empleado.

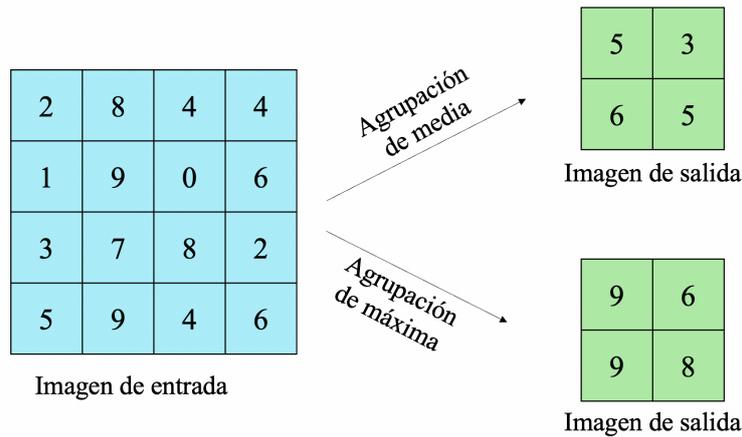


Figura 21. Tipos de operaciones más comunes en una capa de agrupación. En ambos casos se ha empleado un tamaño de 2x2 y avance de 2.

1.4.2.4. Capa de submuestreo o convolución transpuesta

En arquitecturas de codificación-decodificación, muy empleadas en la resolución de problemas de segmentación, en la fase de decodificación, se debe realizar la operación opuesta a la capa de agrupamiento, esto es, ampliar sucesivamente el tamaño de la imagen hasta alcanzar el tamaño original. Para ello se pueden emplear dos tipos de capas: las capas de submuestreo (o, del inglés, capa de *upsampling*) o las capas de convolución transpuesta o deconvolución. La capa de submuestreo redimensiona la imagen a un tamaño superior mediante una interpolación al vecino más cercano. La capa de convolución transpuesta realiza la operación opuesta a la capa convolucional, pero, controlando el campo receptivo, el avance y el relleno, se puede conseguir el tamaño de imagen deseado. La principal diferencia entre ambos métodos es que la capa de convolución transpuesta aprende cómo ampliar la imagen y, por tanto, implica parámetros a ajustar, mientras que la capa de submuestreo aplica siempre el mismo método y no añade nuevos parámetros a ajustar.

1.4.2.5. Capa totalmente conectada

La capa totalmente conectada (o capa densa) equivale a una capa de neuronas de una red neuronal artificial. Por lo tanto, la salida de la última capa convolucional, o de agrupamiento, se vectoriza para convertirla en un vector unidimensional y se introduce en la capa completamente conectada. Generalmente, se emplean varias capas de este tipo para generar la salida final. Se encuentran principalmente en la resolución de problemas de clasificación o regresión.

1.4.3. Optimización del proceso de entrenamiento

Cuando se entrena un modelo de aprendizaje profundo hay que tener en cuenta una serie de consideraciones para optimizar computacionalmente el proceso de entrenamiento, para alcanzar el mejor rendimiento del modelo y para garantizar la generalización y reproducibilidad del modelo entrenado.

1.4.3.1. Entrenamiento en lotes

Como se ha comentado anteriormente, para el ajuste de los parámetros, se emplea el algoritmo de descenso de gradiente. El algoritmo de descenso de gradiente, en su versión más simple, emplea, en cada iteración del proceso de entrenamiento, todas las muestras de entrenamiento de manera simultánea para el cálculo del error, sin embargo, esto supone un elevado coste computacional que, además, no se puede aprovechar de la paralelización. Por esto, surge una variante de este primer algoritmo conocido como descenso de gradiente estocástico que calcula el error y actualiza el modelo para cada ejemplo del conjunto de datos de entrenamiento. Las principales ventajas de este método es que es rápido de calcular, es poco costoso computacionalmente y las actualizaciones frecuentes permiten conocer inmediatamente el rendimiento del modelo y la tasa de mejora. Sin embargo, al realizar el cálculo de error con una única muestra, puede dar lugar a una señal de gradiente ruidosa, que puede hacer que los parámetros y el error del modelo tengan una mayor varianza a lo largo del entrenamiento.

Para alcanzar un equilibrio entre ambas aproximaciones se suele emplear un entrenamiento basado en lotes, esto es, el conjunto de datos de entrenamiento se divide en k lotes que se utilizan para calcular el error del modelo y actualizar los parámetros de este. Así, en cada época, se realizan k iteraciones empleándose, en cada una, n imágenes (tamaño de lote) para calcular el error del modelo. Finalmente, se calcula el error medio que se emplea para la actualización de los parámetros. La actualización de los pesos es más eficiente computacionalmente facilitando, además, la paralelización. Sin embargo, esto introduce la necesidad de definir un nuevo hiperparámetro, el tamaño de lote.

1.4.3.2. Sobreajuste y regularización

Cuando se entrena un modelo se deben evitar dos estados en los que se obtiene un rendimiento inferior al máximo posible (Figura 22), el subajuste, en el que el entrenamiento ha finalizado antes de llegar al error mínimo y, por tanto, no es capaz de

representar adecuadamente los datos; y el sobreajuste, en el que, ya sea por la complejidad excesiva del modelo o por un tiempo de entrenamiento demasiado largo, el modelo se ha adaptado demasiado a los datos de entrenamiento y no es capaz de generalizar en casos nuevos.

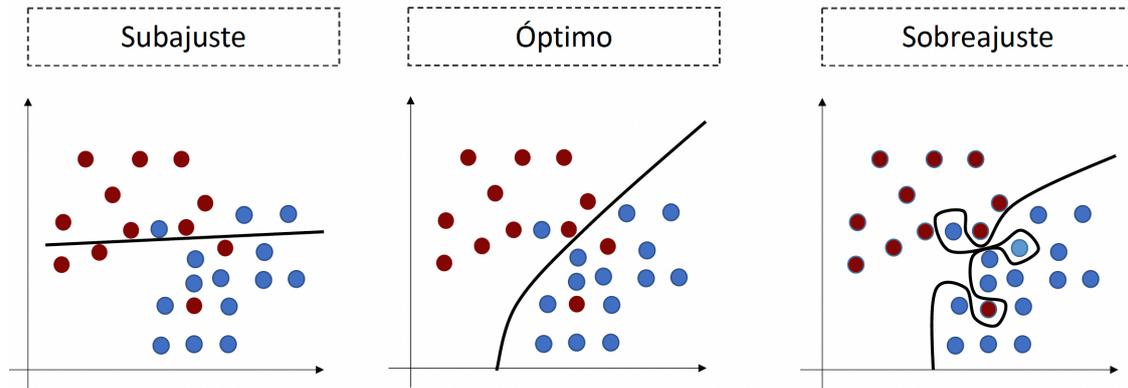


Figura 22. Representación gráfica de tres ajustes diferentes. La figura de la izquierda muestra un modelo subajustado, el cual no es capaz de representar adecuadamente los datos; la figura de la derecha muestra un modelo sobreentrenado, el cual se ha ajustado demasiado a los datos de entrenamiento y no es capaz de generalizar a nuevos datos; y, la figura del centro representa un modelo óptimo, equilibrio entre los dos anteriores.

Para evitar el subajuste se puede alargar el proceso de entrenamiento o incrementar la complejidad de la arquitectura diseñada. Por otro lado, para evitar el sobreajuste se debe aplicar técnicas de regularización que ayuden a disminuir la complejidad del modelo. Entre estas técnicas encontramos:

- **Definición de las bases de datos de entrenamiento, validación y test**

Para garantizar la generalización del modelo a nuevos datos y evitar el sobreajuste a los datos de entrenamiento, se debe emplear parte de la base de datos disponible para comprobar dicha capacidad de generalización. De este modo, se suelen emplear tres subconjuntos de datos:

- **Base de datos de entrenamiento:** Conjunto de imágenes que se emplean para el ajuste de los pesos durante el proceso de entrenamiento, es decir, de las que la red va a extraer los patrones necesarios para aprender a generalizar a nuevas imágenes de entrada.
- **Base de datos de validación:** Conjunto de imágenes que, durante el proceso de aprendizaje, se emplean para evaluar el modelo entrenado sobre un conjunto de datos independiente, es decir, no empleado para el ajuste de los pesos de la red. Este conjunto de datos se suele emplear con diferentes objetivos: (1) Seleccionar los

hiperparámetros óptimos, esto es, tras la evaluación de distintos experimentos donde se combinan diferentes hiperparámetros, se seleccionan aquellos que ofrezcan mejores resultados sobre el conjunto de datos de validación; (2) controlar el sobreajuste; y (3) guardar los pesos del modelo óptimo, esto es, los pesos que se han obtenido en la época que mejor rendimiento se observa sobre el conjunto de datos de validación.

- Base de datos de test: A pesar de que el conjunto de datos de validación contiene imágenes que no se han empleado durante el proceso de entrenamiento para el ajuste de los pesos del modelo generado, se debe realizar un testeo independiente para garantizar la generalización del modelo a imágenes completamente nuevas. Esto es debido a que las imágenes de validación se han empleado para la selección de los hiperparámetros óptimos, así como para el almacenamiento de los pesos que mejores resultados ofrecían. Así, la base de datos de test se trata de un conjunto de imágenes completamente independiente a las empleadas para el entrenamiento y validación del modelo generado.

En ocasiones, en lugar de realizar una única validación, se realizan k validaciones en un proceso de validación cruzada. En estos casos, el conjunto de datos que comprende las muestras de entrenamiento y validación se subdivide en k particiones y, en un proceso de k iteraciones, se emplean $k-1$ particiones para entrenar el modelo y la partición restante para validar. Este proceso se conoce como validación cruzada en k lotes (Figura 23). Este proceso permite evaluar la capacidad de generalización del modelo diseñado ante diferentes datos de entrenamiento y validación. Tras la validación cruzada, se seleccionan los hiperparámetros que mejor rendimiento y estabilidad media ofrezcan.

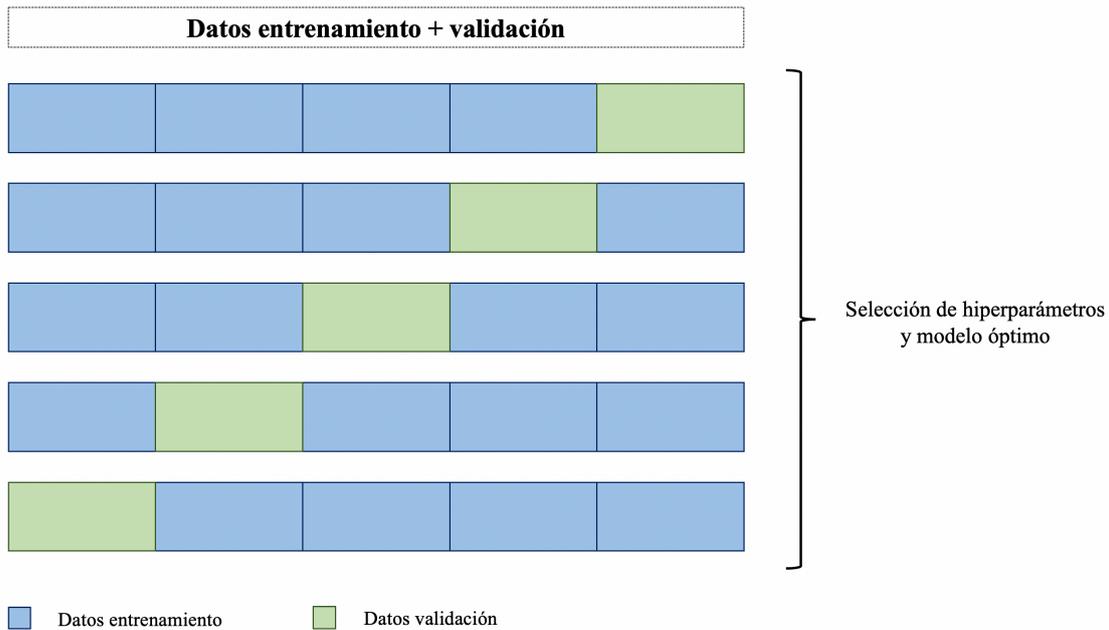


Figura 23. Metodología de un proceso de validación cruzada en k lotes (ejemplo en el que se han definido 5 lotes).

- **Aumento de datos**

Para incrementar la capacidad de generalización y robustez de los modelos generados frente a imágenes nuevas, el uso de técnicas de aumento de datos durante el proceso de entrenamiento es de gran interés. De este modo, diferentes transformaciones se pueden aplicar de forma aleatoria en cada iteración del entrenamiento. Este aumento de datos se puede realizar de dos formas: aplicando transformaciones a la propia imagen como puede ser rotaciones, traslaciones, giros izquierda-derecha, giros arriba-abajo, añadido de ruido o cambio de contraste, entre otras; o bien generando imágenes completamente nuevas de manera sintética mediante, por ejemplo, el entrenamiento de una red generativa adversaria (GAN, del inglés, *Generative Adversarial Network*) [70].

El aumento de datos se puede realizar bien previo al entrenamiento o al vuelo durante el proceso de entrenamiento (*on-the-fly*). En el primer caso, se incrementa el tamaño de la base de datos un factor igual al número de transformaciones aplicadas. Esta opción no se recomienda cuando trabajamos con bases de datos grandes ya que se multiplica el almacenamiento necesario para albergar todas las imágenes originales y transformadas. En el segundo caso, se aplican las transformaciones en cada iteración a cada lote de manera independiente. A cada imagen se le aplican, de manera aleatoria, diferentes transformaciones (e.g., añadido de ruido + rotación; rotación + traslación + cambio de contraste; zoom; etc.), esto hace que, en cada época, a la misma imagen se le apliquen

diferentes transformaciones, por lo que las variaciones que ve la red crecen exponencialmente. El aumento de datos al vuelo durante el proceso de entrenamiento ha sido la técnica empleada en esta tesis doctoral.

- **Regularización L1 y L2**

Se altera la función de pérdidas añadiéndole un término de penalización (Ecuación 22). Esta penalización produce modelos más simples que generalizan mejor.

$$J(W) = \frac{1}{n} \sum_{i=1}^n \mathcal{L}(y'_i, y_i) + \alpha C \quad (22)$$

Donde C refleja la complejidad del modelo y α indica cómo de importante es que el modelo sea simple en relación con cómo de importante es su rendimiento, medido por \mathcal{L} . Al añadir la penalización, se minimiza la complejidad del modelo a la vez que se minimiza la función de pérdidas.

Con la regularización L1, la complejidad se mide con la media del valor absoluto de los pesos del modelo (Ecuación 23). Esta técnica es de utilidad cuando sospechamos que algunas de las características de entrada son irrelevantes. Se favorece que los coeficientes acaben valiendo 0, permitiendo descubrir las variables de entrada relevantes (selección de características). Funciona mejor cuando las variables no están muy correlacionadas entre sí.

$$J(W) = \frac{1}{n} \sum_{i=1}^n \mathcal{L}(y'_i, y_i) + \alpha \frac{1}{m} \sum_{j=1}^m |w_j| \quad (23)$$

Con la regularización L2, la complejidad se mide como la media del cuadrado de los pesos del modelo (Ecuación 24). Esta técnica es de utilidad cuando sospechamos que algunas de las variables de entrada están correlacionadas entre sí. Hace que los coeficientes acaben siendo pequeños, minimizando la correlación entre las variables.

$$J(W) = \frac{1}{n} \sum_{i=1}^n \mathcal{L}(y'_i, y_i) + \alpha \frac{1}{2m} \sum_{j=1}^m w_j^2 \quad (24)$$

Ambos tipos de regularización se pueden emplear tanto en las capas convolucionales como en las capas totalmente conectadas.

- **Dropout**

En cada época, se inactivan algunas neuronas aleatoriamente siendo diferentes, en cada época, las neuronas que se inactivan. Esto hace que ningún nodo tenga un valor elevado y, como resultado, los valores de los parámetros tomarán valores dispersos y la salida de una capa dada no dependerá de un único nodo. Esta técnica de regularización se puede emplear tanto en filtros convolucionales como en capas totalmente conectadas.

- **Finalización automática del entrenamiento**

Muchas herramientas de desarrollo de modelos de aprendizaje profundo tienen un método que permite detener el entrenamiento antes de que el modelo empiece a sobreajustarse. De este modo, se realiza una monitorización de las métricas sobre el conjunto de datos de validación y, tras un número especificado por el desarrollador, de épocas sin mejoras significativas en éstos, detiene el entrenamiento.

1.4.4. Estado del arte en el uso de CNN para resolver problemas de clasificación

La clasificación se basa en, dada una imagen de entrada, atribuirle una etiqueta de salida. Por lo tanto, a la salida de una CNN de clasificación tendremos la probabilidad de que una imagen dada pertenezca a una clase u otra/s.

Las redes de clasificación basadas en CNN han experimentado un gran avance gracias a iniciativas como el reto ILSVRC (*ImageNet Large Scale Visual Recognition Challenge*) [71], se trata de un concurso llevado a cabo entre los años 2010 y 2017 en el que investigadores de todo el mundo compitieron por obtener los mejores resultados en la clasificación de imágenes y detección de objetos de la vida cotidiana. Este concurso se basa en el uso de la base de datos *ImageNet* [72] con más de 14 millones de imágenes etiquetadas en más de 20.000 categorías. De estas, el reto ILSVRC emplea 1,28 millones de imagen de entrenamiento y 50.000 de validación pertenecientes a 1.000 categorías distintas. En la Figura 24, podemos observar cómo estas redes, año tras año, han ido ofreciendo mejores resultados.

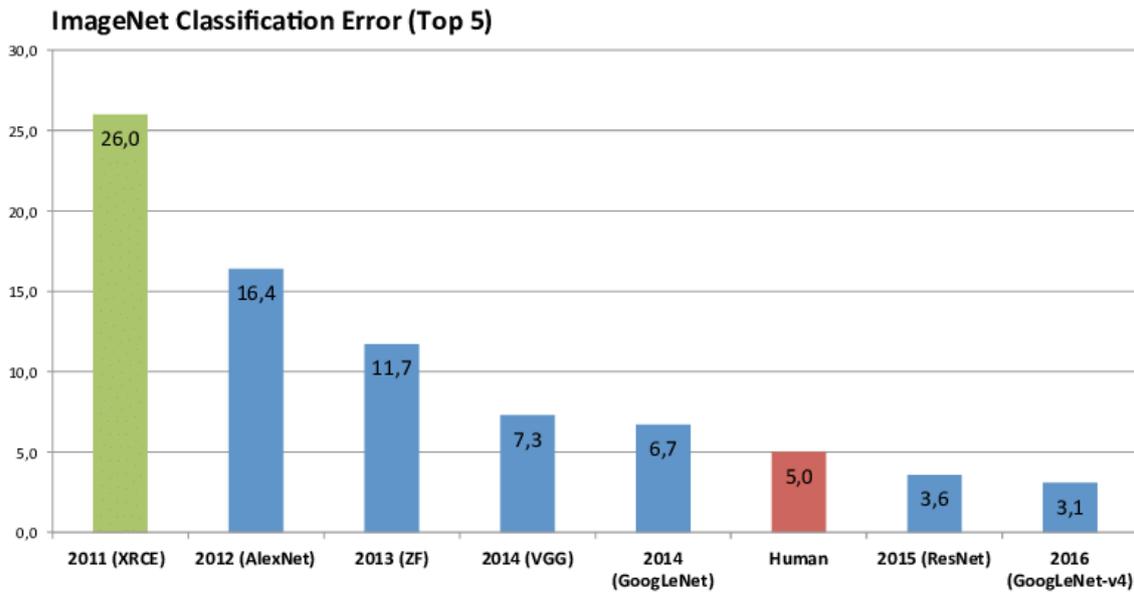


Figura 24. Errores en la clasificación de imágenes de ImageNet a lo largo de las ediciones de ILSVRC. Se puede ver que año tras año este error se ha ido reduciendo hasta incluso superar los resultados de un humano. Top 5 hace referencia a que cada red ha ofrecido la clase correcta entre las 5 clases con más probabilidad. Fuente imagen: seeme.ai.

En el año 2012 se observa una reducción muy significativa del error obtenido por el modelo ganador, en comparación con las tendencias de años anteriores. Esto se produjo gracias al uso, por primera vez en este reto, de CNNs con GPUs.

Además, año tras año se observa una mejora en la clasificación, esta ha ido ligada de un aumento en el número de capas incluidas en cada arquitectura. Así, AlexNet [73] ganadora de la edición de 2012, está formada por 8 capas, mientras que a ResNet [74], de 2015, la construyen 152 capas, aumentando así la complejidad de la red. Este aumento de capas ha ido ligado de las mejoras en los equipos de computación, la mejora en las GPU (unidades de procesamiento gráfico, o, del inglés, *Graphics Processing Unit*) y la introducción de las TPU (unidades de procesamiento de tensores, o, del inglés, *Tensor Processing Unit*) han permitido el entrenamiento de redes más profundas de manera más rápida.

Con el avance de las redes de clasificación han ido apareciendo arquitecturas nuevas que mejoran configuraciones de la arquitectura tradicional.

ResNet [74] introduce el concepto de bloque residual. Esta arquitectura fue la primera que permitió crear redes muy profundas (con muchas capas) controlando el problema del desvanecimiento de gradiente. Este problema se produce en redes muy profundas ya que, durante el proceso de entrenamiento, los pesos, según el algoritmo de propagación hacia atrás, se actualizan a un valor proporcional a la derivada parcial de la función de pérdidas. Éstos, pueden llegar a tomar valores muy bajos impidiendo, en el peor de los casos, que

la red neuronal continúe su aprendizaje, ofreciendo, por tanto, peores resultados que una red con un menor número de capas. Mediante el uso de bloques residuales, este problema se reduce. En estos bloques, se añade una conexión puente que facilita el flujo de información entre una capa y la siguiente (Figura 25). Con la incorporación de estas conexiones puentes, se garantiza que el rendimiento de estos modelos no disminuya, pero puede aumentar ligeramente:

- Añadir capas nuevas no empeorará el rendimiento del modelo ya que la regularización saltará aquellas capas que no sean útiles.
- Si la nueva capa introducida es útil, incluso con la presencia de regularización, los pesos de las capas convolucionales serán distintos de cero y el rendimiento del modelo podrá incrementarse ligeramente.

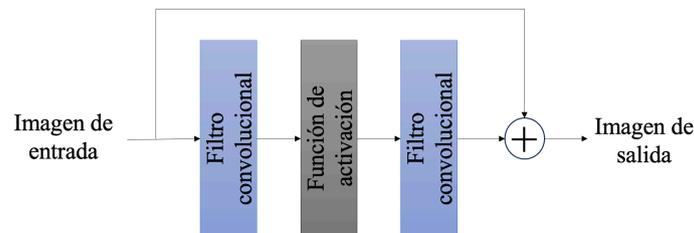


Figura 25. Bloque residual en el que a un bloque convolucional estándar se le añade una conexión puente, combinando la salida de ambas para generar la imagen de salida

Otra arquitectura de alto impacto fue DenseNet [75], presentada en 2017. Ésta sigue la misma idea que ResNet de conectar distintas capas, pero, DenseNet recibe la salida de todas las capas precedentes y las concatena, mientras que ResNet suma la salida de un bloque con la salida del bloque anterior (Figura 26). Se ha visto que, para valores similares de precisión, DenseNet es más eficiente en términos de número de parámetros a ajustar y carga computacional [75].

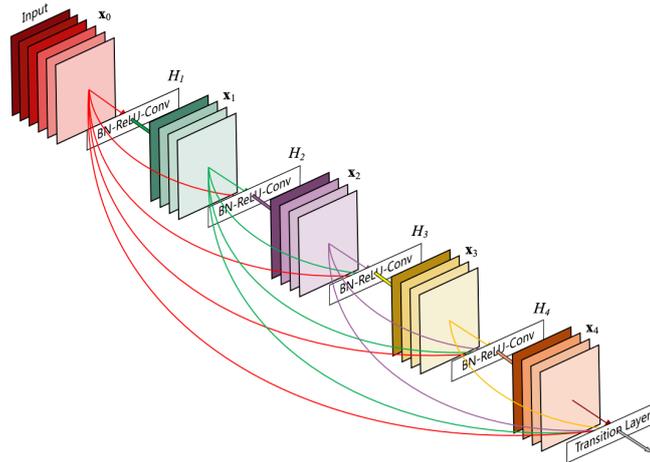


Figura 26. Bloque denso de 5 capas con una tasa de crecimiento $k = 4$. Cada capa toma todos los mapas de características previos [75].

Una de las grandes aplicaciones que tienen estos clasificadores en el sector radiológico es en la mejora del flujo radiológico mediante el desarrollo de herramientas que faciliten el triaje de los pacientes. Por ejemplo, se han propuesto herramientas basadas en CNN para la clasificación de imágenes como, por ejemplo, una radiografía de tórax [76] o una mamografía [77] en función de si presentan patología o no, permitiendo, de este modo, generar listas de priorización. Así, se generan listas de informado en las que en las posiciones superiores se sitúan aquellos pacientes en los que la probabilidad de presencia de patología obtenida con el algoritmo de clasificación es mayor que otras con menor probabilidad que se situarán en posiciones inferiores.

1.4.5. Estado del arte en el uso de CNN para resolver problemas de segmentación

Segmentar una imagen consiste en detectar todos aquellos píxeles que corresponden a una región de interés en concreto. En radiología, esta tarea es de especial relevancia ya que es una tarea que generalmente desempeñan los radiólogos de manera manual o semi-automática, ralentizando, por lo tanto, su flujo de trabajo. Esta segmentación es necesaria para estudios volumétricos de los órganos y lesiones, así como técnica de preproceso para la posterior cuantificación de la región de interés. No solo para el diagnóstico y seguimiento de pacientes en el ámbito radiológico, en radioterapia, es una práctica común la delineación de distintas áreas (isocontornos) para establecer qué regiones del paciente y con qué dosis se van a radiar; así como para la planificación quirúrgica.

Los primeros modelos de CNN resolvían un problema de segmentación como un problema de clasificación. Para ello, se dividía la imagen en distintas regiones, con un

determinado tamaño de ventana (algoritmos de ventana deslizante), y, el píxel central de cada ventana se clasificaba como perteneciente o no a la región de interés [78]. Sin embargo, esta metodología tiene dos inconvenientes principales, por un lado, al dividir la imagen en ventanas, se pierde información contextual del resto de la imagen, por otro lado, al tener que realizar una clasificación píxel a píxel de la imagen, el tiempo de computación se alarga. Además, la clasificación de píxeles vecinos, dada la similitud entre ellos, es redundante. En 2015, Ronneberger et al. introdujeron la arquitectura conocida como U-Net [79] (Figura 27) en la comunidad científica. Esto supuso un cambio de paradigma en la resolución de problemas de segmentación. U-Net es una red neuronal totalmente convolucional (FCN, del inglés *Fully Convolutional Network*), ya que únicamente cuenta con bloques convolucionales. Esta red está formada por una primera parte de codificación, donde, dada una imagen de entrada, se crea un vector de características multidimensional mediante la extracción de características a distintos niveles; y una segunda parte de decodificación, donde el vector generado anteriormente se decodifica a distintos niveles para, finalmente, generar la máscara de segmentación. Esta decodificación se realiza mediante capas de deconvolución. Otra característica fundamental de esta arquitectura es que se emplean conexiones puente entre las capas de codificación y las de decodificación, esto permite tener información a múltiples niveles a la hora de generar la máscara de segmentación.

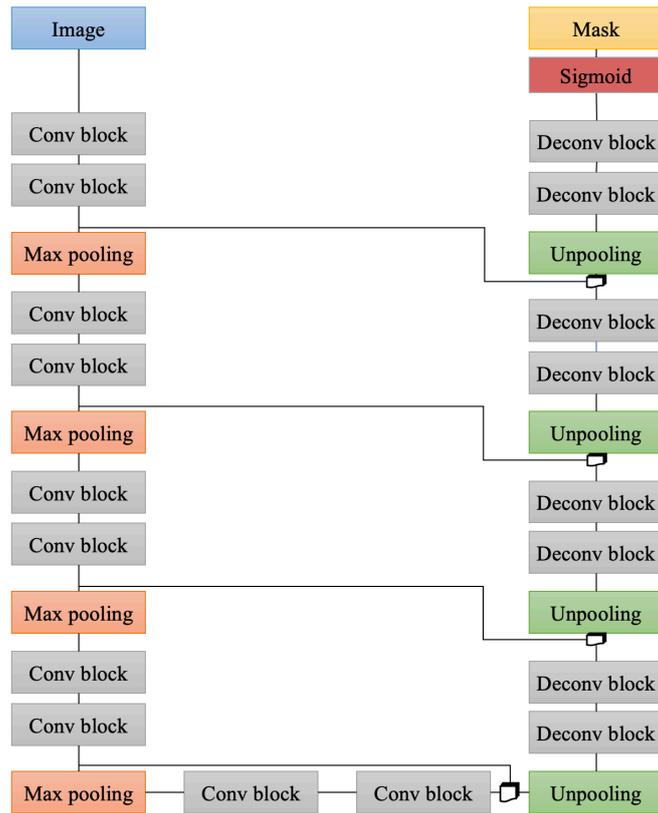


Figura 27. Arquitectura U-Net. A la izquierda tenemos el bloque de codificación, a la derecha el bloque de decodificación y, entre ambos, se incluyen conexiones puentes.

Esta arquitectura, permitió la introducción de algoritmos de segmentación basados en algoritmos de ventana deslizante, pero, con la ventaja de poder segmentar todos los píxeles de la ventana de una, sin necesidad de hacerlo píxel a píxel [80]. Por lo tanto, esta solución permite eliminar el problema de clasificar píxeles redundantes, reduciendo, de esta forma, el tiempo de computación.

La solución de los algoritmos basados en ventana deslizante es de gran utilidad cuando el número de imágenes de las que se dispone para entrenar es reducido, ya que cada ventana cuenta como una muestra independiente. Sin embargo, mantiene el problema de la pérdida de información contextual de la imagen. Por lo tanto, se tiende a ir al uso de redes basadas en FCN en las que se segmenta la imagen en su totalidad [81].

A partir de la introducción de las FCN, diferentes variaciones se han realizado de la misma, entre ellas encontramos:

- Bloques convolucionales 3D: En 2016 Milletari et al., siguiendo la misma idea que Ronneberger et al. en [79], introdujo V-Net [82]. Su principal aportación es el uso de bloques convolucionales 3D. Esto es de especial interés en imagen médica ya que tanto en estudios de TC como de RM se emplean imágenes

volumétricas. Sin embargo, tiene la desventaja de que, así como en las redes 2D cada corte de la imagen cuenta como una muestra independiente a la hora de entrenar un modelo, en las redes 3D, toda la imagen es una única muestra de entrada. Reduciéndose, de este modo, el tamaño del conjunto de datos de entrenamiento. Además, el coste computacional requerido para entrenar un modelo basado en bloques convolucionales 3D es mucho mayor que el correspondiente basado en bloques convolucionales 2D, especialmente en términos de RAM.

- Supervisión profunda [83, 84]: Mediante esta técnica, mapas de características obtenidos previamente en distintos puntos de la red se emplean para generar nuevos mapas de características. Estos se emplean de dos formas distintas: se combinan con la segmentación final y las pérdidas asociadas a estos mapas de segmentación, se ponderan y se suman a la función de pérdidas final. Esto se ha visto que permite una convergencia más rápida de la red durante el proceso de entrenamiento, y una precisión mayor en la segmentación de los bordes de la región de interés [83].
- Puertas de atención [85]: su principal objetivo es hacer que la red preste especial atención en aquellas regiones de interés, eliminando aquellas de menor interés. Son de gran utilidad en aquellos casos en los que hay un gran desbalanceo de los datos, significando esto en un problema de segmentación, que el número de píxeles correspondientes a la máscara (región segmentada) difiere mucho del número de píxeles correspondientes al fondo.
- Uso de otro tipo de bloques convolucionales:
 - Bloques residuales [74], que, como hemos visto en el apartado anterior, están diseñados para permitir el entrenamiento de CNN muy profundas, es decir, con un número elevado de capas convolucionales.
 - Filtros convolucionales dilatados [86] en los que el filtro convolucional se genera introduciendo relleno entre los elementos del filtro. La tasa de dilatación define cuándo se dilata el filtro. De esta forma, se incrementa el campo receptivo y, por tanto, la información contextual que se tiene en cuenta es mayor.

Otra arquitectura de segmentación que también ha demostrado ofrecer buenos resultados es Mask R-CNN [87]. Ésta se basa en la arquitectura R-CNN [88] cuyo objetivo es la

Capítulo 1. Introducción

detección y clasificación de objetos, por lo tanto, a la salida de ésta se obtiene las coordenadas de la región detectada y su clase. Mask R-CNN añade la segmentación dentro de la región detectada. Mientras que U-Net realiza una segmentación semántica, en la que la red le asigna una clase a cada píxel de la imagen, Mask R-CNN realiza segmentación por instancia, en la que la red asigna una etiqueta a cada objeto individual dentro de la imagen. De esta forma, si en una misma imagen hay varias ocurrencias de un mismo objeto, cada una de ellas se etiquetará con una clase diferente (Figura 28).

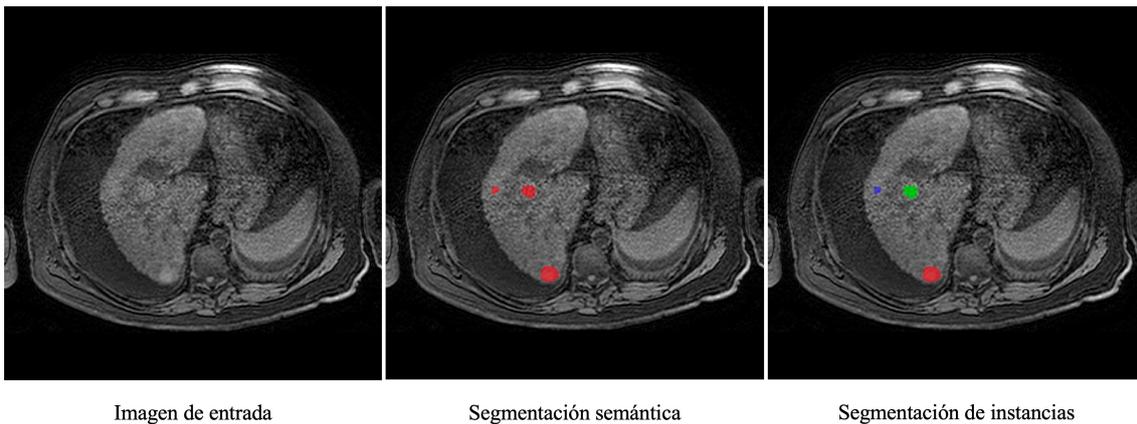


Figura 28. Diferencia entre segmentación semántica (medio) y segmentación de instancias (derecha).

Capítulo 2. Hipótesis y objetivos

Hipótesis

Los biomarcadores de imagen y su combinación con técnicas de aprendizaje profundo permiten la automatización de la caracterización y cuantificación de determinados órganos y lesiones. Todo esto, permite la generación de flujos de análisis automático que se pueden integrar en la práctica clínica para la mejora y optimización del flujo radiológico. El síndrome metabólico y las enfermedades hepáticas difusas y focales pueden favorecerse de este tipo de técnicas mediante la generación de las conocidas biopsias virtuales, una estimación precisa de los cambios anatomopatológicos que suceden en el organismo debido a estas enfermedades. Así, en la presente tesis doctoral, se pretende desarrollar una metodología novedosa para la caracterización automática del parénquima hepático en pacientes con enfermedad hepática difusa y focal, así como para el análisis de la composición corporal mediante la combinación de técnicas de aprendizaje profundo y modelado matemático.

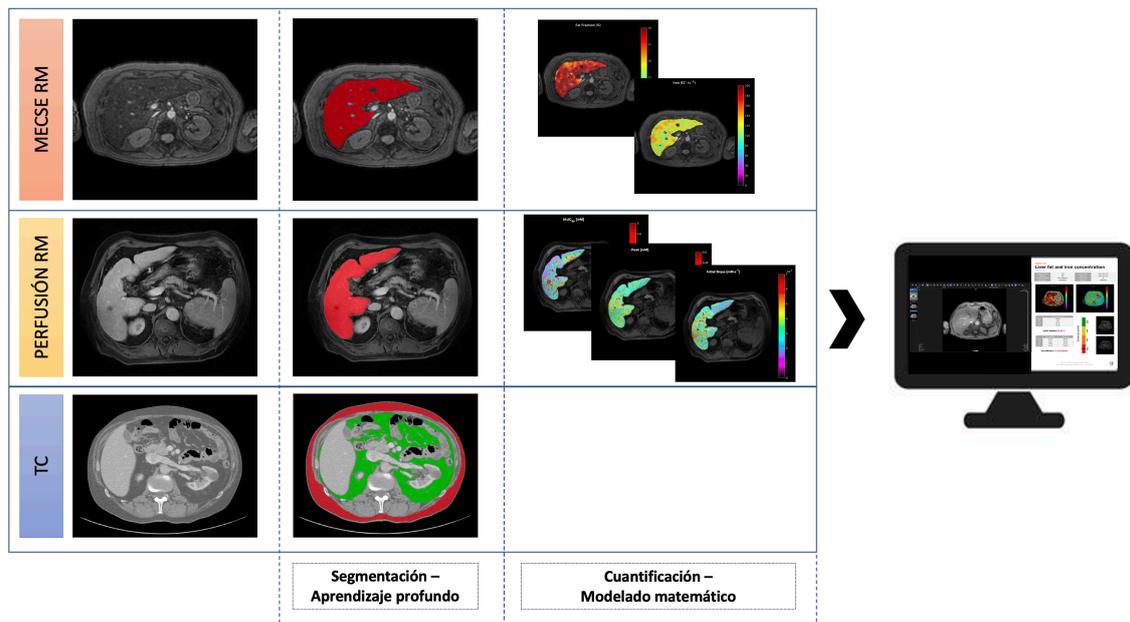


Figura 29. Diagrama del flujo de la tesis doctoral.

Objetivos

El principal objetivo de la presente tesis doctoral es:

- Diseñar, implementar y validar metodologías de análisis de imagen basadas en técnicas de inteligencia artificial (IA) y extracción de biomarcadores de imagen para la caracterización automática del parénquima hepático y el análisis de la distribución de grasa para el estudio de enfermedades hepáticas, tanto difusas de origen metabólico como focales.

Los objetivos secundarios que se establecieron para la correcta consecución del objetivo principal son:

- Recopilar y anotar las bases de datos necesarias para el desarrollo de los diferentes algoritmos implicados. Esto es, recolección de estudios y segmentación manual del parénquima hepático en estudios de resonancia magnética (RM) con secuencia multi-eco con desplazamiento químico (MECSE) en sujetos sanos y con enfermedad hepática difusa; recolección de estudios y segmentación manual del parénquima hepático en estudios de resonancia magnética con secuencia de perfusión (DCE) en pacientes con carcinoma hepatocelular; y recolección de estudio y segmentación manual de la grasa visceral y subcutánea en estudios de tomografía computarizada (TC) de sujetos sanos y pacientes con síndrome metabólico.
- Entrenar, validar y testear modelos de inteligencia artificial para la segmentación automática de las regiones de interés definidas en el punto anterior.
- Implementar métodos de cuantificación para la caracterización del parénquima hepático dependiendo de la secuencia disponible, esto es, cuantificación de la fracción de grasa y sobrecarga de hierro en secuencias de RM MECSE; y estudio semicuantitativo de perfusión en secuencias de RM de perfusión.
- Integrar en práctica clínica todos los desarrollos realizados a lo largo de la tesis doctoral de forma que los resultados generados por los modelos se encuentren en formatos estándar y puedan ser utilizados en los centros médicos.

Capítulo 3. Materiales y métodos

Capítulo 3. Materiales y métodos

En este capítulo se profundiza en los materiales de los que partimos para el desarrollo de los modelos propuestos, así como, en la metodología diseñada y seguida para la consecución de los objetivos definidos.

Los desarrollos realizados se pueden englobar en tres grandes bloques cuyas metodologías se encuentran representadas en las Figuras 30-32.

- Segmentación y cuantificación hepática en secuencias MECSE de RM.

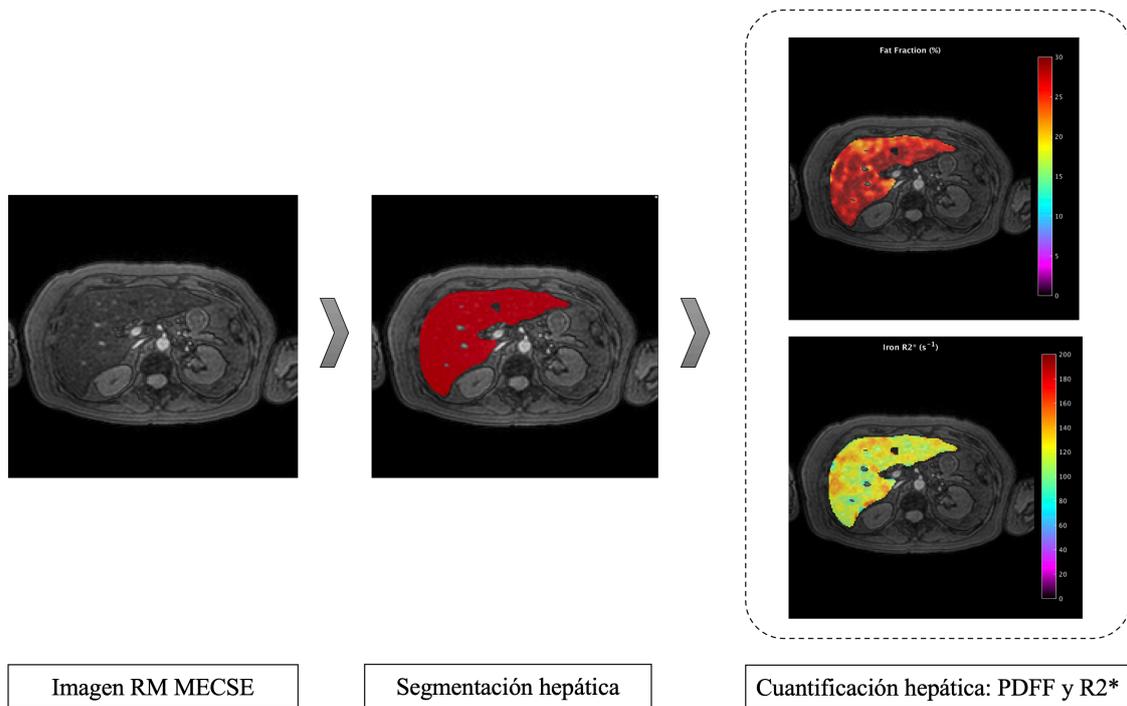


Figura 30. Metodología para la caracterización hepática en secuencias MECSE de RM. Dada dicha secuencia, se emplea un modelo basado en CNN para la segmentación del hígado y, a continuación, se cuantifica la fracción de grasa el R2* (hierro). Esta cuantificación se realiza vóxel a vóxel dentro de toda el área segmentada generando mapas paramétricos.

- Segmentación y volumetría de la grasa visceral y subcutánea en imágenes de TC.

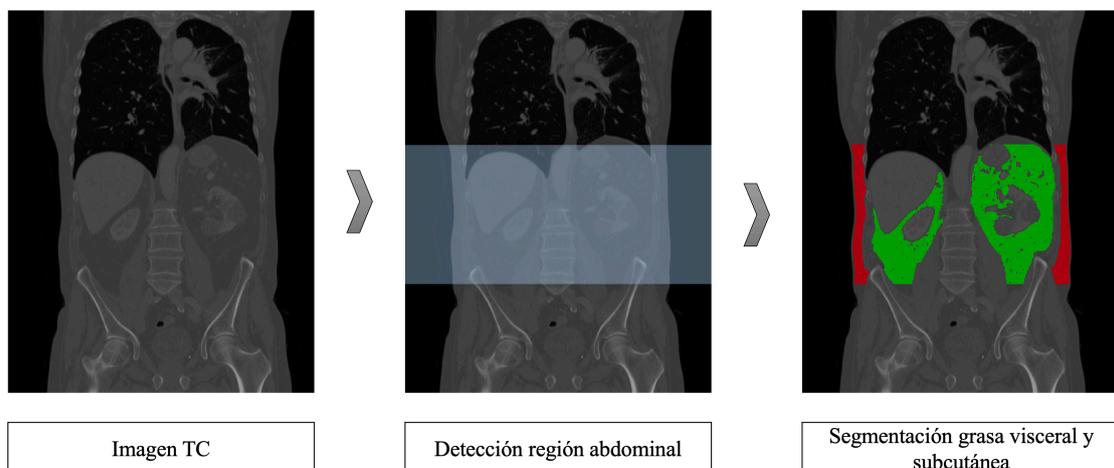


Figura 31. Metodología para la cuantificación de los volúmenes de grasa visceral y subcutánea en imágenes de TC. Dada una imagen de TC con campo de visión arbitrario, se emplea un modelo basado en CNN para identificar la

Capítulo 3. Materiales y métodos

región abdominal seguido de otro modelo de CNNs para la segmentación automática de la grasa visceral (verde) y subcutánea (rojo). Finalmente, se puede cuantificar el volumen de cada una de estas regiones.

- Segmentación y cuantificación hepática en secuencias de perfusión de RM.

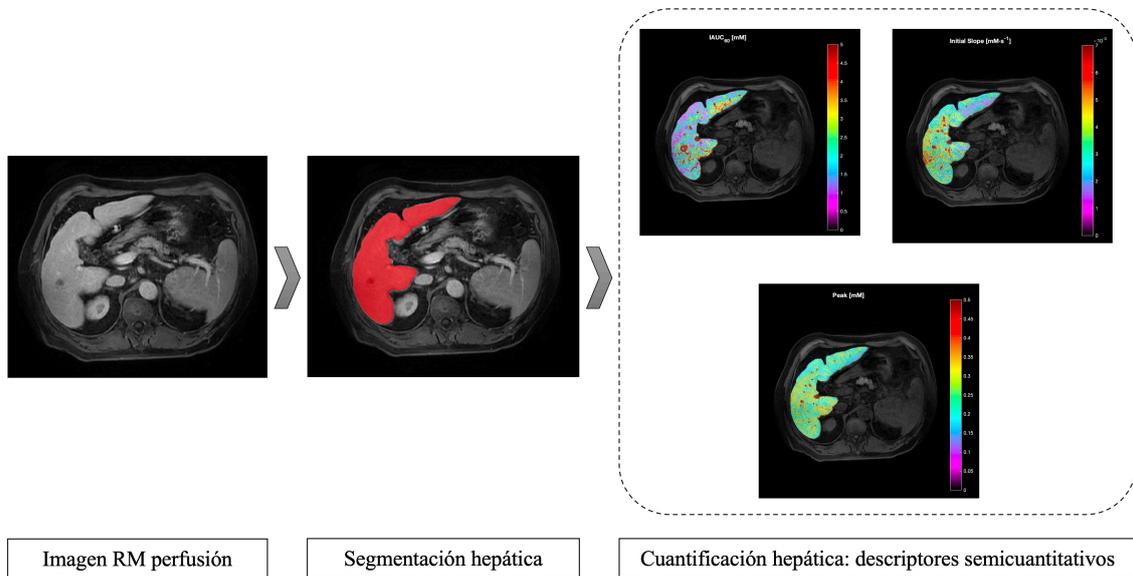


Figura 32. Metodología para el análisis automático del hígado en una secuencia de RM de perfusión abdominal. Dada dicha secuencia, se emplea un modelo basado en CNN para la segmentación del hígado y, a continuación, se cuantifica el área bajo la curva de captación de contraste a los 60 segundos (arriba izquierda), la pendiente inicial de captación de contraste (arriba derecha) y el pico máximo de concentración de contraste (abajo). Esta cuantificación se realiza vóxel a vóxel dentro de toda el área segmentada generando mapas paramétricos.

3.1. Bases de datos

En esta sección se describen las bases de datos recopiladas y cómo se han particionado para el entrenamiento, validación y testeo de cada uno de los desarrollos realizados en la tesis doctoral.

La creación de la base de datos que se va a emplear para el entrenamiento de los modelos de aprendizaje profundo es una de las fases más importantes ya que de ella va a depender la capacidad de aprendizaje de las CNN. A continuación, vamos a describir el proceso de generación de la base de datos utilizada para el entrenamiento y validación de cada uno de los modelos desarrollados en la presente tesis doctoral (Figura 33).

En cada uno de los problemas a resolver, partimos de una base de datos etiquetada, de ésta se realizaron tres particiones de datos diferentes para garantizar el correcto diseño de la arquitectura dada, la ausencia de sobreajuste y la capacidad de generalización del modelo entrenado. Estos subconjuntos de datos son:

- **Base de datos de entrenamiento y validación**

Se siguió un proceso de validación cruzada en k lotes, concretamente se emplearon 5 lotes. Por lo tanto, el 80% del total de datos etiquetado se subdividió en 5 particiones y, en un proceso iterativo durante 5 iteraciones, se emplearon 4 particiones para entrenar y la restante para validar. Esto permitió evaluar la capacidad de la arquitectura diseñada a generalizar independientemente del conjunto de datos de entrenamiento. De este modo, para seleccionar la mejor configuración se eligió aquella que, de media, ofreciera mejores resultados durante el proceso de validación cruzada con elevada estabilidad, es decir, que existieran pocas diferencias entre los resultados de las iteraciones con peores y mejores resultados.

- **Base de datos de test**

Conjunto de imágenes completamente independiente a las empleadas para el entrenamiento y validación del modelo generado. Esta base de datos, en cada caso, supuso el 20% del total de estudios de imagen anotados disponible.

En el conjunto de datos etiquetado de imágenes MECSE de RM, disponíamos de imágenes procedentes de distintos centros y/o distintos escáneres y protocolos de adquisición. En este caso, se emplearon las imágenes adquiridas en estos centros

independientes para evaluar la reproducibilidad del modelo desarrollado ante imágenes procedentes de centros y escáneres distintos a los empleados para el ajuste del modelo.

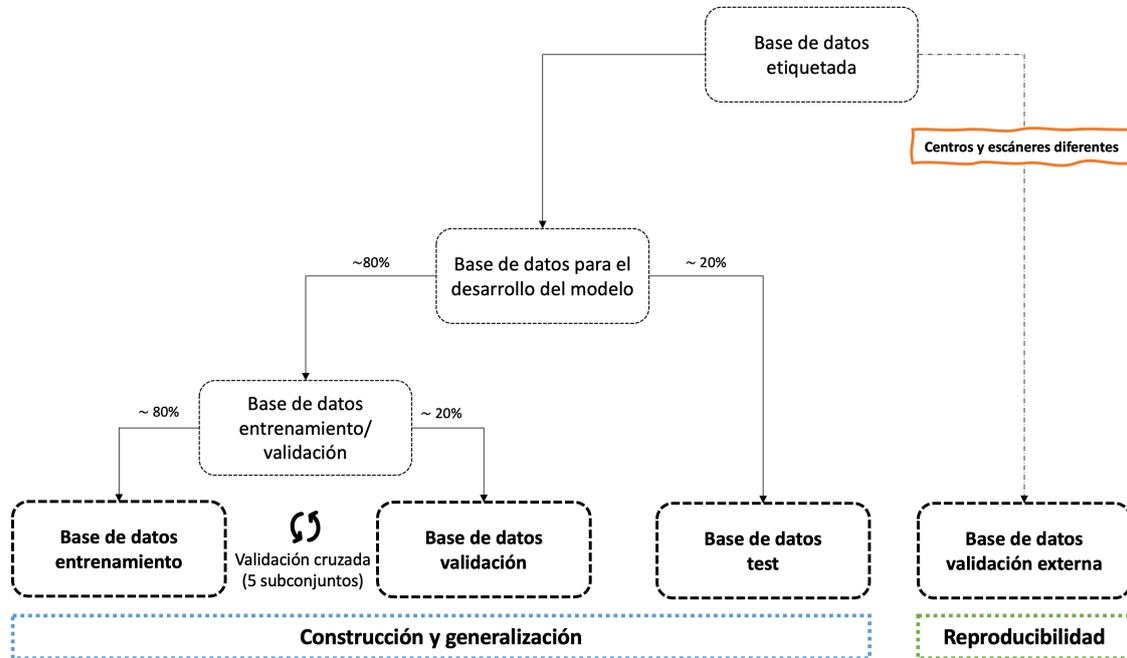


Figura 33. Diagrama creación base de datos para el desarrollo y apropiada validación del modelo generado.

A continuación, vamos a describir las bases de datos disponibles para la resolución de cada uno de los desarrollos propuestos.

3.1.1. Segmentación automática hepática en secuencias MECSE

Esta base de datos consta de imágenes provenientes de cuatro centros distintos (Hospital Universitario y Politécnico La Fe - HULAFE, Valencia, España; Centro Hospitalar do Porto - PORTO, Oporto, Portugal; Laboratorio Diagnomed, - DIAGNOMED Concepción, Chile; y ARGUS Diagnóstico Médico - ARGUS, Buenos Aires, Argentina) adquiridas en tres escáneres de RM diferentes (Philips 3T, Philips 1,5T y Siemens 3T). Podemos observar, por lo tanto, que disponíamos de imágenes de dos fabricantes diferentes (Philips y Siemens) y dos campos magnéticos diferentes (1,5T y 3T). Las bases de datos de PORTO y HULAFE forman parte de las bases de datos presentadas en [26] y [89], respectivamente.

El entrenamiento de la CNN se desarrolló con imágenes provenientes de dos de los centros (HULAFE y PORTO) adquiridas todas ellas con un escáner de Philips de 3T. El

Capítulo 3. Materiales y métodos

testeo de la red, por otro lado, se llevó a cabo con imágenes de los cuatro centros. De este modo, se evaluó la capacidad de generalización de la CNN entrenada a imágenes provenientes de otros centros, con distintos escáneres, distintos campos magnéticos y distintos protocolos de adquisición a los empleados para el entrenamiento del modelo (validación externa). En la Tabla 2 se encuentra un resumen de las diferencias en los parámetros de adquisición entre las imágenes disponibles en la base de datos. Así, se emplearon 107 casos para el entrenamiento y validación del modelo, 46 para el test interno y, para la validación externa se emplearon dos subconjuntos de 19 y 10 casos respectivamente.

Tabla 2. Protocolos de adquisición de los estudios de RM empleados tanto para el entrenamiento/validación, testeo y validación externa del modelo de segmentación automática hepática en secuencias MECSE. La primera columna incluye los parámetros empleados en los escáneres de dos centros diferentes (HULAFE y PORTO) empleados para el entrenamiento y testeo del modelo, la segunda y tercera columna incluye los parámetros empleados en ARGUS y DIAGNOMED, respectivamente, ambos empleados para la validación externa del modelo desarrollado.

Parámetro	Valor		
	Entrenamiento / Validación y testeo	Validación externa	
		Institución 1	Institución 2
País	España y Portugal	Argentina	Chile
Pacientes, n	153 (107 + 46)	19	10
Hombres (Mujeres), n	81 (73)	12 (7)	6 (4)
Edad en años, mediana [rango]	56 [48-63]	53 [27-66]	59.5 [26-79]
Fabricante	Philips	Siemens	Philips
Modelo	Achieva	Skyra	Achieva
Campo magnético, T	3	3	1.5
Ángulo de magnetización, °	10	10	5
Tiempo de repetición, s	9 – 12	32 – 120	15
Ancho de banda, Hz/píxel	2132 – 2880	1565	2681
Tamaño matriz	160x160 – 224x224	256x216	192x192
Número de cortes	24 – 34	4 – 5	31 – 40
Tamaño píxel, mm²	1,7 x 1,7 – 2,0 x 2,0	1,6 x 1,6	2,3 x 2,3

Capítulo 3. Materiales y métodos

Espaciado entre cortes,	6,0 – 7,3	15,0 – 17,1	5,0 – 5,2
Espesor de corte, mm	6,0 – 7,0	7,0 – 8,0	10,0 – 10,5
Número de ecos	12	6 ó 12	16
Primer TE, s	0,9 – 1,0	1,0 – 1,2	1,1 – 1,2
Último TE, s	7,7 – 8,7	9,7 – 14,4	13,8
Espaciado entre ecos, s	0,6 – 0,7	1,2 – 1,4	0,8
Duración adquisición, s	13,5 – 35,2	14,4 – 27,5	15,3 – 19,1

3.1.1.1. Cuantificación y validación clínica

Adicionalmente a la base de datos empleada para el entrenamiento del modelo de segmentación automática, para el ajuste y validación del método de cuantificación, así como para la validación clínica de toda la metodología (segmentación seguida de cuantificación), se empleó una base de datos independiente. Se trata de una base de datos de 165 estudios de RM con la secuencia MECSE proveniente del Hospital Universitario y Politécnico La Fe de Valencia. Estos pacientes, diagnosticados con enfermedad hepática crónica, disponían tanto de biopsia hepática como de estudio de RM [89,90]. En [91] podemos encontrar la etiología de las patologías, así como otros datos clínicos que podemos encontrar en la base de datos. En este trabajo nos centraremos en el análisis de la RM y su relación con lo obtenido a partir de la biopsia.

En la Tabla 3 encontramos los parámetros de adquisición de las secuencias MECSE incluidas en esta base de datos.

Tabla 3. Protocolos de adquisición de las secuencias MECSE de RM empleadas tanto para el desarrollo y validación del método de cuantificación, así como para la validación clínica de la metodología desarrollada (segmentación + cuantificación).

Parámetro	Valor
País	España
Pacientes, n	165
Hombres (Mujeres), n	69 (96)
Edad en años, mediana [rango]	54 [18-78]

Capítulo 3. Materiales y métodos

Fabricante	Philips
Modelo	Achieva
Campo magnético, T	3
Ángulo de magnetización, °	10
Tiempo de repetición, s	9 – 9,32
Ancho de banda, Hz/píxel	2714 – 2939
Tamaño matriz	192x192
Número de cortes	23 – 35
Tamaño píxel, mm²	1,95 x 1,95
Espaciado entre cortes, mm	6,00
Espesor de corte, mm	6,00
Número de ecos	12
Primer tiempo de eco, s	0,88 – 1,00
Último tiempo de eco, s	7,89 – 8,23
Espaciado entre ecos, s	0,64 – 0,66
Duración adquisición, s	8,53 – 8,89

Esta base de datos estaba formada tanto por estudios de RM como de muestra histológica de la biopsia hepática que se realizó a los pacientes. De la evaluación histológica se disponía de dos tipos de información:

- Estadiaje histológico de la esteatosis y depósitos de hierro:
 - Esteatosis: estadios S0-S3 empleando el sistema de gradación de la esteatosis de acuerdo con el sistema *Non-alcoholic Steatohepatitis (NASH) Clinical Research Network (CRN)* [92]. En la base de datos disponible disponíamos del siguiente número de casos por estadio: 70 (S0), 33 (S1), 33 (S2), 28(S3).
 - Depósitos de hierro: estadios Fe0-Fe4 empleando el sistema Scheuer [93]. Para los depósitos de hierro, disponíamos del siguiente número de casos por estadio: 129 (Fe0), 10 (Fe1), 8 (Fe2), 11 (Fe3), 6 (Fe4).

- Proporción de área (%) con presencia de grasa y hierro obtenida tras la digitalización de la muestra histológica [90].

En [91] podemos encontrar los resultados obtenidos de este análisis histológico.

3.1.2. Segmentación automática de grasa abdominal en imágenes de TC

Para el desarrollo del modelo de segmentación automática de grasa abdominal en imágenes de TC, se empleó una base de datos etiquetada formada por 59 estudios. De estos, 47 se emplearon para el entrenamiento y validación de los modelos, mientras que los 12 estudios restantes se emplearon para el test externo.

Esta base de datos se empleó tanto para el modelo de clasificación y detección de la región abdominal como para la propia segmentación de la grasa visceral y subcutánea.

Los estudios de TC provienen de dos centros, el Hospital Universitario y Politécnico La Fe (Institución 1) y otro centro (Institución 2) cuya procedencia se desconocía al tratarse de casos provenientes de un estudio multicéntrico completamente anonimizados cuyo centro de procedencia también se eliminó de los metadatos de las imágenes. Los principales parámetros de las imágenes incluidas en este estudio se encuentran resumidos en la Tabla 4.

Tabla 4. Protocolos de adquisición de los estudios de TC empleados tanto para el entrenamiento/validación como para el testeo del modelo de segmentación automática de grasa visceral y subcutánea. El signo “-“ indica valor desconocido.

Parámetro	Institución 1	Institución 2
País	España	-
Sujetos, n	33	26
Hombres (Mujeres), n	-	16 (10)
Edad en años, mediana (rango)	-	68 [37 – 78]
Fabricante	-	GE Medical Systems
Modelo	-	BrightSpeed S
Corriente del tubo (mA)	-	162 – 318
Tamaño matriz	512x512 / 768x768	512x512
Número de cortes	167 – 870	28 – 133
Tamaño píxel, mm ²	0,55x0,55 – 0,98x0,98	0,65x0,65 – 0,83x0,83

Capítulo 3. Materiales y métodos

Espaciado entre cortes, mm	1 / 1,5	5
Espesor de corte, mm	-	5

3.1.2.1. Clasificación de la región abdominal

El modelo de clasificación de la región abdominal trataba de diferenciar cortes transversales (imagen 2D) abdominales de cortes transversales del resto del cuerpo. Así, la base de datos de entrenamiento estaba formada por 3.155 imágenes de la región abdominal y 9.972 del resto del cuerpo. Para balancear ambas clases, se seleccionaron de manera aleatoria 3.155 imágenes del grupo de imágenes del resto del cuerpo. De este modo, finalmente se realizó el entrenamiento y validación con 6.310 imágenes.

En cuanto al test externo, se emplearon un total de 2.997 imágenes 2D, de las cuales 694 pertenecían a la región abdominal mientras que las 2.303 restantes no pertenecían a la región abdominal.

3.1.2.2. Segmentación de grasa visceral y subcutánea

Para el entrenamiento y validación del modelo de segmentación automática se emplearon 47 estudios de TC de los cuales se disponía de 680 imágenes 2D con la grasa visceral y subcutánea manualmente segmentada. Por otro lado, para el testeo del modelo final obtenido, se emplearon 12 estudios de TC con 170 imágenes 2D segmentadas.

3.1.3. Segmentación automática hepática en secuencias de perfusión

Para el estudio de lesiones hepáticas focales se desarrolló otro método de segmentación automática del hígado en secuencias de perfusión (DCE) ya que se trata de una secuencia adquirida en el estudio de este tipo de lesiones.

Se recopiló una base de datos de 109 estudios de DCE provenientes del Hospital Universitario y Politécnico La Fe de Valencia. De éstos, 76 se emplearon para el entrenamiento y validación de los modelos mientras que los 33 estudios restantes se emplearon para el test final. En la Tabla 5 podemos encontrar un resumen de los parámetros de adquisición empleados en la base de datos recopilada.

Capítulo 3. Materiales y métodos

Tabla 5. Protocolos de adquisición de los estudios de RM empleados para el entrenamiento/validación y testeo del modelo de segmentación automática hepática en secuencias de perfusión.

Parámetro	Valor				
	Entrenamiento / Validación			Testeo	
Fabricante	GE	Siemens	Philips	GE	Siemens
Pacientes, n	47	27	2	28	5
Modelo (n casos)	Optima MR360 (20) Signa HDxt (19) Signa Excite (5) Optima MR450w (3)	Symphony (14) Magnetom (7) Avanto (3) Sonata (3)	Achieva	Optima MR360 (19) Signa HDxt (7) Signa Excite (2)	Symphony
Campo magnético, T (n casos)	1,5 (36) 3 (11)	1,5	1,5	1,5	1,5
Hombres (Mujeres), n	36 (11)	21 (6)	2 (0)	21 (7)	3 (2)
Edad en años, mediana [rango]	60 [49, 81]	59 [51, 81]	53,5 [51, 56]	65 [49, 84]	74 [59, 83]
Ángulo de magnetización, °	12 – 80	10 – 70	10	12 – 15	10
Tiempo de repetición, s	3,3 – 6,2	2,9 – 5,0	4,0	3,3 – 4,7	4,3 – 5,0
Tiempo de eco, s	1,3 – 3,0	1,1 – 4,8	2,0	1,6 – 2,1	2,0 – 2,4
Ancho de banda, Hz/píxel	244,1 – 976,6	250,0 – 700,0	433,0	244,1 – 488,3	300,0 – 415,0
Tamaño matriz	256x256 – 512x512	168x224 – 448x512	448x448	256x256 – 512x512	384x512 – 448x512
Número de cortes	24 – 244	18 – 128	100	92 – 124	72 – 112
Tamaño píxel, mm²	0,7x0,7 – 1,7x1,7	0,7x0,7 – 1,6x1,6	1,0x1,0	0,7x0,7 – 1,7x1,7	0,7x0,7 – 0,8x0,8
Espaciado entre cortes, mm	1,7 – 9,0	9,7 – 11,6	2,0	1,7 – 2,2	2,0
Espesor de corte, mm	3,4 – 7,5	2,0 – 8,0	4,0	3,4 – 4,4	2,2 – 3,0
Número dinámicos	4 – 9	3 – 6	4	5 – 8	4 – 5

3.2. Anotación

En esta sección se describe el proceso seguido para la anotación de las bases de datos disponibles, esto es, la segmentación manual de las regiones de interés que se van a emplear como referencia para el entrenamiento, validación y testeo de los diferentes métodos propuestos.

Para poder entrenar los modelos de segmentación basados en aprendizaje supervisado, fue necesario realizar su anotación. De este modo, cada uno de los casos de las bases de datos se segmentaron de manera manual empleando el software ITKSnap². Este proceso se realizó de manera iterativa, es decir, en primer lugar, se segmentó el 10% de los casos de manera completamente manual, con éstos se entrenó un modelo preliminar para realizar la pre-segmentación del siguiente 10%, el cual se editó adecuadamente, y así, sucesivamente, hasta que se obtuvo la segmentación de la totalidad de casos. Mediante este procedimiento se disminuyó el tiempo de etiquetado. Las anotaciones resultantes fueron supervisadas por un radiólogo con más de 20 años de experiencia.

A continuación, se presentan en detalle las anotaciones empleadas para cada uno de los problemas a resolver.

3.2.1. Segmentación automática hepática en secuencias MECSE

Para el entrenamiento del modelo de segmentación hepática en secuencias MECSE, se segmentó el hígado corte a corte en la imagen correspondiente al primer TE ya que ofrece una mayor relación señal-ruido. En un estudio previo [81], también se demuestra un mejor rendimiento empleando una CNN entrenada con la imagen del primer tiempo de eco.

La segmentación se realizó delineando el parénquima hepático, evitando la vesícula biliar y vasos de elevado calibre, como la aorta, la vena cava, las arterias hepáticas, la vena porta o las principales ramas intrahepáticas.

La Figura 34 muestra un ejemplo de segmentación del parénquima hepático en diferentes cortes anatómicos.

² <http://www.itksnap.org/>

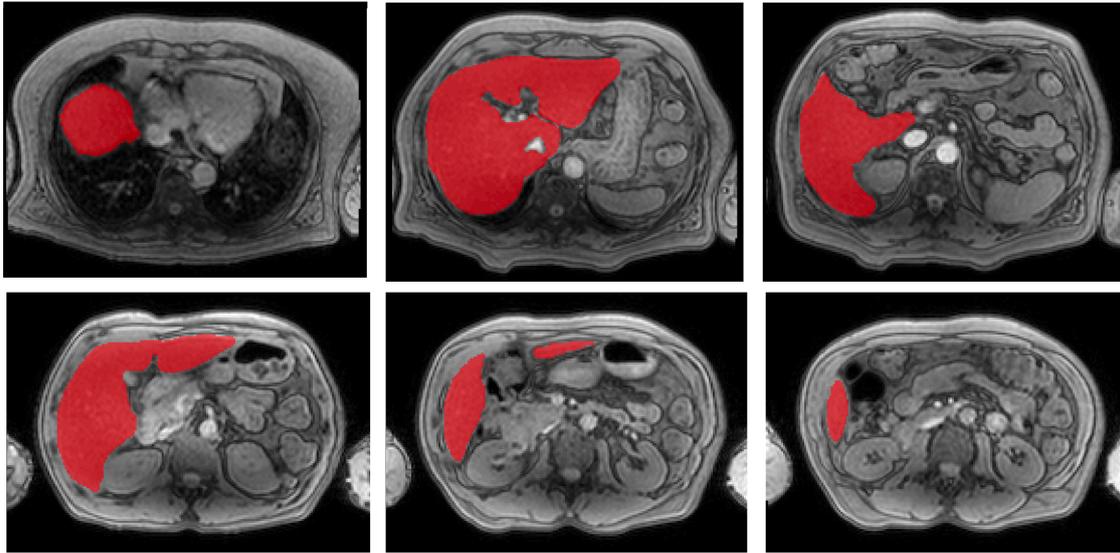


Figura 34. Ejemplo de segmentación manual del parénquima hepático en secuencia MECSE en diferentes cortes anatómicos.

3.2.2. Segmentación automática de grasa abdominal en imágenes de TC

Para el entrenamiento del modelo de segmentación automática de grasa abdominal (grasa visceral y grasa subcutánea) se realizó un etiquetado en dos pasos. En primer lugar, dado que las imágenes de TC suelen tener coberturas anatómicas diferentes, se realizó una diferenciación de cortes pertenecientes a la región abdominal, definida como aquella que cubre la región entre el diafragma (primer corte en el que no vemos región pulmonar) y el disco intervertebral situado entre las vértebras L4-L5, de aquellos que no lo eran. De este modo, se formaron dos grupos de imágenes anotadas: imágenes de la región abdominal e imágenes de fuera de la región abdominal (Figura 35). En segundo lugar, se realizó la segmentación manual de ambas estructuras en cada uno de los casos únicamente en cortes pertenecientes a la región de interés (Figura 36). Dado que las imágenes de TC suelen tener una elevada resolución entre cortes, cortes consecutivos ofrecen información similar. De este modo, con el objetivo de aumentar el número de sujetos diferentes, en lugar de segmentar toda la región abdominal se realizó la segmentación de uno de cada cinco cortes. Como se ha comentado anteriormente, se segmentaron dos compartimentos grasos: la grasa visceral (para la que se empleó la etiqueta 1) y la grasa subcutánea (para la que se empleó la etiqueta 2).

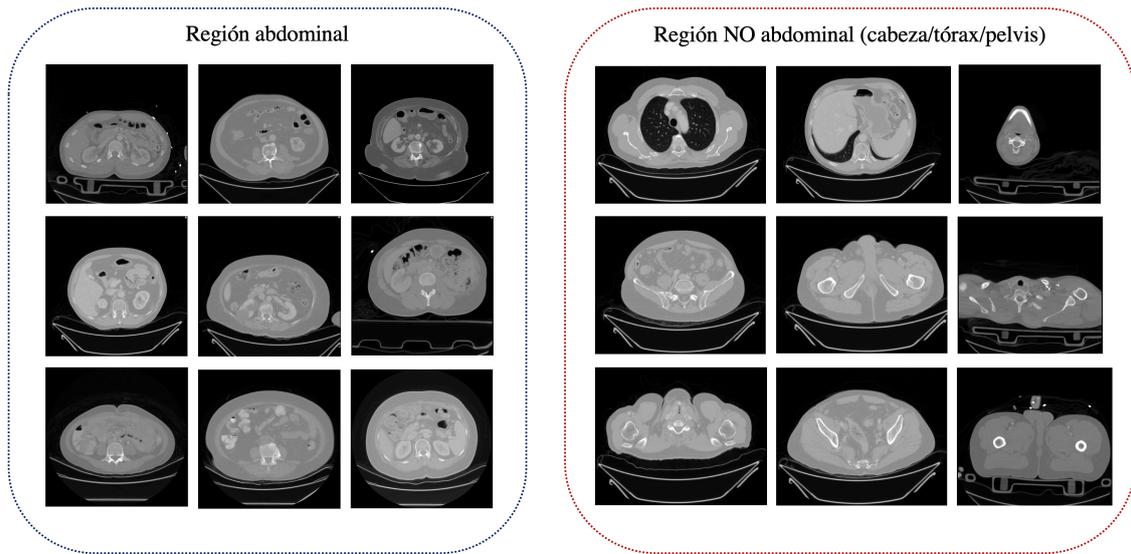


Figura 35. Generación de la base de datos para diferenciar los cortes de TC pertenecientes a la región abdominal (región entre el final del diafragma y el disco intervertebral L4-L5) frente aquellos fuera de esta región.

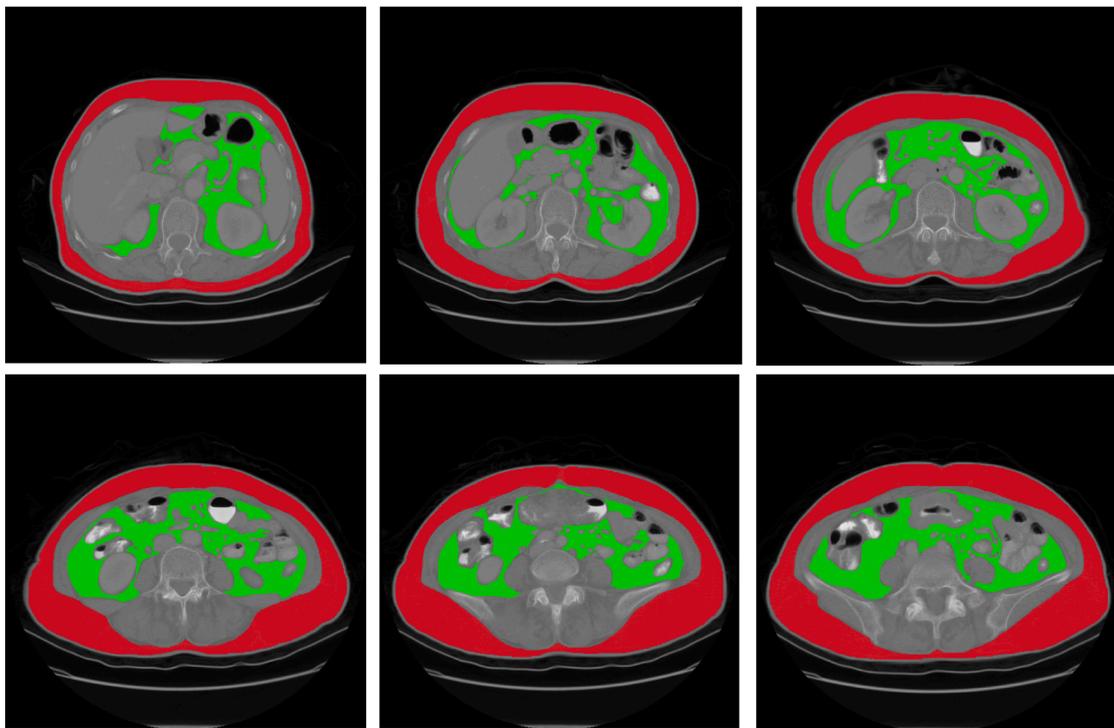


Figura 36. Ejemplo de segmentación manual de la grasa visceral (verde) y subcutánea (rojo) en una imagen de TC en diferentes cortes anatómicos.

3.2.3. Segmentación automática hepática en secuencias de perfusión

Para la segmentación hepática en secuencias de perfusión se seleccionó como imagen de referencia la fase arterial y, de igual modo que para la segmentación hepática en

Capítulo 3. Materiales y métodos

secuencias MECSE, se segmentó manualmente el parénquima hepático evitando la vesícula biliar y los vasos de gran calibre.

La Figura 37 muestra un ejemplo de segmentación hepática en secuencias de perfusión en distintos cortes anatómicos.

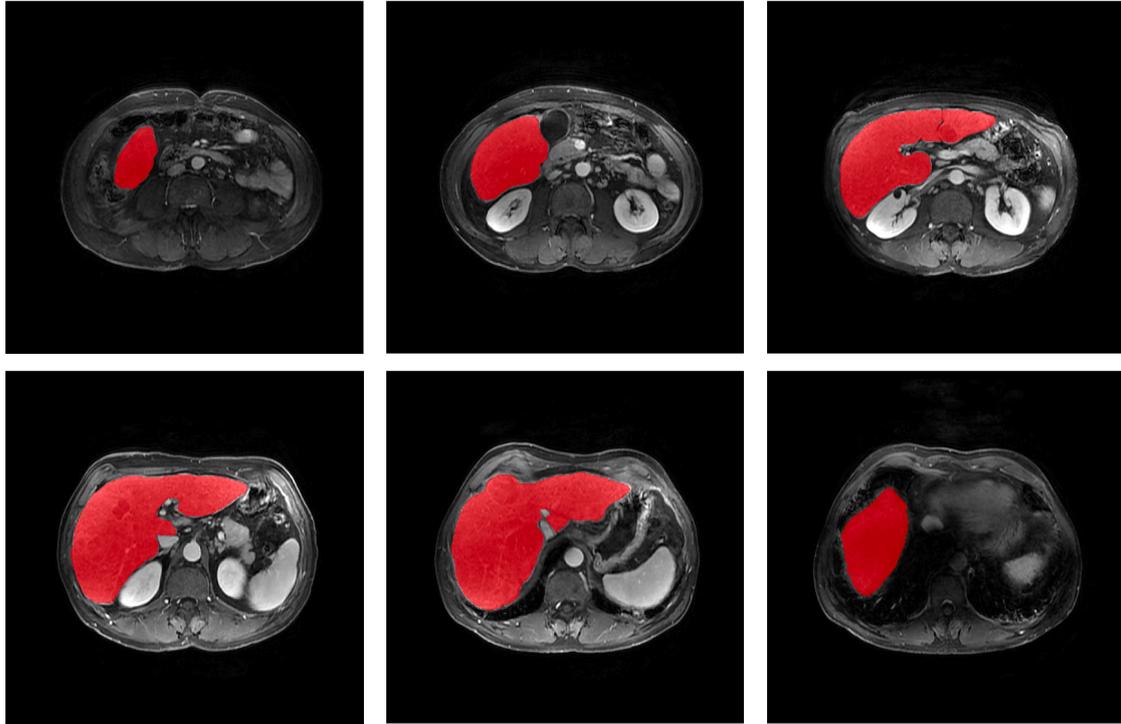


Figura 37. Ejemplo de segmentación manual del tejido hepático en una secuencia de perfusión a distintos cortes anatómicos.

3.3. Preprocesado

En esta sección se describe la metodología seguida para la preparación de las imágenes previo a su uso en las CNN. Este preprocesado incluye, por un lado, la normalización de la imagen y, por otro, su re-escalado.

3.3.1. Normalización

Dada una misma secuencia podemos encontrarnos diferencias en los protocolos de adquisición en los distintos centros, esto es, diferentes tiempos de eco, tiempos de repetición, ángulos de magnetización, corriente del tubo en TC, tamaño de matriz de adquisición, resolución de la imagen, entre otros. Esto, junto otros factores como el propio algoritmo de reconstrucción empleado por los diferentes fabricantes o el campo magnético de la RM, hace que existan diferencias en los niveles de intensidad de la imagen. Por este motivo es necesario un proceso de normalización para llevar todas las imágenes al mismo rango de niveles de intensidad. Para cada uno de los problemas, se ha empleado un método de normalización diferente. El mejor método de normalización se definió de manera empírica tras un proceso de experimentación en el que se probaron diferentes opciones (normalización entre 0 y 1, estandarización serie a serie o considerando todo el conjunto de datos de entrenamiento, o normalización entre 0 y 1 seguido de estandarización).

- **Segmentación hepática en secuencias MECSE**

A cada una de las series (conjunto de cortes 2D de un mismo paciente) se le aplicó una normalización en dos pasos:

1. En primer lugar, todos los vóxeles se normalizaron entre 0 y 1 aplicando la Ecuación 25.

$$I'_{i,j,k} = \frac{I_{i,j,k} - \min(I)}{\max(I) - \min(I)} \quad (25)$$

Donde $I'_{i,j,k}$ es el valor normalizado de intensidad del vóxel (i,j,k); $I_{i,j,k}$ es el valor de intensidad del vóxel (i,j,k); $\min(I)$ es el valor de intensidad mínimo de la imagen I y $\max(I)$ es el valor de intensidad máximo de la imagen I.

2. A la imagen resultante se le aplicó una estandarización haciendo uso de los valores de media y desviación estándar de la intensidad de todo el conjunto de datos de entrenamiento de acuerdo con la Ecuación 26.

$$I''_{i,j,k} = \frac{I'_{i,j,k} - \mu(D)}{\sigma(D)} \quad (26)$$

Donde $I''_{i,j,k}$ es el valor normalizado de intensidad del vóxel (i,j,k); $I'_{i,j,k}$ es el valor de intensidad normalizado obtenido del paso anterior del vóxel (i,j,k); $\mu(D)$ es el valor de intensidad medio considerando todas las imágenes del conjunto de datos de entrenamiento y $\sigma(D)$ es la desviación estándar de las intensidades de todos los vóxeles del conjunto de datos de entrenamiento. Cabe resaltar que estos valores de media y desviación estándar son lo que se emplearon para normalizar tanto las imágenes de validación como las de test.

- **Segmentación hepática en secuencias de perfusión**

A cada una de las series se le aplicó una estandarización, es decir, a cada vóxel se le restó el valor de intensidad medio de la serie y se dividió entre la desviación estándar (Ecuación 27).

$$I'_{i,j,k} = \frac{I_{i,j,k} - \mu(S)}{\sigma(S)} \quad (27)$$

Donde, en este caso, S hace referencia a todas las imágenes que forman la serie.

- **Segmentación de grasa abdominal en imágenes de TC**

Se aplicaron métodos de normalización diferentes al modelo de clasificación de imágenes abdominales y al propio de segmentación de grasa visceral y subcutánea.

- Clasificación de imágenes abdominales: se realizó una estandarización considerando todo el conjunto de imágenes de entrenamiento.
- Segmentación de grasa visceral y subcutánea: en primer lugar, se le aplicó un enventanado a las imágenes entre [-150, 250] UH, siendo ésta la ventana empleada en abdomen que permite visualizar mejor los contrastes

entre la grasa y el resto de las vísceras [94]. A continuación, se aplicó la misma normalización que la empleada para la segmentación hepática en secuencias de perfusión.

3.3.2. Re-escalado

A la hora de diseñar una CNN, el tamaño de la imagen de entrada hay que fijarlo, por lo tanto, previo a la introducción de las imágenes a la CNN fue necesario redimensionarlas al tamaño establecido en cada caso. Para el modelo de segmentación automática hepática en secuencias MECSE se empleó un tamaño de entrada de 192x192, ya que gran parte de las imágenes empleadas tenían este tamaño o similar (Tabla 2). Sin embargo, tanto para la segmentación hepática en secuencias de RM de perfusión como de grasa en TC se empleó un tamaño de entrada de 256x256. Esta elección se tomó con el objetivo de alcanzar una relación de compromiso entre resolución de la imagen y carga computacional.

El redimensionado, además, se realizó de forma que se mantuvo un tamaño de voxel isotrópico en el plano (i.e., mismo tamaño de voxel en x e y) para no producir una deformación de la imagen. Por lo tanto, para alcanzar el tamaño objetivo de filas y columnas (n_{CNN}), se redimensionaron las imágenes de la siguiente forma.

1. El primer paso fue convertir la imagen para que tuviera un voxel isotrópico en el plano, en caso de que no lo tuviera. Para ello, se tomó como referencia el tamaño de la dirección que tuviera un tamaño de voxel más pequeño. Con este, se calculó el nuevo tamaño de filas y columnas correspondiente haciendo uso de las Ecuaciones 28-30.

$$\Delta v_{iso} = \min\{\Delta x, \Delta y\} \quad (28)$$

$$n'_{fil} = n_{fil} \cdot \frac{\Delta y}{\Delta v_{iso}} \quad (29)$$

$$n'_{col} = n_{col} \cdot \frac{\Delta x}{\Delta v_{iso}} \quad (30)$$

Donde Δv_{iso} hacer referencia al tamaño de voxel isotrópico en el plano al que se redimensionó la imagen; Δx y Δy son el tamaño original del voxel en el eje x e y, respectivamente; n_{fil} y n_{col} son, respectivamente, el número de filas y columnas

de la imagen original; y n'_{fil} y n'_{col} indican el nuevo número de filas y columnas, respectivamente, que adquirió la imagen.

2. A continuación, se redimensionó la imagen de forma que al menos uno de los ejes adquiriera el tamaño objetivo. De este modo, se detectó el eje de mayor tamaño y ambos ejes se redimensionaron con el objetivo que el mayor eje alcanzara el tamaño objetivo (Ecuaciones 31-33).

$$n_{ref} = \max\{n'_{fil}, n'_{col}\} \quad (31)$$

$$n''_{fil} = \left\lfloor \frac{n'_{fil} \cdot n_{CNN}}{n_{ref}} \right\rfloor \quad (32)$$

$$n''_{col} = \left\lfloor \frac{n'_{col} \cdot n_{CNN}}{n_{ref}} \right\rfloor \quad (33)$$

Donde n_{ref} es el tamaño del eje de mayor dimensión y que se empleó de referencia; y n''_{fil} y n''_{col} hacen referencia al nuevo número de filas y columnas que adquirió la imagen, respectivamente.

3. Una vez conocido el tamaño que debía adquirir la imagen, ésta se redimensionó empleando un algoritmo de interpolación bicúbica.
4. Finalmente, quedó ajustar el eje con menor dimensión para que adquiriera un tamaño n_{CNN} , para esto, se añadió un margen de ceros (valor de fondo) a ambos lados de la imagen, de forma equitativa, en el eje correspondiente hasta que se alcanzó el tamaño deseado.

Este redimensionado es de especial importancia en el modelo de segmentación hepática en imágenes de perfusión donde empleamos, como imagen de entrada a la CNN, la reconstrucción de las vistas coronales y sagitales. Estas reconstrucciones suelen resultar en imágenes no isotrópicas y con diferencia de resolución y variaciones en el número de filas y columnas.

3.4. Aprendizaje profundo

En esta sección se realiza una revisión de la metodología empleada para el desarrollo y validación de los distintos métodos de segmentación y clasificación basados en aprendizaje profundo.

Concretamente, se introducen las distintas técnicas de aumento de datos empleadas; se describen las arquitecturas diseñadas para la resolución de los distintos problemas a resolver; se detalla la metodología de selección de los hiperparámetros a definir como la función de pérdidas o la tasa de aprendizaje; así como la metodología de entrenamiento y validación interna de los modelos desarrollados; y, finalmente, se listan las métricas empleadas para evaluar la bondad de los distintos modelos entrenados, haciendo distinción en aquellas empleadas para evaluar un método de clasificación o segmentación.

3.4.1. Software y hardware

Para el desarrollo de los algoritmos se hizo uso del lenguaje de programación Python (versión 3.7).

Para el entrenamiento de los modelos se empleó un *cluster* con dos nodos de cálculo, cada nodo disponía de dos tarjetas gráficas (GPU) NVIDIA QUADRO GP100 ACTIVA con 16 GB de memoria RAM.

Para la fase de testeo e implementación de los algoritmos y métodos de cuantificación se empleó un ordenador con un procesador Intel Core i7 de 6 núcleos (2,6 GHz) y 16 GB de memoria RAM.

Todos las CNN se han implementado haciendo uso de la librería Tensorflow (versión 2.4).

3.4.2. Aumento de datos

Para incrementar la capacidad de generalización y robustez de los modelos generados frente a imágenes nuevas, el uso de técnicas de aumento de datos durante el proceso de entrenamiento es de gran interés. De este modo, diferentes transformaciones se pueden aplicar de forma aleatoria en cada iteración del entrenamiento.

En la presente tesis doctoral se ha empleado un método de aumento de datos al vuelo, es decir, en cada iteración se aplican las transformaciones oportunas sobre el lote de datos correspondiente con el que se van a ajustar los pesos del modelo.

Cuando diseñamos las técnicas de aumento de datos que se aplicarán a las imágenes de entrenamiento es de gran importancia aplicar transformaciones que tengan cierto sentido en el contexto en el que nos encontramos. Por ejemplo, una técnica de aumento de datos muy común es la generación de la imagen especular, sin embargo, en el cuerpo humano, no encontramos simetría en muchas regiones (como es el caso del abdomen) y, por tanto, debe evaluarse cuidadosamente cuando aplicar esta transformación.

Las transformaciones que se han empleado en la presente tesis doctoral son las siguientes:

- **Ruido gaussiano**

Es común encontrar ruido en las imágenes adquiridas, por lo tanto, es de gran utilidad entrenar una red para que sea capaz de ofrecer resultados adecuados independientemente de la presencia de ruido. De este modo, de forma aleatoria, se añadió ruido gaussiano a las imágenes de entrenamiento con $\mu = 0$; $\sigma = [0,01 - 0,1]$.

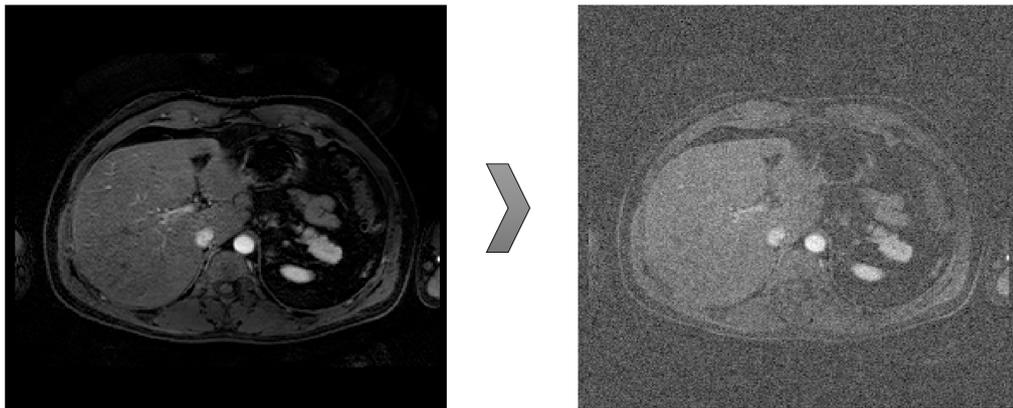


Figura 38. Aumento de datos: añadido de ruido gaussiano.

- **Rotaciones**

Las imágenes se pueden encontrar ligeramente rotadas debido, por ejemplo, al movimiento del paciente. Sin embargo, es raro encontrar grandes rotaciones. De este modo, pequeñas rotaciones (entre -10° y $+10^\circ$) se aplicaron a las imágenes de entrenamiento de forma aleatoria.

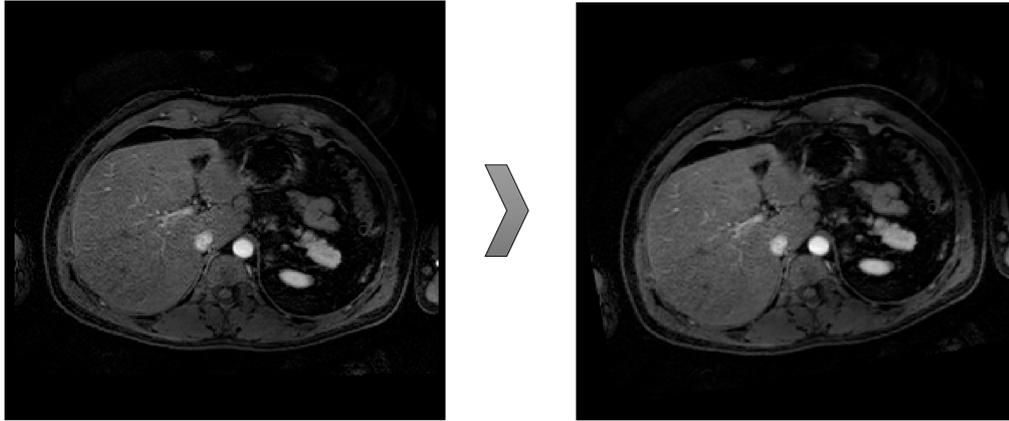


Figura 39. Aumento de datos. Rotación.

- **Traslaciones**

Del mismo modo que ocurre con las rotaciones debido a, por ejemplo, el movimiento del paciente, la imagen puede adquirirse ligeramente desplazada en lugar de encontrarse centrada. Por esto, es de gran utilidad encontrar este tipo de alteraciones en el conjunto de datos de entrenamiento. Así, se aplicaron pequeñas traslaciones a las imágenes (entre -50 y 50 píxeles en x e y) de forma aleatoria.

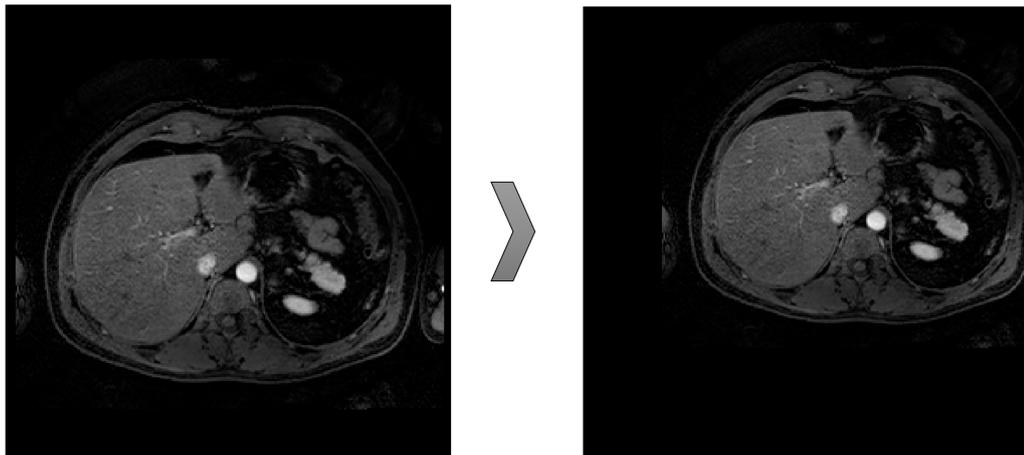


Figura 40. Aumento de datos: Traslaciones.

- **Zoom**

Dependiendo del protocolo de adquisición empleado, la cobertura anatómica en el plano puede variar. De este modo, las imágenes se pueden adquirir centrándose en una región de interés en concreto o cubriendo un área más grande. Por ejemplo, en estudios de TC, podemos encontrarnos estudios abdominales en los que la imagen se encuentra centrada en el abdomen o estudios de cuerpo completo en los que, junto al abdomen, encontramos, también, los brazos del paciente, viéndose

en este segundo caso, el abdomen más alejado. Por esta razón, es de gran interés incorporar zooms (hacia dentro y hacia fuera) a las imágenes de entrenamiento para mejorar el rendimiento de la red cuando tratamos con imágenes con coberturas anatómicas diferentes. Así, se aplicó zoom con una escala entre 0,5 y 2 de forma aleatoria.

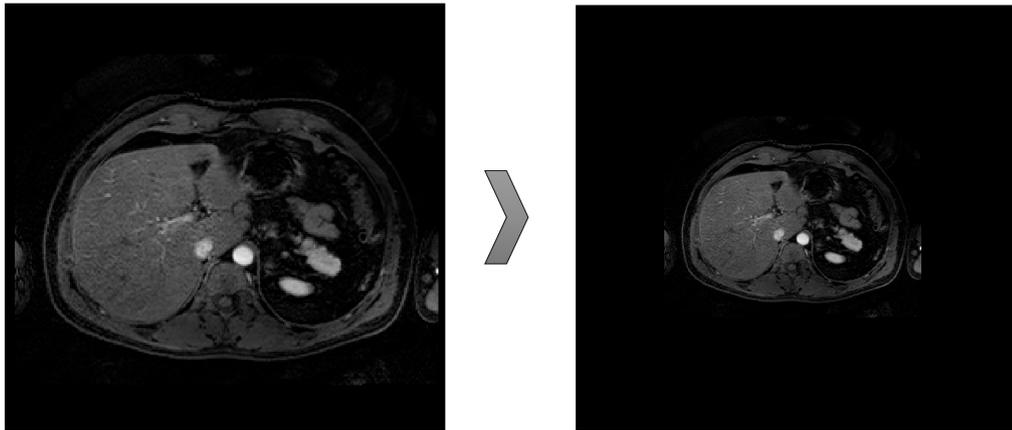


Figura 41. Aumento de datos: Zoom.

Cada una de estas transformaciones se aplicó con una probabilidad del 30%.

3.4.3. Selección de la arquitectura

Todos los modelos entrenados en la presente tesis doctoral emplean la misma arquitectura (Figura 42). Se trata de una arquitectura FCN, basada en la arquitectura U-Net [79]. A la configuración propuesta originalmente se le realizaron pequeñas modificaciones como es el uso de capas de normalización del lote [95] (BN, del inglés *Batch Normalization*) después de cada capa convolucional y la incorporación de un bloque de supervisión profunda en la fase de decodificación y generación de la máscara final. La capa de BN coge la salida del filtro convolucional y lo normaliza restando la media del lote y dividiéndolo entre la desviación estándar (normalización *z-score*), esto tiene un efecto de incremento de la generalización del modelo y reducción del sobreajuste. Las capas de supervisión profunda, como ya se introdujo previamente, se ha visto que ofrecen una convergencia más rápida y una mejor delineación de los bordes.

Para todas las arquitecturas se empleó un campo receptivo de 3x3 en las capas convolucionales y un relleno tal que la imagen de salida tenía el mismo tamaño que la imagen de entrada (*padding same*). La función de activación empleada fue la ReLU

(*Rectified Linear Unit*, Figura 17). En la fase de codificación de características se emplearon capas de agrupación máxima con un tamaño de 2×2 y avance de 2, de forma que el tamaño de la imagen se reducía a la mitad. Del mismo modo, en la fase de decodificación se emplearon capas de convolución traspuesta de tamaño 2×2 y avance de 2 para multiplicar el tamaño de la imagen al doble. Finalmente, el número de filtros convolucionales empleados en el primer bloque convolucional fue de 32 y se multiplicó por dos tras cada capa de agrupación durante la compresión de la imagen hasta un valor de 512 filtros, sin embargo, durante la descompresión de la imagen se siguió una estrategia inversa, tras cada capa de convolución traspuesta, el número de filtros convolucionales empleados se redujo a la mitad. Para el entrenamiento de esta arquitectura, para cada caso mencionado posteriormente, se ajustaron 7.765.860 parámetros.

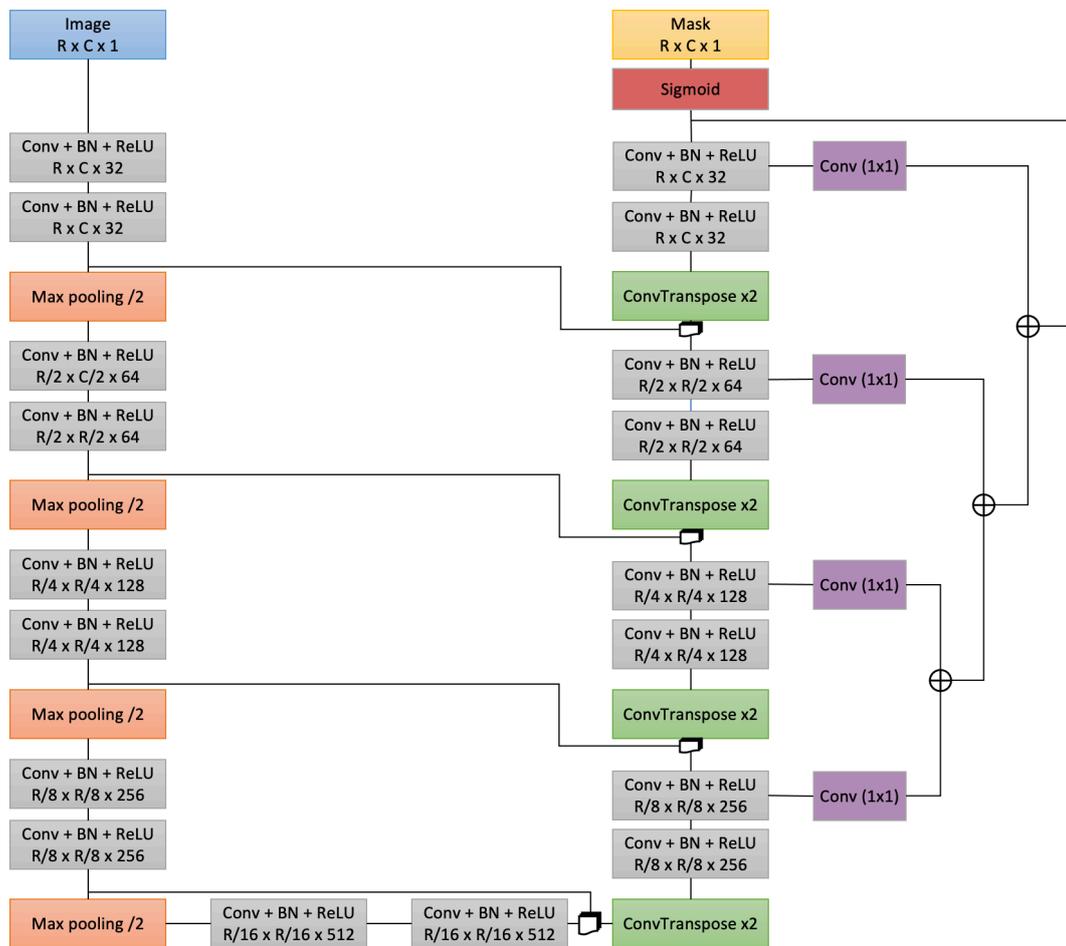


Figura 42. Arquitectura base empleada para el entrenamiento de todos los modelos de segmentación propuestos. *R*: número de filas; *C*: número de columnas; *Conv*: filtro convolucional; *BN*: capa de normalización de lote; *ReLU*: función de activación *Rectified Linear Unit*; *Max pooling*: capa de agrupación máxima; *ConvTranspose*: capa de convolución traspuesta; *Sigmoid*: función de activación sigmoide.

Capítulo 3. Materiales y métodos

3.4.3.1. Segmentación automática hepática en secuencias MECSE

Para la segmentación hepática en secuencias MECSE se hizo uso de la arquitectura presentada en la Figura 42, con un tamaño de imagen de entrada de 192x192x1.

3.4.3.2. Segmentación automática de grasa abdominal en imágenes de TC

Para la segmentación automática de grasa abdominal, como se ha comentado anteriormente, se ha desarrollado un modelo previo para la selección de la región abdominal, es decir, se han clasificado todas las imágenes transversales que componen el TC como pertenecientes a la región abdominal (sí/no). Aquellas pertenecientes a la región abdominal son las que posteriormente se segmentaron en grasa visceral y subcutánea. De este modo, se han diseñados dos arquitecturas diferentes, una para la clasificación y otra para la segmentación.

3.4.3.2.1. Clasificación de imágenes de TC en región abdominal o no abdominal

Para la resolución de este problema se ha diseñado una arquitectura formada por cinco bloques convolucionales y una capa densa de salida con dos neuronas. Cada bloque convolucional está compuesto por un filtro convolucional con un campo receptivo de 3x3, un relleno tal que la imagen de salida tenía el mismo tamaño que la imagen de entrada (*padding same*) y una función de activación ReLU. Tras el filtro se incluyó una capa de agrupación máxima con un tamaño de 2x2 y avance de 2, de forma que la imagen quedaba reducida a la mitad. El número de filtros empleados en cada filtro convolucional se duplicaba cada vez que la imagen se reducía a la mitad empezando con 32 filtros y terminando con 512. Tras los cinco bloques convolucionales se añade una capa de aplanamiento (*Flatten*, de su nombre en inglés) que convierte la salida de la capa anterior en un vector que acaba, finalmente, en una capa densa o totalmente conectada de dos neuronas, correspondientes al número de salidas deseadas. Para definir la clasificación final se emplea una función de activación *Softmax*.

Para este caso, se empleó un tamaño de imagen de entrada de 256x256x1. Cabe destacar que para la definición de la clase objetivo se definió la salida en codificación *One Hot* donde [1, 0] pertenecía a la clase abdomen y [0, 1] a la clase no abdomen.

En la Figura 43 encontramos el esquemático de la arquitectura empleada que consta de un total de 1.633.538 parámetros a entrenar.

Capítulo 3. Materiales y métodos

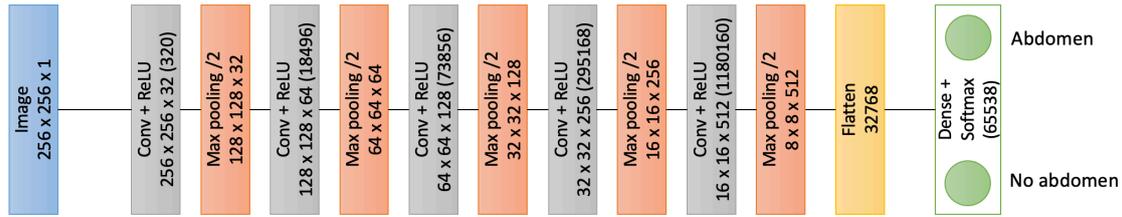


Figura 43. Arquitectura empleada para la clasificación de las imágenes de TC de una serie en abdomen y no abdomen con el objetivo de definir la región a segmentar. Para cada capa se muestra el tamaño de la imagen de salida, así como el número de parámetros a ajustar en dicha capa entre paréntesis. Se puede ver cómo tanto en las capas de agrupación como en las de aplanamiento no hay parámetros a ajustar. Conv: Filtro convolucional; ReLU: Función de activación Rectified Linear Unit; Max pooling: Capa de agrupación máxima; Flatten: Capa que convierte una matriz en vector.

3.4.3.2.2. Segmentación de grasa visceral y subcutánea en imágenes de TC

Para la segmentación de la grasa visceral y subcutánea en imágenes de TC se ha empleado la misma arquitectura que la mostrada en la Figura 42 con algunas pequeñas diferencias. Estas diferencias surgen por el hecho de tener que segmentar dos regiones en lugar de una única. De este modo, para realizar el entrenamiento se convirtió la máscara en una imagen con tres canales, de forma que el canal cero correspondía al fondo y teníamos a 1 los píxeles correspondientes al fondo (todo lo que no es grasa visceral ni subcutánea) y a 0 el resto; el canal uno correspondía a la grasa visceral y teníamos, por tanto, los píxeles correspondientes a la grasa visceral a 1 y el resto a 0; y el canal dos correspondía a la grasa subcutánea donde teníamos a 1 los píxeles correspondientes a la grasa subcutánea y a 0 el resto. Así, a la salida obteníamos un tamaño de máscara, de $R \times C \times 3$, siendo la función de activación que se empleó en este caso, la función *Softmax* al tratarse de un problema multiclase.

A la entrada, al igual que para la clasificación, se definió un tamaño de imagen de $256 \times 256 \times 1$.

3.4.3.3. Segmentación automática hepática en secuencias de perfusión

La secuencia de perfusión ofrece buena resolución de imagen entre los cortes (Tabla 5), esto hace que podemos reconstruir la imagen en otras orientaciones manteniendo, así, buena resolución de imagen. Las imágenes médicas tomográficas se pueden orientar en tres direcciones, ortogonales entre sí y que se conocen como:

- Dirección transversal: plano que corta al paciente de delante hacia atrás.
- Dirección coronal: plano que corta al paciente de izquierda a derecha.
- Dirección sagital: plano que corta al paciente de arriba hacia abajo.

Este proceso de reorientar la imagen adquirida en otras orientaciones se conoce como reconstrucción multiplanar (MPR, del inglés *Multi-Planar Reconstruction*).

De este modo, se entrenaron tres modelos diferentes, todos ellos basados en la arquitectura de la Figura 42, con la única diferencia de la imagen empleada a la entrada de cada una de ellas. Así, como se puede observar en la Figura 44, se entrenó un modelo con imágenes en la dirección transversal, otro con imágenes en la dirección coronal y un último con imágenes en la dirección sagital.

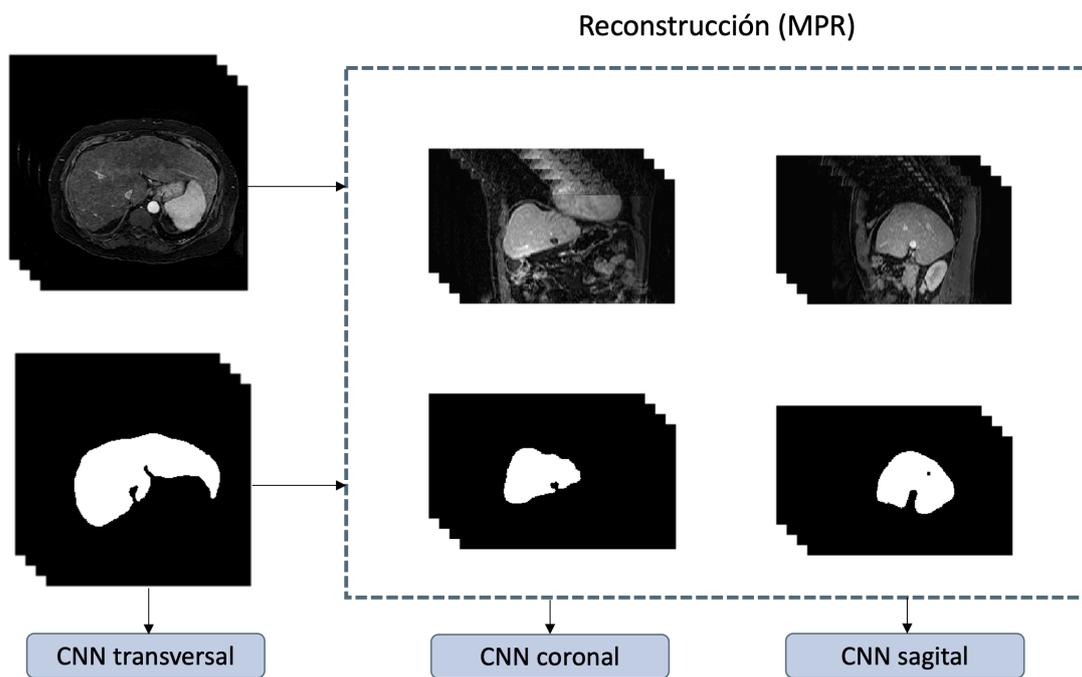


Figura 44. Reconstrucción de imagen para el entrenamiento de tres modelos diferentes de segmentación hepática en las tres orientaciones ortogonales (transversal, coronal y sagital).

En este caso, para los tres modelos, se empleó un tamaño de imagen de entrada de 256x256x1.

3.4.4. Entrenamiento y validación

Una vez teníamos tanto las imágenes que íbamos a emplear para el ajuste de la red como las correspondientes anotaciones listas, se procedió a realizar el entrenamiento de los modelos.

Este entrenamiento se desarrolló siguiendo una estrategia de ajuste de hiperparámetros en el que diferentes configuraciones se entrenaron y se seleccionó aquella que mejores resultados ofrecía en el proceso de validación cruzada.

La Figura 45 muestra el proceso seguido para el entrenamiento y validación interna de los modelos finales. Este proceso se puede dividir en las siguientes etapas:

1. El conjunto de datos de entrenamiento y validación se subdividió en cinco subconjuntos que se emplearon para realizar la validación cruzada de los modelos entrenados. De este modo, en un proceso iterativo de cinco iteraciones, en cada iteración, se emplearon cuatro subconjuntos para el entrenamiento de los modelos y el restante para su validación interna.
2. En cada iteración se entrenaron todas las configuraciones definidas en la sección 3.4.4.1. *Selección de los hiperparámetros* para seleccionar tanto la tasa de aprendizaje como la función de pérdidas óptimas.

Para resolver el problema de clasificación de imágenes de TC entre abdomen y no abdomen se ajustó únicamente la tasa de aprendizaje siguiendo el mismo procedimiento, esto es debido a que la función de pérdidas, al tratarse de un problema de clasificación, se empleó la entropía cruzada.

3. Para definir los hiperparámetros que se emplearon en el modelo final, se seleccionaron aquellos que ofrecían un mayor coeficiente Dice (DC) medio a lo largo de las 5 iteraciones y mayor estabilidad, es decir, que la diferencia de los DC a lo largo de las iteraciones fuera baja. Para ello se estableció un umbral de diferencia máximo de DC del 3%.

Para el clasificador de imágenes de TC se empleó la exactitud como métrica de referencia para seleccionar el mejor modelo.

4. Una vez seleccionada la mejor configuración, finalmente se entrenaron los modelos finales. Se decidió entrenar un ensamblado de cinco CNN, cada una de ellas entrenada con 4 particiones diferentes del conjunto de datos de entrenamiento y validación.

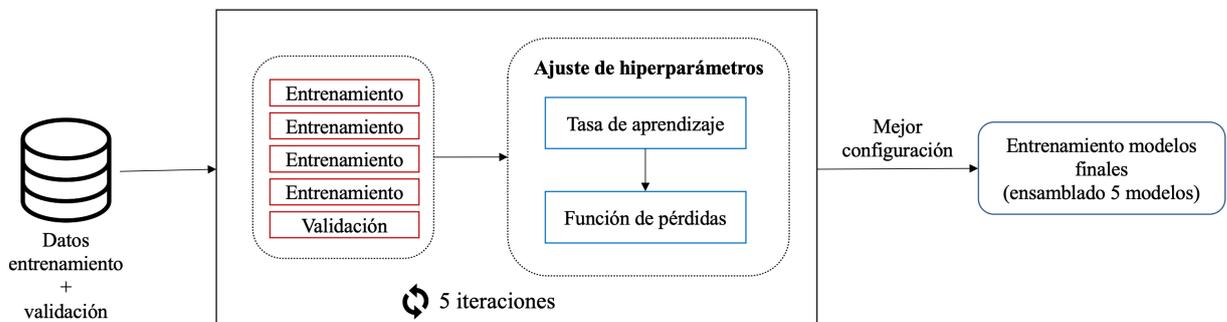


Figura 45. Proceso de entrenamiento y validación cruzada para la selección de la mejor configuración.

3.4.4.1. Selección de los hiperparámetros

Como se ha comentado anteriormente, se realizó una selección de los hiperparámetros que optimizaran el rendimiento del modelo en un proceso de validación cruzada. Entre los hiperparámetros que se ajustaron fueron el tamaño de lote, el número de épocas, la tasa de aprendizaje y la función de pérdidas.

3.4.4.1.1. *Tamaño de lote y número de épocas*

El tamaño de lote y el número de épocas son dos de los hiperparámetros a definir previo al inicio del entrenamiento de los modelos.

- Tamaño de lote: hace referencia al número de imágenes que emplea el modelo de una (en una iteración) para ajustar los pesos de la red. Generalmente, el tamaño de lote va a ser inferior al número de imágenes de entrenamiento lo que hará, que el número de iteraciones y de épocas sea diferente. En muchos casos, especialmente trabajando con imágenes, el tamaño de lote va a depender de la memoria RAM que tenga la GPU sobre la que se produce el entrenamiento. En nuestro caso, se ha empleado siempre un tamaño de lote de 40 imágenes.
- Número de épocas: número de pasos completos sobre el conjunto de datos de entrenamiento. Este hiperparámetro debe de ajustarse debidamente para evitar tanto el subajuste (falta de entrenamiento) como el sobreajuste (exceso de entrenamiento). Durante el proceso de selección de los hiperparámetros se empleó un número de épocas fijo de 70. Sin embargo, para el entrenamiento de los modelos finales, se realizó una selección dinámica del número de épocas, de forma que el entrenamiento se detuvo, en cada caso, cuando no había una mejora en las métricas de validación durante más de 5 épocas seguidas.

3.4.4.1.2. *Tasa de aprendizaje*

Durante el proceso de entrenamiento podemos emplear una tasa de aprendizaje constante o variable. Generalmente, emplear una tasa de aprendizaje variable ofrece mejores resultados ya que permite adaptar la magnitud de esta a cada etapa del aprendizaje. De este modo, conviene emplear tasas de aprendizaje más elevadas al inicio del entrenamiento e ir reduciéndola conforme el entrenamiento avanza para un ajuste más fino de los pesos de la red.

Para determinar cuál es la tasa de aprendizaje óptima para cada problema se realizó una fase de experimentación donde se evaluó el rendimiento de distintas aproximaciones. Para cada una de ellas se empleó el coeficiente Dice como función de pérdidas (3.4.4.1.3. *Función de pérdidas*).

1. Tasa de aprendizaje constante: en primer lugar, se ha evaluado el rendimiento de los modelos entrenados empleando una tasa de aprendizaje constante. Para ello, se desarrolló una rutina para evaluar el rendimiento de la red empleando ocho valores diferentes de tasa de aprendizaje: 1, 10^{-1} , 10^{-2} , 10^{-3} , 10^{-4} , 10^{-5} , 10^{-6} , 10^{-7} .
2. Tasa de aprendizaje variable: a continuación, se analizó el rendimiento de los modelos entrenados empleando una tasa de aprendizaje variable durante el proceso de entrenamiento. El cambio de este hiperparámetro se puede realizar de dos formas diferentes, a través de planificadores o empleando diferentes optimizadores.
 - a. Planificadores: varían la tasa de aprendizaje de acuerdo con una función predefinida. Se establece una tasa de aprendizaje inicial y, en función del número de iteraciones de entrenamiento, se va variando la misma en cada iteración. De este modo, se desarrolló una rutina de experimentación que evalúa el rendimiento de distintos planificadores empleando distintas tasas de aprendizaje iniciales. Por lo tanto, se realizaron distintos entrenamientos con distintos planificadores que empleaban los siguientes valores de tasa de aprendizaje inicial: 1, 10^{-1} y 10^{-2} .

Los planificadores que se evaluaron son los siguientes:

- i. Basado en el tiempo:

La tasa de aprendizaje en una iteración dada viene definida por la Ecuación 34:

$$TA = \frac{TA_{prev}}{1 + declive \cdot \acute{e}poca} \quad (34)$$

Donde TA_{prev} hace referencia a la tasa de aprendizaje previa, $\acute{e}poca$ es la iteración del proceso de aprendizaje y $declive$ es un hiperparámetro que viene definido por la Ecuación 35.

Capítulo 3. Materiales y métodos

$$declive = \frac{TA_{ini}}{N^{\circ} \text{ épocas}} \quad (35)$$

Donde TA_{ini} es la tasa de aprendizaje inicial establecida.

ii. Lineal

La tasa de aprendizaje, en cada iteración del proceso de entrenamiento, disminuye de manera lineal de acuerdo con la Ecuación 36.

$$TA = TA_{ini} \cdot \left(1 - \frac{\text{época}}{N^{\circ} \text{ épocas}}\right) \quad (36)$$

La Figura 46 muestra la variación de la tasa de aprendizaje durante el proceso de entrenamiento dada una tasa de aprendizaje inicial de 0,1 y un número total de épocas de 50.

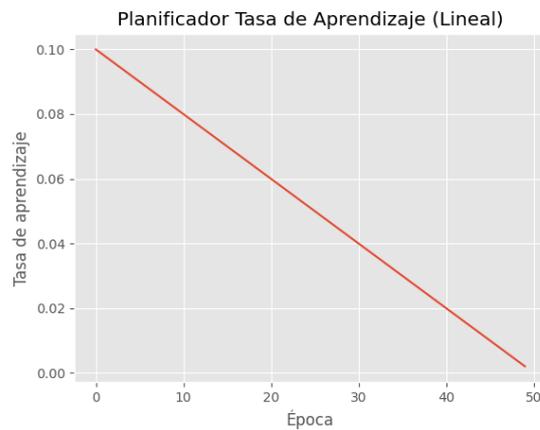


Figura 46. Planificador de la tasa de aprendizaje lineal (Tasa de aprendizaje inicial: 0,1; número de épocas: 50).

iii. Polinomial

Similar al planificador anterior, pero en este caso, la rapidez con que la tasa de aprendizaje cae viene definido por el factor del polinomio. De este modo, la tasa de aprendizaje durante el entrenamiento viene definida por la Ecuación 37.

Capítulo 3. Materiales y métodos

$$TA = TA_{ini} \cdot \left(1 - \frac{\text{época}}{N^{\circ} \text{ épocas}}\right)^{\text{factor}} \quad (37)$$

Para este planificador, además, se evaluaron cuatro factores diferentes: 2, 3, 5 y 7.

La Figura 47 muestra la variación de la tasa de aprendizaje dada una tasa inicial de 0,1, un número total de épocas de 50 y un factor de 3.

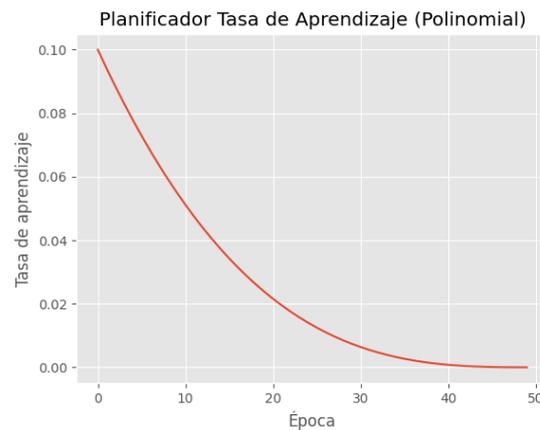


Figura 47. Planificador de la tasa de aprendizaje polinomial (Tasa de aprendizaje inicial: 0,1; número de épocas: 50; factor: 3).

iv. Exponencial

Este planificador permite reducir la tasa de aprendizaje de manera exponencial de acuerdo con la Ecuación 38.

$$TA = TA_{ini} \cdot e^{-k \cdot \text{época}} \quad (38)$$

Donde k es el factor de la exponencial y se le ha dado un valor constante de 0,1.

La Figura 48 representa la tasa de aprendizaje durante el proceso de entrenamiento al emplear un planificador exponencial donde la tasa de aprendizaje inicial es 0,1.

Capítulo 3. Materiales y métodos

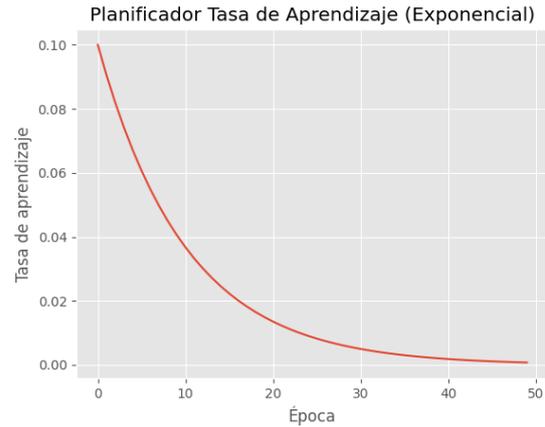


Figura 48. Planificador de la tasa de aprendizaje exponencial (Tasa de aprendizaje inicial: 0,1; número de épocas: 50, k : 0,1).

v. Por saltos

Tras un número concreto de épocas se reduce el valor de la tasa de aprendizaje. La tasa de aprendizaje viene dada por la Ecuación 39.

$$TA = TA_{ini} \cdot tasa^{[(1+época)/\Delta épocas]} \quad (39)$$

Donde *tasa* define cuánto cambia la tasa de aprendizaje cada vez y $\Delta épocas$ hace referencia a cada cuántas épocas se produce este cambio.

Para este planificador, además se experimentaron las siguientes configuraciones: caídas cada 5, 10 y 25 épocas con un factor de disminución de 0,25 y 0,5.

La Figura 49 representa la tasa de aprendizaje durante el proceso de entrenamiento dada una tasa inicial de 0,1 y una tasa de disminución de 0,5 cada 10 épocas.

Capítulo 3. Materiales y métodos

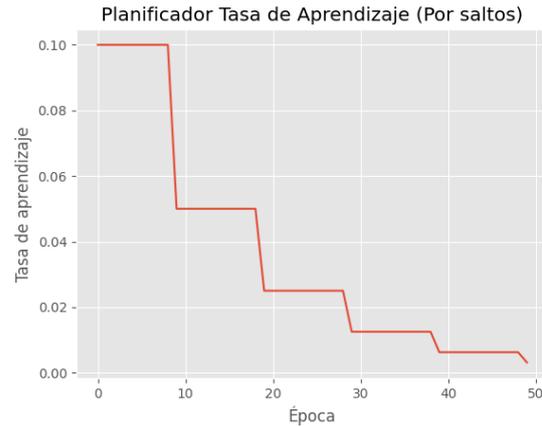


Figura 49. Planificador de la tasa de aprendizaje por saltos (Tasa de aprendizaje inicial: 0,1; número de épocas: 50; tasa de disminución: 0,5 cada 10 épocas).

vi. Constante y exponencial

Se mantiene una tasa de aprendizaje constante durante las primeras n épocas del proceso de entrenamiento y, a continuación, se produce una caída exponencial de la misma.

Se estableció un valor de 10 en el número de épocas con tasa de aprendizaje constante y un factor de 0,1 en la caída exponencial.

La Figura 50 representa las variaciones de la tasa de aprendizaje durante el proceso de entrenamiento siendo la tasa inicial de 0,1, el número de épocas con tasa de aprendizaje constante de 10 y un factor exponencial de 0,1.



Figura 50. Tasa de aprendizaje constante y exponencial (Tasa de aprendizaje inicial: 0,1; número de épocas: 50; número de épocas con tasa de aprendizaje constante: 10; k : 0,1).

Capítulo 3. Materiales y métodos

vii. De un ciclo

Se produce un aumento gradual de la tasa de aprendizaje desde un valor inicial hasta un valor máximo (0,1) durante un número determinado de épocas (10). Tras esta subida, se mantiene constante un número determinado de épocas (10) y, finalmente, disminuye de manera exponencial ($k=0,8$) hasta un valor mínimo establecido (10^{-8}). La Figura 51 muestra la tasa de aprendizaje cíclica empleada.

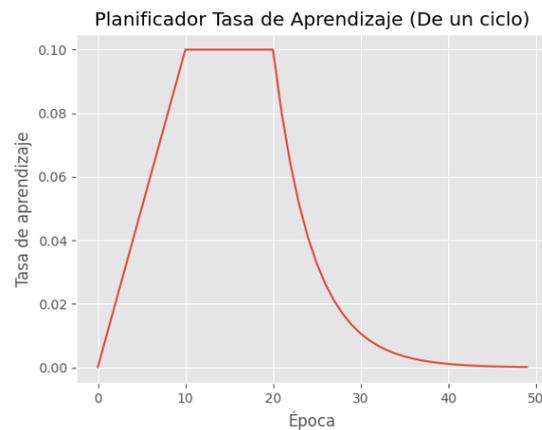


Figura 51. Tasa de aprendizaje de un ciclo con una tasa de aprendizaje inicial de 10^{-4} , durante 10 iteraciones se produce un aumento lineal hasta una tasa máxima de 10^{-2} que se mantiene constante durante 10 iteraciones y disminuye exponencialmente con $k=0,8$ hasta un valor mínimo de 10^{-5} .

viii. Dinámico

En este caso la tasa de aprendizaje va disminuyendo de manera dinámica durante el proceso de entrenamiento. De este modo, cuando las pérdidas en validación se mantienen constantes durante un número determinado de épocas, se disminuye la tasa de aprendizaje un determinado factor.

Se evaluaron diferentes números de épocas tras las que disminuir la tasa de aprendizaje: 5, 10, 15.

El factor de disminución se dejó constante en un valor de 0,1.

b. Optimizadores:

Todo lo visto anteriormente emplea el algoritmo de gradiente de descenso estocástico (SGD, del inglés, *Stochastic Gradient Descent*) para la

actualización de los pesos de la red durante el proceso de entrenamiento.

Sin embargo, como hemos podido ver, hay una serie de retos:

- La selección de la tasa de aprendizaje adecuada puede resultar complicado y requiere de experimentación.
- La selección del planificador adecuado para cada problema puede resultar tedioso. Además, se añaden nuevos hiperparámetros (e.g., factor de la caída en el planificador exponencial).
- La misma tasa de aprendizaje se aplica para la actualización de todos los pesos. Si nuestros datos son dispersos y nuestras características tienen diferentes frecuencias, puede que no deseemos actualizar todos los pesos de la misma forma.
- Debemos evitar quedarnos “atrapados” en un mínimo local.

Para hacer frente a estos retos surgen una serie de algoritmos de optimización, entre los que encontramos:

- Adagrad [96]: Este algoritmo adapta la tasa de aprendizaje a los pesos, actualizando en menor medida (tasas de aprendizaje más bajas) los pesos asociados a características que ocurren con mayor frecuencia, y en mayor medida (tasas de aprendizaje más elevadas) pesos asociados a características poco frecuentes. Una de las principales ventajas de Adagrad es que elimina la necesidad de seleccionar la tasa de aprendizaje de forma manual, sin embargo, cuenta con el inconveniente de que la tasa de aprendizaje cae rápido a valores infinitésimamente pequeños, en cuyo momento la red no es capaz de seguir aprendiendo.

Este optimizador se inicializó con los siguientes parámetros:

$$TA_{ini} = 0,01; \varepsilon = 10^{-8}.$$

- Adadelta [97]: Este algoritmo busca eliminar el principal problema de Adagrad en el que la tasa de aprendizaje cae a valores muy bajos.

Se emplearon los siguientes parámetros: $TA_{ini} = 1; \rho = 0,95; \varepsilon = 10^{-8}$

- RMSprop [98]: Al igual que Adadelta, presentadas en fechas similares, pretende eliminar el problema de Adagrad.

Los siguientes parámetros se emplearon al hacer uso de este optimizador: $TA_{ini} = 0,001$; $\gamma = 0,9$; $\varepsilon = 10^{-8}$, recomendados por el autor del algoritmo.

- Adam [99]: *Adaptive Moment Estimation* (Adam), se presenta como actualización de Adadelta y RMSprop. Al igual que todas ellas, busca calcular una tasa de aprendizaje diferente para cada uno de los pesos de la red.

Se emplearon los siguientes parámetros: $TA_{ini} = 0,001$; $\beta_1 = 0,9$; $\beta_2 = 0,999$; $\varepsilon = 10^{-8}$. Parámetros recomendados por los propios autores del algoritmo a usar por defecto.

3.4.4.1.3. Función de pérdidas

La función de pérdidas, también conocida como función de coste, es la función que se desea minimizar durante el proceso de entrenamiento y que, de este modo, determina el error que tiene la red en cada iteración.

Dependiendo del problema que se desea resolver la función de pérdidas óptima a emplear variará. Así, en un problema de clasificación, la función de pérdidas más empleada es la entropía cruzada, mientras que, en un problema de segmentación, son diversas las funciones de pérdidas que se pueden emplear, entre ellas destacamos las siguientes:

1. Entropía cruzada binaria

La entropía cruzada binaria (BCE, del inglés, *Binary Cross Entropy*) compara la probabilidad predicha con la clase real y penaliza en función de la distancia entre ambos valores. Se basa en la Ecuación 40.

$$BCE = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)) \quad (40)$$

Donde N es el número de vóxeles de la imagen, y_i es la etiqueta (i.e. 1 para los vóxeles que definen la región segmentada y 0 para el fondo) y $p(y_i)$ es la probabilidad predicha de que un vóxel i pertenezca a la etiqueta dada.

2. Pérdidas Dice

Las pérdidas Dice (DL, del inglés *Dice Loss*) se calculan como el complementario del coeficiente Dice (DC, del inglés *Dice Coefficient*) [100, 101] o *F1 score*. El coeficiente Dice mide el grado de similitud o solapamiento entre dos objetos y se calcula como la relación entre el número de vóxeles de solapamiento entre los dos objetos y el número total de vóxeles de ambos (Ecuación 41).

$$DL = 1 - DC = 1 - \frac{2 \cdot |P \cap R|}{|P| + |R|} = 1 - \frac{2 \cdot VP}{2 \cdot VP + FP + FN} \quad (41)$$

Siendo P la máscara predicha y R la máscara de referencia. VP (Verdaderos Positivos) hace referencia a los píxeles segmentados como región de interés correctamente; FN (Falsos Negativos) hace referencia a los píxeles establecidos como fondo que realmente corresponden a región de interés; FP (Falsos Positivos), hace referencia a los vóxeles segmentados como región de interés que en realidad pertenecen al fondo.

3. Entropía cruzada binaria y pérdidas Dice

Esta función de pérdidas permite combinar el resultado de las dos funciones de pérdidas anteriores (Ecuación 42).

$$BCE_DL = BCE + DL \quad (42)$$

4. Pérdidas intersección sobre unión

Las pérdidas intersección sobre unión o distancia Jaccard se define como el complementario del coeficiente de la intersección sobre la unión (IoU, del inglés, *Intersection Over Union*), también conocido como índice Jaccard. La IoU mide la similitud entre dos objetos y se define como el tamaño de la intersección dividido por el tamaño de la unión entre ambos objetos (Ecuación 43).

$$IoU = 1 - \frac{|P \cap R|}{|P \cup R|} = 1 - \frac{|P \cap R|}{|P| + |R| - |P \cap R|} \quad (43)$$

5. Pérdidas Tversky

La función de pérdidas Tversky [102] (TL, del inglés *Tversky Loss*) aparece para mejorar el rendimiento de los modelos de segmentación en la resolución de problemas en los que los datos están desbalanceados, como ocurre, generalmente, en imagen médica. Este efecto ocurre cuando el número de vóxeles pertenecientes a la región de interés es significativamente inferior a aquellos pertenecientes a fondo de imagen. De este modo, la función de pérdidas Tversky emplea un parámetro $\alpha \in [0, 1]$ que, en función de su valor, penaliza los falsos positivos o falsos negativos. Así, $\alpha < 0,5$ penalizará en mayor medida los falsos negativos, mientras que $\alpha > 0,5$ penalizará los falsos positivos. $\alpha = 0,5$ pondera por igual los falsos negativos y los falsos positivos, y equivaldría al coeficiente Dice. De este modo la función de pérdidas Tversky, complementario del índice Tversky [103], se rige por la Ecuación 44.

$$TL = 1 - TI = 1 - \frac{|P \cap R|}{|P \cap R| + \alpha \cdot |P \setminus R| + (1 - \alpha) \cdot |R \setminus P|} \quad (44)$$

Para evaluar el efecto de α en cada uno de los modelos se evaluaron diferentes valores de esta: 0,1, 0,3, 0,7 y 0,9.

6. Pérdidas *Focal Tversky*

La función de pérdidas Focal Tversky [104] (FTL, del inglés *Focal Tversky Loss*) surge como ampliación de la función de pérdidas Tversky bajo la hipótesis de que regiones de interés pequeñas no contribuyen de forma significativa a la función de pérdidas. Este problema proponen controlarlo a través del parámetro $\gamma \in [1, 3]$ para focalizar el entrenamiento en aquellas regiones pequeñas detectadas con menor probabilidad. En la práctica, si un vóxel se segmenta erróneamente con un índice de Tversky elevado, la FTL no se verá alterada, sin embargo, si un vóxel se segmenta erróneamente con un índice de Tversky bajo, la FTL disminuirá significativamente.

En [104], tras realizar diferentes experimentos, encuentran que el mejor resultado lo obtienen con $\gamma = 4/3$ y, por tanto, es el que se ha empleado en la presente tesis doctoral en los experimentos correspondientes.

La función de pérdidas Focal Tversky viene definida por la Ecuación 45.

$$FTL = (1 - TI)^{1/\gamma} \quad (45)$$

Para evaluar el efecto de usar esta función de pérdidas, para cada modelo desarrollado, se evaluó el efecto de incluir el parámetro γ para cada α definida en el apartado anterior.

7. Pérdidas Combo

La función de pérdidas Combo [105] (CL, del inglés, *Combo Loss*) se presenta para solventar el problema del desbalanceo entre clases en problemas de segmentación multiclase (e.g., segmentación de diferentes órganos). Esta función de pérdidas es una combinación de las pérdidas Dice y una entropía cruzada modificada (BCE_m – Ecuación 46) de acuerdo con la Ecuación 47. Así, incorpora diferentes parámetros para penalizar los falsos positivos o los falsos negativos, en función de las necesidades de cada problema.

$$BCE_m = -\frac{1}{N} \sum_{i=1}^N \beta \cdot y_i \cdot \log(p(y_i)) + (1 - \beta) \cdot (1 - y_i) \cdot \log(1 - p(y_i)) \quad (46)$$

$$CL = \alpha \cdot BCE_m - (1 - \alpha) \cdot DC \quad (47)$$

Donde α controla la contribución de cada componente (BCE_m y DC) a la función de pérdidas y $\beta \in [0, 1]$ controla la penalización de falsos positivos o negativos. Así, para $\beta < 0,5$ se penalizan en mayor medida los falsos positivos, mientras que para $\beta > 0,5$ se penalizan mayoritariamente los falsos negativos.

Para un ajuste óptimo de α y β cada uno de los modelos desarrollados se evaluaron con diferentes valores de estos: α : 0,3, 0,5 y 0,7; β : 0,1, 0,3, 0,5, 0,7 y 0,9.

Para cada uno de los modelos de segmentación desarrollados en la presente tesis doctoral se evaluó el funcionamiento de cada una de las funciones de pérdidas presentadas para seleccionar aquella que proporcione mejores resultados en la resolución de cada uno de los problemas a solventar.

Cabe resaltar que para el entrenamiento de los modelos de clasificación de imágenes abdominales se empleó la entropía cruzada como función de pérdidas. Además, para el entrenamiento de los modelos de segmentación visceral y subcutánea, al contar con tres clases (grasa visceral, grasa subcutánea y fondo) en lugar de dos, se calcularon las funciones de pérdidas para cada clase y se computó la media como valor de pérdidas final.

3.4.5. Test

Como hemos visto anteriormente, tras el entrenamiento y validación de cada uno de los modelos, con cada una de las configuraciones mencionadas en apartados anteriores, y la selección de aquel que mejores resultados ofrecía, se realizó un testeo adicional para garantizar la generalización y reproducibilidad de dicho modelo ante casos completamente nuevos.

3.4.5.1. Procesamiento de estudios de test

A continuación, se define el procesamiento aplicado a cada uno de los estudios de test, en función del objetivo, y las métricas que se emplearon para analizar la bondad de las predicciones.

- Para la *segmentación hepática en secuencias MECSE*, con el modelo final entrenado se obtuvo un mapa de probabilidad donde, en cada vóxel, teníamos la probabilidad de que éste perteneciera a hígado a no. Por lo tanto, para obtener la máscara final se realizó un estudio del umbral que, manteniendo una buena precisión, minimizaba los falsos positivos con el objetivo de minimizar los sesgos introducidos en la caracterización del tejido si se incorporaban, en la región analizada, vóxeles que no pertenecían a hígado. Una vez obtenida la máscara final, ésta se postprocesó haciendo una búsqueda de todas las componentes conectadas independientes que se habían obtenido y seleccionando únicamente aquella que mayor volumen tenía, correspondiente a la región hepática. De este modo, se eliminaban pequeñas regiones de ruido que se pudieran haber incluido en la máscara anterior.
- Para la *segmentación hepática en secuencias de perfusión*, para cada una de las vistas, se entrenó un ensamblado de cinco CNN, empleando cada una de las particiones durante el proceso de validación cruzada. De este modo, para obtener la máscara final, en primer lugar, para cada orientación se generó el mapa de

probabilidad con las cinco CNN independientemente y, a continuación, se calculó el mapa de probabilidad medio haciendo uso de la media aritmética. Después, los tres mapas resultantes correspondientes a cada una de las orientaciones, tras su reconstrucción al plano transversal, se promediaron para obtener un único mapa de probabilidad. Finalmente, se empleó un umbral de 0,5 para generar la máscara binaria final. En este caso no se realizó un estudio para seleccionar el umbral óptimo ya que el principal objetivo de esta máscara es emplearse como base para la detección de lesiones hepáticas, por lo tanto, se consideró que el umbral de 0,5 establecía la relación de compromiso óptima para minimizar los falsos positivos sin eliminar un número significativo de verdaderos positivos que pudieran corresponder a lesión. Una vez definida la máscara binaria se realizó el mismo postproceso que en el caso de la segmentación hepática en secuencias MECSE.

- Para la *segmentación de grasa visceral y subcutánea en TC* se realizó una concatenación de dos modelos:
 1. En primer lugar, de toda la serie de TC, se debían de seleccionar los cortes pertenecientes a la región abdominal, para ello, al igual que en el anterior modelo, se entrenó un ensamblado de cinco CNN. Así, para cada imagen, se obtuvo la probabilidad de cada clase (abdomen o no abdomen) con cada uno de los cinco modelos y, tras realizar la media aritmética se seleccionó como predicción la clase que tuviera mayor probabilidad. Tras esto se realizó una fase de postproceso para definir la región abdominal:
 - Si se clasificaba un corte aislado como abdomen (corte anterior y siguiente clasificados como no abdomen), se le cambiaba la clase a no abdomen. Así, el primer corte que definía la región abdominal era el primero en el que, al menos, el inmediatamente siguiente también se clasificaba como abdomen.
 - Tras el corte identificado en el punto anterior, si se clasificaba un corte aislado como no abdomen (corte anterior y siguiente clasificados como abdomen), se le cambiaba la clase a abdomen. Así, el último corte que definía la región abdominal era el primero en el que, al menos, el inmediatamente siguiente también se clasificaba como no abdomen.
 - Si entre los dos cortes definidos anteriormente, que definían la región abdominal, seguía habiendo algún corte clasificado como no abdomen, se le cambiaba la clase a abdomen.

2. Una vez definida la región abdominal se segmentaban los cortes pertenecientes a ésta para identificar la grasa visceral y subcutánea. Así, siguiendo la misma idea anterior, se entrenó un ensamblado de cinco modelos. De cada uno de ellos se obtenía un mapa de probabilidad formado por los tres canales definidos anteriormente (fondo, grasa visceral y grasa subcutánea). Para obtener el mapa de probabilidad final, para cada canal, se realizó la media aritmética de las cinco salidas y la clase final de cada vóxel quedaba definida como la clase de mayor probabilidad. Las máscaras obtenidas se postprocesaron de la siguiente forma:

- La máscara correspondiente a la grasa subcutánea se esperaba que estuviera formada por una única región conectada. De este modo, se buscaron todas las componentes independientes conectadas entre sí y se seleccionó únicamente aquella que tuviera el mayor volumen.
- Tanto para la máscara correspondiente a la grasa subcutánea resultante del punto anterior como para la correspondiente a la grasa visceral, se realizó un postproceso corte a corte. Así, para cada corte se buscaron las componentes interconectadas y se eliminaron aquellas que estuvieran formadas por menos de 4 píxeles. Este paso permitió eliminar regiones ruidosas que se segmentaban de manera aislada (Figura 52).

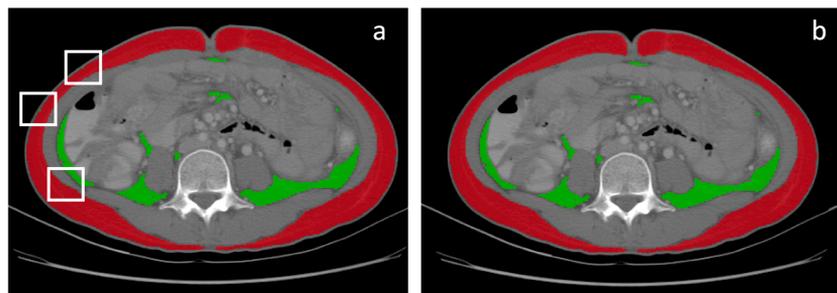


Figura 52. Resultado de eliminar componentes interconectadas corte a corte formadas por menos de 4 píxeles. a) Máscara antes de aplicar el filtrado, en blanco se resaltan las zonas donde encontramos las regiones ruidosas a eliminar. b) Resultado de aplicar el postprocesado.

- Finalmente, se eliminaron de las máscaras todos aquellos píxeles segmentados que no se encontraban dentro de los rangos de UH esperados, esto es, entre -190 y -30 UH para la grasa subcutánea y -150 y -50 UH para la grasa visceral.

Para evaluar la precisión de las máscaras obtenidas sobre el conjunto de datos de entrenamiento se emplearon una serie de métricas dependiendo del problema a resolver.

3.4.5.2. Métricas utilizadas en los modelos de segmentación

Existen diferentes métricas que se pueden emplear para evaluar la precisión de la máscara generada para la segmentación. Éstas se basan en la comparación entre la máscara predicha (P), generada automáticamente por la CNN, y una máscara generada de forma manual, considerada la máscara de referencia (R), y se pueden clasificar en tres grupos principales: basadas en el solapamiento, basadas en la diferencia de volúmenes o basadas en la distancia entre las máscaras.

Métricas basadas en el solapamiento

Estas métricas se emplean principalmente para analizar la proporción de vóxeles correctamente clasificados como región de interés y/o fondo. La principal desventaja de estas métricas es que penalizan errores en máscara pequeñas en mayor medida que en máscaras grandes.

- **Coefficiente Dice (DC, del inglés *Dice Coefficient*)**

El DC mide la desalineación entre dos formas midiendo la región de superposición entre ellas haciendo uso de la Ecuación 48. Adquiere valores entre 0 y 1, donde 0 significa que no existe superposición entre dos máscaras binarias y 1 indica que hay un solapamiento total entre ambas. Esta métrica es la más empleada en la validación de métodos de segmentación en imagen médica.

$$DC(P, R) = \frac{2 \cdot |P \cap R|}{|P| + |R|} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (48)$$

- **Intersección sobre la unión (IoU, del inglés *Intersection over Union*)**

La IoU, también conocida como índice Jaccard, mide la similitud entre dos objetos y se define como el tamaño de la intersección dividido por el tamaño de la unión entre ambos objetos (Ecuación 49).

$$IoU(P, R) = \frac{|P \cap R|}{|P \cup R|} = \frac{|P \cap R|}{|P| + |R| - |P \cap R|} \quad (49)$$

- **Sensibilidad**

La sensibilidad, también conocida como Relación de Verdaderos Positivos (RVP) o *recall*, mide la proporción de vóxeles pertenecientes a la región de interés en la máscara de referencia que se han clasificado correctamente como región de interés (VP, Verdaderos Positivos) de acuerdo con la Ecuación 50. Toma valores entre 0 y 1, donde un valor de 1 indica que todos los vóxeles segmentados como región de interés pertenecen a ella.

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (50)$$

Donde FN se refiere a los Falsos Negativos, es decir, los vóxeles clasificados erróneamente como fondo.

- **Especificidad**

La especificidad, también conocida como Fracción de Verdaderos Negativos (FVN), mide la proporción de vóxeles pertenecientes al fondo en la máscara de referencia que se han identificado correctamente como fondo de la máscara predicha (VN, Verdaderos Negativos) de acuerdo con la Ecuación 51. Toma valores entre 0 y 1, donde 1 significa que todos los vóxeles clasificados como fondo realmente pertenecen al fondo.

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (51)$$

Donde FP se refiere a los Falsos Positivos, es decir, los vóxeles clasificados erróneamente como pertenecientes a la región de interés.

- **Precisión**

La precisión mide la proporción de vóxeles clasificados por la CNN como pertenecientes a la región de interés que realmente pertenecen a esta (VP, Verdaderos Positivos) de acuerdo con la Ecuación 52. Toma valores entre 0 y 1, donde un valor de 1 indica que la máscara generada de forma automática se superpone completamente con la máscara de referencia.

$$Precisión = \frac{VP}{VP + FP} \quad (52)$$

- **Tasa de falsos aciertos**

La tasa de falsos aciertos (FDR, del inglés *False Discovery Rate*) mide, de todos los vóxeles segmentados por la CNN, cuantos se han segmentado erróneamente como pertenecientes a la región de interés de acuerdo con la Ecuación 53. Esta métrica adquiere importancia cuando la máscara predicha debe de ser empleada para una posterior cuantificación, donde interesa que todos los vóxeles analizados pertenezcan a la región de interés (verdaderos positivos), siendo menos importante que algunos vóxeles, pertenecientes a la región, dejen de segmentarse (falsos negativos).

$$FDR = \frac{FP}{VP + FP} \quad (53)$$

Métricas basadas en el volumen

Estas métricas se basan en la comparación del volumen calculado sobre la máscara de referencia (volumen real) y el calculado sobre la máscara predicha (volumen predicho). Estas métricas no consideran la posición de las máscaras, es decir, su solapamiento. Por lo tanto, las diferencias en volumen pueden ser bajas, pero los vóxeles segmentados encontrarse en posiciones diferentes. Estas métricas adquieren relevancia cuando el volumen de la región de interés es importante, pero no nos importa el contorno de ésta o la correcta segmentación de la forma en general. Entre estas métricas encontramos:

- **Diferencia de volumen relativo (RVD)**

La diferencia de volumen relativo (RVD, del inglés *Relative Volume Difference*) mide la diferencia de volumen de la máscara generada automáticamente (V_P) y de la máscara de referencia (V_R), relativo al volumen de referencia de acuerdo con la Ecuación 54.

$$RVD(P, R) = \frac{|V_P - V_R|}{V_R} \quad (54)$$

- **Coefficiente de correlación (r)**

Análisis de la correlación entre los volúmenes reales y predichos. Esta correlación se realizará con el coeficiente de Pearson o Spearman, en función de si las distribuciones de volumen son normales o no, respectivamente. Para conocer si la distribución es normal, o no, se empleará la prueba de Shapiro-Wilk, definiéndose como normales aquellas distribuciones con un $p > 0,05$.

Métricas basadas en la distancia

Estas métricas se definen como funciones de la distancia Euclídea entre los vóxeles segmentados automáticamente y los vóxeles de la máscara de referencia. Entre este tipo de métricas encontramos:

- **Distancia media simétrica superficial (ASSD, del inglés *Average Symmetric Surface Distance*)**

La ASSD mide la distancia Hausdorff promediada sobre todos los puntos de la máscara. La ASD directa desde la máscara G a la máscara P viene dada por la suma de todas las distancias mínimas de todos los vóxeles segmentados en G a todos aquellos segmentados en P, dividido por el número de vóxeles segmentados en G. La ASD directa se puede calcular desde la máscara de referencia a la máscara predicha y viceversa. Por lo tanto, para calcular la ASSD, ambas opciones se calculan y se promedian (Ecuación 55).

$$ASSD(P, R) = \left[\frac{1}{P} \sum_{p \in P} \min_{r \in R} d(p, r) + \frac{1}{R} \sum_{r \in R} \min_{p \in P} d(p, r) \right] / 2 \quad (55)$$

Donde $d(p, r)$ hace referencia a la distancia Euclídea entre dos puntos. La ecuación anterior se simplifica como la Ecuación 56.

$$ASSD(P, R) = \frac{\frac{PaR}{P} + \frac{RaP}{R}}{2} \quad (56)$$

- **Máxima distancia superficial (MSD, del inglés, *Maximum Surface Distance*)**

La MSD, también conocida como distancia Hausdorff, mide la mayor diferencia entre dos superficies de acuerdo con la Ecuación 57, donde un valor más elevado

de MSD indica una distancia mayor entre los contornos de ambas máscaras y donde un valor de 0 significa que ambos volúmenes comparten el mismo contorno.

$$MSD(P, R) = \max \left\{ \max_{p \in P} \min_{r \in R} d(p, r), \max_{r \in R} \min_{p \in P} d(p, r) \right\} \quad (57)$$

Aquellos modelos cuyo principal objetivo es obtener una segmentación para, posteriormente, realizar una cuantificación de la región analizada, podemos extraer métricas que nos den información sobre la precisión en la cuantificación. Entre estas encontramos:

- **Coefficiente de correlación (r)** de Pearson (distribuciones normales) o Spearman (distribuciones no normales) entre cada uno de los valores cuantitativos (e.g., PDFF, R2*, AUC₆₀) obtenidos sobre la máscara de referencia y sobre la máscara predicha. Al igual que para el volumen, para conocer si la distribución es normal o no se empleará la prueba de Shapiro-Wilk, definiéndose como normales aquellas distribuciones con un $p > 0,05$.
- **Error relativo (ER)** en la cuantificación al comparar los valores cuantitativos obtenidos sobre ambas máscaras. Dado un biomarcador BMK, el error relativo se mide de acuerdo con la Ecuación 58.

$$ER_{BMK} = \left| \frac{BMK_R - BMK_P}{BMK_R} \right| \quad (58)$$

- **Análisis de Bland-Altman**, se trata de un método gráfico que compara dos técnicas de medición sobre una misma variable cuantitativa. Este método mide la diferencia media (sesgo) y un rango del 95% confianza entre ambos métodos. Se va a emplear para comparar los valores cuantitativos obtenidos al emplear la segmentación manual de referencia frente a emplear la segmentación generada automáticamente por los distintos algoritmos desarrollados.

3.4.5.3. Métricas utilizadas en la clasificación

A continuación, se van a listar las métricas que se emplearon para estudiar el rendimiento del clasificador entrenado para la identificación de las imágenes pertenecientes a la región abdominal en las series de TC, donde se definen dos clases: clase abdomen (clase positiva) y clase no abdomen (clase negativa).

- **Exactitud**

La exactitud, con un valor entre 0 y 1, mide la proporción de imágenes correctamente clasificadas entre el total de imágenes analizadas (Ecuación 59).

$$Exactitud = \frac{VP + VN}{VP + FP + VN + FN} \quad (59)$$

- **Sensibilidad, Especificidad y Precisión**

Medidas del mismo modo que el descrito en la sección de segmentación de acuerdo con las Ecuaciones 45, 46 y 47, respectivamente.

- **F1 Score**

El *F1 Score*, con un valor entre 0 y 1, mide la media armónica de la precisión y la sensibilidad (Ecuación 60). El F1 Score es una buena métrica para emplearse cuando queremos desarrollar un modelo que tenga tanto buena precisión como sensibilidad.

$$F1\ Score = 2 \cdot \frac{precisión \cdot sensibilidad}{precisión + sensibilidad} \quad (60)$$

- **Área bajo la curva ROC**

El área bajo la curva ROC (AUC, del inglés *Area Under the Curve*), con un valor entre 0 y 1, indica cómo de bien, las probabilidades de la clase positiva se separan de la clase negativa.

3.5. Cuantificación

En este apartado se detalla la implementación realizada con relación a los métodos de cuantificación. Principalmente nos centramos en la cuantificación de la grasa y el hierro en secuencias MECSE de RM, para la caracterización de la enfermedad difusa hepática, y en la cuantificación de los descriptores de las curvas de contraste (análisis semicuantitativo) en secuencias de perfusión de RM, para la caracterización hepática en casos con CHC.

3.5.1. Cuantificación PDFF y R2* en secuencias MECSE de RM

Como ya introdujimos en el apartado 1.3.2 *Biomarcadores de imagen para la caracterización de la enfermedad difusa hepática*, para la caracterización de la enfermedad difusa hepática se realiza la cuantificación de la fracción de grasa y concentración de hierro en el parénquima hepático. Así, empleando la máscara obtenida con el modelo entrenado previamente, se cuantifica, vóxel a vóxel, la PDFF y R2* haciendo uso de la Ecuación 61.

$$|S(TE)| = \sqrt{S_w^2 \cdot e^{-2 \cdot R2_w^* \cdot TE} + S_f^2 \cdot e^{-2 \cdot R2_f^* \cdot TE} + 2 \cdot S_w \cdot S_f \cdot e^{-(R2_w^* + R2_f^*) \cdot TE} \cos(\omega \cdot TE)} \quad (61)$$

Donde la PDFF se calcula de acuerdo con la Ecuación 62.

$$PDFF (\%) = \frac{S_f}{S_f + S_w} \cdot 100 \quad (62)$$

De este modo, se desarrolló un algoritmo que permitiera ajustar las variables:

$S_w \rightarrow$ Señal del agua.

$S_f \rightarrow$ Señal de la grasa.

$R2_w^* \rightarrow$ Coeficiente de relaxometría R2* en agua, siendo la variable que se relaciona con la concentración de hierro.

$R2_f^* \rightarrow$ Coeficiente de relaxometría R2* en grasa.

Para la implementación del código se hizo uso del lenguaje de programación Python, empleando, para el ajuste de las variables definidas anteriormente, la librería NLOpt³. Se trata de una librería de código abierto para la optimización no lineal que integra diferentes algoritmos. Estos se basan en la minimización de una función objetivo permitiendo, además, establecer límites (mínimo y/o máximo) para las variables a ajustar.

En este problema, la función a minimizar fue la definida según la Ecuación 63.

$$\min_{x \in \mathbb{R}^n} f(x) = \sum_{i=0}^n \left[S_i - \sqrt{S_w^2 \cdot e^{-2 \cdot R2_w^* \cdot TE} + S_f^2 \cdot e^{-2 \cdot R2_f^* \cdot TE} + 2 \cdot S_w \cdot S_f \cdot e^{-(R2_w^* + R2_f^*) \cdot TE} \cos(\omega \cdot TE)} \right]^2 \quad (63)$$

Donde n es el número de vóxeles que definen el parénquima hepático.

Para definir el algoritmo óptimo a emplear, se evaluaron 15 algoritmos disponibles en la librería NLOpt y, para cada uno de ellos, se establecieron dos criterios de parada: número de iteraciones y tolerancia. Así, para cada algoritmo, se evaluaron 9 tolerancias diferentes (10^{-2} hasta 10^{-10} en salto de 0,1) combinadas con 10 números de iteraciones (100 hasta 1000 en saltos de 100). Los algoritmos evaluados fueron:

- COBYLA (*Constrained Optimization BY Linear Approximations*).
- BOBYQA.
- NEWUOA con restricción de límites.
- PRAXIS (*Principal Axis*).
- Nelder-Mead Simplex.
- Sbxplx: re-implementación del método Subplex.
- MMA (*Method of Moving Asymptotes*).
- CCSAQ (*Conservative Convex Separable Quadratic Approximation*).
- SLSQP (*Sequential Least-Squares Quadratic Programming*).
- *Low-storage BFGS*.
- *Preconditioned truncated Newton*: de este se generan tres algoritmos diferentes:
 - Variante preconditionada por el algoritmo *low-storage BFGS* con reinicio de la pendiente más pronunciada.
 - Igual que la anterior, pero sin reinicio.
 - Igual que la anterior, pero sin reinicio y preconditionación.
- *Shifted limited-memory variable-metric*: de este surgen dos algoritmos:

³ <https://nlopt.readthedocs.io/en/latest>

- VAR1: Empleando el método Rank-1
- VAR2: Empleando el método Rank-2

De todos estos algoritmos, empleando la base de datos presentada en el apartado 3.1.1.1. *Cuantificación y validación clínica*, se seleccionó aquel que ofrecía un menor error en la minimización de la función definida por la Ecuación 63. Para cada paciente se ajustó la señal media perteneciente al hígado. Para el ajuste de los parámetros, en todos ellos, se estableció el valor mínimo para los cuatro parámetros a ajustar en 0, dejando libre el límite superior.

Una vez seleccionado el algoritmo que mejor se ajustaba a nuestro problema, éste se empleó para cuantificar la grasa (PDFF) y el hierro ($R2^*$) vóxel a vóxel en todo el parénquima hepático. Esto se realizó tanto sobre la segmentación manual de referencia como sobre la obtenida por el método de segmentación automática. De cada biomarcador, en cada segmentación independientemente, se calculó la media, mediana, desviación estándar y percentiles 25 y 75. La mediana de ambas máscaras, se empleó, en primer lugar, para analizar la correlación y el ER obtenido.

A continuación, se realizó una validación clínica del método desarrollado. Para esta validación se relacionaron los resultados extraídos de la RM haciendo uso de la segmentación automática, con la información obtenida a partir de la muestra histológica de la biopsia hepática, que se tomó como referencia. De la muestra histológica disponíamos de dos tipos de información que permitieron realizar diferentes análisis:

- Estadiaje histológico de la esteatosis y depósitos de hierro:
 - Se calculó la mediana del PDFF y $R2^*$ para cada paciente y se diferenciaron aquellos pertenecientes a los diferentes estadios de esteatosis y depósitos de hierro. Para cada caso, se analizó si había diferencias significativamente estadísticas entre los diferentes grupos (S0-S3 para la esteatosis, Fe1-Fe4 para el hierro) mediante la prueba Kruskal-Wallis, una prueba no paramétrica para evaluar si las muestras tienen las mismas medianas (hipótesis nula).
 - Se realizó una categorización de los grupos y se analizó la capacidad de clasificación en éstos mediante el cálculo del área bajo la curva ROC (AUC) y, tras definir el umbral de PDFF y $R2^*$ respectivamente, la sensibilidad y la especificidad. Los grupos para la esteatosis, como se ha descrito en la *sección 3.1.1.1. Cuantificación y validación clínica*, fueron:

Capítulo 3. Materiales y métodos

- Esteatosis leve: S0 vs. S1-3
- Esteatosis moderada: S0-1 vs. S2-3
- Esteatosis grave: S0-2 vs. S3

Mientras que los grupos para la sobrecarga de hierro fueron:

- Sobrecarga presente: Fe0 vs. Fe1-4
 - Sobrecarga leve: Fe0-1 vs. Fe2-4
 - Sobrecarga moderada: Fe0-2 vs. Fe3-4
 - Sobrecarga grave: Fe0-3 vs. Fe4
- Comparativa del PDFF y R2* con la proporción de área (%) con presencia de grasa y hierro obtenida tras la digitalización de la muestra histológica. Para ello, se realizó un estudio de correlación entre la mediana del PDFF y R2* y la información de la patología digital. Dado que se trataba de distribuciones que no seguían una distribución normal, se empleó el coeficiente de Spearman para el estudio de esta correlación.

3.5.1.1. Paralelización

La cuantificación se realiza para cada uno de los vóxeles que forman la región hepática segmentada, por lo tanto, son miles las curvas que hay que ajustar en este proceso para un único caso. Para acelerar el proceso de cuantificación se implementó un sistema de paralelización de forma que se realizó la cuantificación de varios vóxeles en paralelo, en función del número de núcleos empleados en cada caso. Así, haciendo uso de una máquina con un procesador 2,6 GHz Intel Core i7 de 6 núcleos se analizaron 10 casos con un número medio de 66.397 vóxeles a analizar empleando desde 1 hasta 12 núcleos para el cálculo.

3.5.2. Descriptores de la curva concentración en secuencias de perfusión de RM

En las secuencias de perfusión de RM podemos realizar el análisis descrito en el apartado

1.3.3.2. Análisis de perfusión, en éste se describen dos tipos de análisis, el semicuantitativo y el farmacocinético. Para el análisis farmacocinético un requerimiento es tener secuencias de perfusión con múltiples dinámicos (>10) y con buena resolución temporal, esto es, tiempo corto entre dinámicos. Esto no se cumple en la base de datos

empleada, ya que contamos con secuencias de perfusión con 4-5 dinámicos como máximo. Por este motivo, únicamente se realizó un análisis semicuantitativo.

Como se comenta en el apartado

1.3.3.2. Análisis de perfusión el primer paso es la conversión de la curva de intensidad-tiempo a concentración-tiempo, para ello, dado que se empleó Gadolinio como contraste, se estableció un valor de relajatividad (R) de $3,5 \text{ mM}^{-1} \cdot \text{s}^{-1}$ para estudios de 3T y $3,6 \text{ mM}^{-1} \cdot \text{s}^{-1}$ para estudios de 1,5T (DOTAREM) [106]. De acuerdo con [60], el tiempo de T1 del hígado (T1(0)) empleado fue de 586 ms en estudios de 1,5T y de 809 ms en estudios de 3T.

Finalmente, una vez se obtuvo la curva concentración-tiempo se calcularon, para cada caso de la base de datos descrita en *3.1.3. Segmentación automática hepática en secuencias de perfusión*, los descriptores principales (

1.3.3.2. Análisis de perfusión – Análisis semicuantitativo):

- Concentración máxima: concentración del dinámico que presenta el valor máximo.
- Tiempo al pico: tiempo del dinámico en el que se produce la concentración máxima.
- Pendiente de captación: se calcula la pendiente de cada punto de la curva y se selecciona el valor máximo.
- Área bajo la curva a los 60 segundos (AUC₆₀): en primer lugar, se identifica el momento en el que empieza la captación de contraste, una vez identificada, se calcula el área de la región de la curva comprendida entre ese punto y 60 segundos adicionales.

Estos descriptores se calcularon vóxel a vóxel en toda la región hepática empleando, por un lado, la segmentación manual de referencia y, por otro, la segmentación automática. De cada biomarcador se calculó la media, desviación estándar, mediana y percentiles 25 y 75. Los valores de mediana de cada descriptor se emplearon para analizar la correlación y ER obtenido al emplear la segmentación automática.

Capítulo 4. Resultados y discusión

En este apartado se presentan los resultados obtenidos en cada uno de los desarrollos realizados a lo largo de la presente tesis doctoral. Por lo tanto, se muestran los resultados obtenidos para cada uno de los métodos de segmentación automática entrenados, así como la configuración de hiperparámetros que mejores resultados ofrecieron. Además, se muestra la validación realizada en relación a la cuantificación de los diferentes biomarcadores de imagen introducidos.

4.1. Segmentación hepática y cuantificación en secuencias MECSE

En esta sección se muestran los resultados obtenidos en el método de segmentación y cuantificación automática de secuencias MECSE. En primer lugar, se muestran los resultados obtenidos mediante el proceso de validación y ajuste de hiperparámetros y, a continuación, se muestran los resultados obtenidos en la fase de test. Esta última está formada por dos test independientes, uno empleando imágenes provenientes de la misma fuente que las empleadas para el entrenamiento del modelo y, otro subconjunto nuevo obtenido de centros y escáneres diferentes a los empleados en el desarrollo del modelo. En segundo lugar, se muestra la validación realizada al método de cuantificación de PDFF y $R2^*$, tanto técnicamente mediante la selección del algoritmo que minimiza el error de cuantificación, como clínicamente mediante el estudio de correlación entre dichos parámetros y los resultados de los análisis de la biopsia hepática.

4.1.1. Validación del método de segmentación hepática

Para la construcción de este modelo se empleó el optimizador Adam con una tasa de aprendizaje inicial de 10^{-5} . Además, se realizó una comparativa entre el uso de dos funciones de pérdidas: la entropía cruzada y las pérdidas Dice. En la Figura 53 se puede observar la diferencia en el coeficiente Dice durante el proceso de entrenamiento tanto sobre el conjunto de datos de entrenamiento como sobre el conjunto de datos de validación empleando ambas funciones de pérdidas. Podemos ver cómo se obtiene un mejor rendimiento (convergencia más rápida y mayor asíntota) empleando las pérdidas Dice. Esto puede deberse al desbalanceo que existe entre el número de píxeles que pertenecen a la región de interés (hígado) y el fondo. En estos casos se ha visto, que funciones de pérdidas que le dan atención al objeto de interés frente al fondo, ofrecen mejores resultados [107].

Capítulo 4. Resultados y discusión

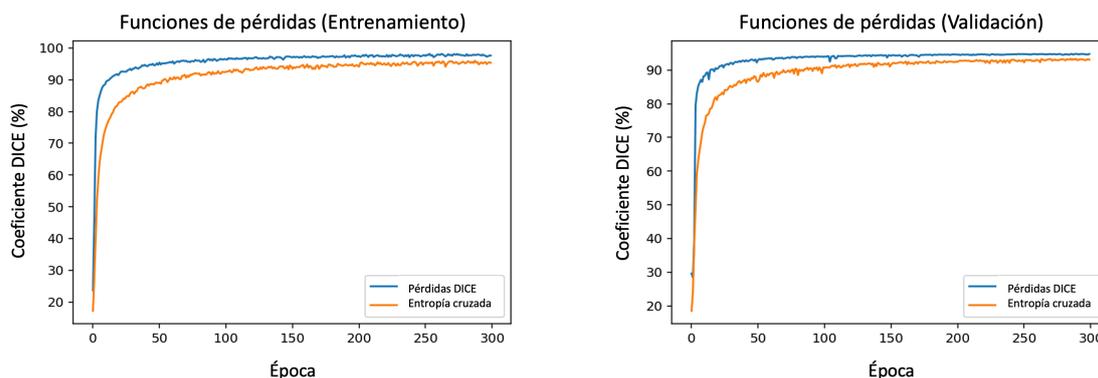


Figura 53. Coeficiente Dice obtenido durante el proceso de entrenamiento en la evaluación del modelo sobre el conjunto de datos de entrenamiento (izquierda) y sobre el de validación (derecha) empleando una función de pérdidas basada en la entropía cruzada (naranja) y en las pérdidas Dice (azul).

En la Tabla 6 podemos encontrar, para cada una de las iteraciones del proceso de validación cruzada, el valor de coeficiente Dice máximo obtenido sobre el conjunto de datos de validación, así como el valor correspondiente sobre el conjunto de datos de entrenamiento. Estos resultados son los correspondientes a emplear una función de pérdidas Dice. Podemos observar un valor de coeficiente Dice en validación elevado (>94% en todas las iteraciones) y estable, esto es, se obtienen resultados similares independientemente del conjunto de datos de entrenamiento y validación, por lo tanto, podemos decir que nuestro modelo ofrece resultados reproducibles.

Tabla 6. Coeficiente Dice máximo obtenido sobre el conjunto de datos de validación en cada iteración (it.) del proceso de validación cruzada, así como el correspondiente valor sobre los datos de entrenamiento.

	It. 1	It. 2	It. 3	It. 4	It. 5
DC validación (%)	95,05	95,17	95,26	94,92	95,38
DC entrenamiento (%)	98,07	98,02	98,14	97,87	97,87

4.1.2. Test del método de segmentación hepática

Para la resolución de este problema disponíamos de dos conjuntos diferentes de test. Uno formado por estudios independientes, pero adquiridos en los mismos centros y escáneres que los empleados para el desarrollo del modelo (test interno), y otro formado por estudios adquiridos en centros y máquinas diferentes (test externo). Para el test externo, además, se han diferenciado las métricas por institución/máquina.

La metodología desarrollada, así como los resultados han sido previamente publicados en [108].

Capítulo 4. Resultados y discusión

4.1.2.1. Test interno

En la Tabla 7 se encuentran las métricas obtenidas en el test del modelo sobre un conjunto de datos independiente al de entrenamiento pero que se adquirieron empleando los mismos escáneres (n=46).

Tabla 7. Métricas obtenidas en la comparativa de la máscara de referencia (manual) y la generada automáticamente por la CNN en el conjunto de datos de test interno (n=46). Para cada métrica se ha calculado la mediana, media y desviación estándar.

Métrica	Mediana	Media±Std
DC	0,95	0,94±0,02
IoU	0,90	0,89±0,04
Sensib.	0,95	0,94±0,04
Especif.	1,00	1,00±0,00
Precisión	0,96	0,95±0,02
FDR	0,04	0,05±0,02
RVD	0,03	0,03±0,03
ASSD (mm)	1,39	1,62±0,75
MSD (mm)	26,27	26,99±7,49

Podemos ver cómo hay un elevado grado de solapamiento entre las segmentaciones predichas y de referencia, representado con una mediana de DC del 95%. Además, podemos ver como la mediana de la tasa de falsos aciertos es del 4%, lo que supone que el 96% de los vóxeles segmentados corresponden realmente a hígado, minimizando, de este modo, el sesgo introducido en cuantificaciones posteriores.

También se realizó un estudio de la correlación entre el volumen medido sobre la máscara realizada manualmente y la obtenida de la CNN entrenada. Con la prueba Shapiro-Wilk se vio que la distribución de los volúmenes de este conjunto de datos no seguía una distribución normal ($p < 0,05$), por lo tanto, se realizó el análisis de correlación haciendo uso del coeficiente de Spearman. Así, se obtuvo una $r = 0,98$ ($p < 0,001$), es decir, se observó una elevada correlación entre ambos volúmenes. Además, se realizó el análisis Bland-Altman para comparar las diferencias entre ambos volúmenes (Figura 54). Este análisis muestra un sesgo bajo (0,07 L) y límites de acuerdo estrechos (-0.11-0.26 L).

Capítulo 4. Resultados y discusión

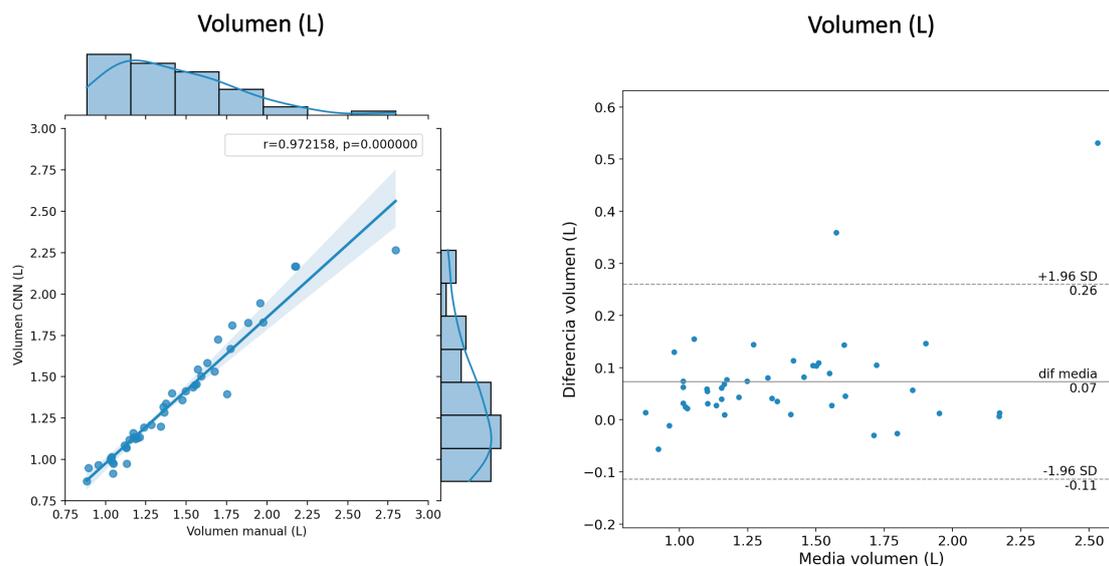


Figura 54. Análisis de correlación entre el volumen hepático calculado sobre la segmentación de referencia (manual) y la predicha por la CNN sobre una imagen de RM MECSE. A la izquierda vemos el diagrama de pares entre ambos volúmenes; en los ejes superior y derecho se encuentran los diagramas de los volúmenes de la segmentación manual y predicha, respectivamente. A la derecha se muestra el diagrama Bland-Altman.

A continuación, se evaluó el sesgo introducido en la cuantificación cuando se emplea como región de interés la máscara obtenida automáticamente con la CNN entrenada frente al uso de la máscara realizada manualmente (Figura 55). Para ello, empleando el algoritmo de cuantificación de la PDFF y el $R2^*$, introducido anteriormente, se analizaron todos los casos y se realizó el estudio de correlación entre los valores de mediana obtenidos a lo largo de todo el parénquima hepático en ambas máscaras. Con la prueba Shapiro-Wilk se observó que tanto la distribución de valores de PDFF como de $R2^*$ no seguían una distribución normal ($p < 0,05$), por lo tanto, se empleó el coeficiente de Spearman para el estudio de la correlación. Finalmente, se obtuvo una correlación (r) de 1 para ambos casos ($p < 0,001$). De este modo, pese a que la correlación en volumen entre ambas máscaras no fue perfecta, sí que lo fue la correlación de PDFF y $R2^*$, es decir, no se introdujo sesgo en la cuantificación al emplear un método de segmentación automático basado en CNN. Además, al igual que para el volumen, se realizó el análisis Bland-Altman obteniendo un sesgo bajo (PDFF: 0,12% y $R2^*$: $0,27s^{-1}$) y límites de acuerdo estrechos (PDFF: $[-0,12 - 0,36]$ % y $R2^*$: $[-0,18 - 0,71]$ s^{-1}).

Capítulo 4. Resultados y discusión

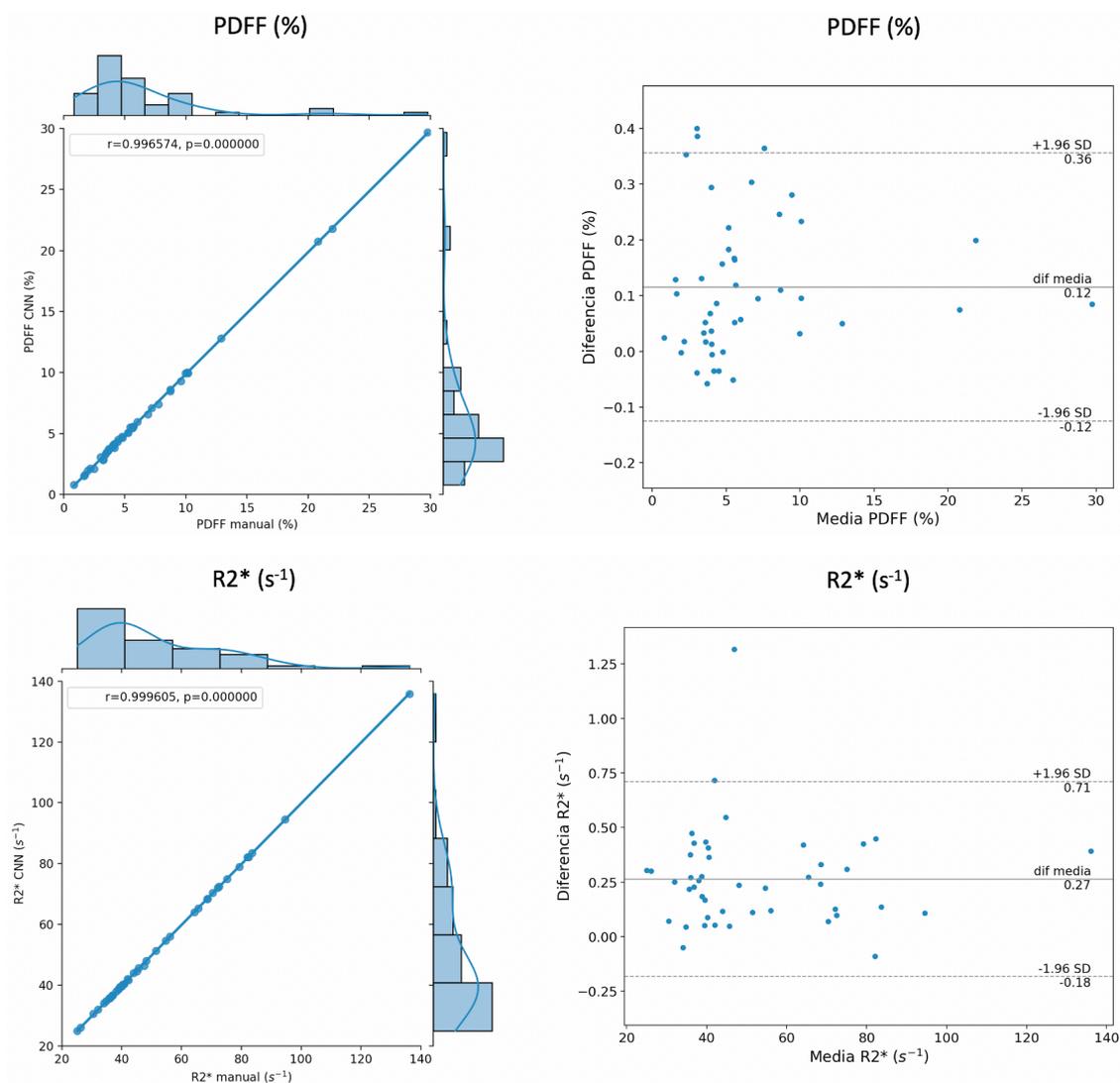


Figura 55. Análisis de correlación entre el PDFFF (fila superior) y R2* (fila inferior) cuantificados sobre el parénquima hepático empleando la segmentación de referencia (manual) y la predicha por la CNN. A la izquierda vemos el diagrama de pares entre ambas máscaras; en los ejes superior y derecho se encuentran los diagramas de cada uno de los biomarcadores cuantificados sobre la segmentación manual y predicha, respectivamente. A la derecha se muestra el diagrama Bland-Altman.

Finalmente, se analizó el ER en la cuantificación de la PDFFF y R2*, con un valor medio \pm desviación estándar de $2,8 \pm 3,25\%$ para la PDFFF y de $0,59 \pm 0,52\%$ para el R2*; y una mediana de $1,4\%$ y $0,47\%$, respectivamente.

La Figura 56 muestra el rendimiento de la red entrenada al segmentar tres casos diferentes pertenecientes al conjunto de datos de test interno. Para cada caso se muestra la comparativa entre la segmentación generada de forma manual (rojo) y la predicha por la CNN (amarillo). Se muestra, además, el rendimiento de la red al segmentar un caso en el que el paciente le realizaron una hepatectomía (extirpación del hígado) parcial. En este caso se puede observar cómo, pese a dejarse ciertas regiones del parénquima hepático sin segmentar (falsos negativos), la región extirpada del hígado se excluyó correctamente de

la segmentación. También podemos ver como grandes vasos, como la vena cava inferior, se excluyeron correctamente de la máscara.

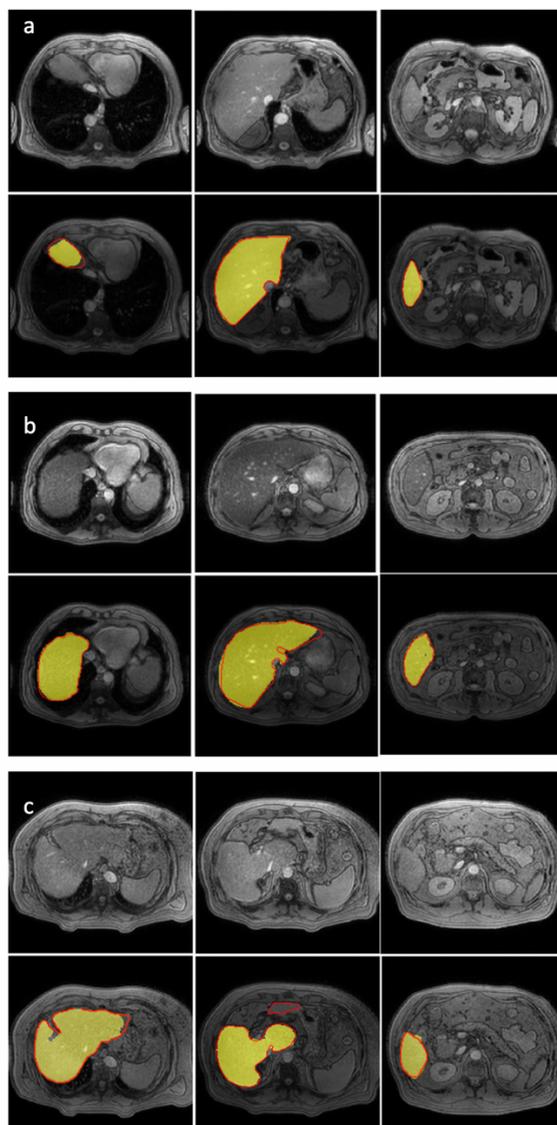


Figura 56. Segmentación obtenida en tres casos diferentes del test interno mostrando, para cada caso, tres niveles anatómicos diferentes. En amarillo se representa la máscara predicha por la CNN y en rojo el contorno de la máscara de referencia realizada de manera manual. a) Segmentación en un estudio adquirido en uno de los centros incluidos para el desarrollo del modelo. b) Segmentación en un estudio proveniente del otro centro empleado para el desarrollo del modelo. c) Segmentación obtenida en un estudio que pertenece a un paciente con hepatectomía parcial.

4.1.2.2. Test externo

El mismo estudio que el anterior se realizó sobre un conjunto de datos adicional, también independiente al empleado para el entrenamiento del modelo, pero, en este caso, de dos centros y escáneres diferentes. Mientras que el conjunto de datos de entrenamiento y test interno provenían de dos centros (España y Portugal) ambos con una RM Philips de 3T, el conjunto de validación externo estaba formado por imágenes provenientes de un centro

Capítulo 4. Resultados y discusión

en Argentina con una RM Siemens de 3T (n=19) y otro centro en Chile con una RM Philips de 1,5T (n=10). De esto modo, se evaluó la capacidad de generalización del modelo a otros protocolos de adquisición, fabricantes y campos magnéticos.

En primer lugar, se analizaron las métricas obtenidas para el primer conjunto de datos del test externo, es decir, el correspondiente al centro de Argentina con una RM Philips de 3T (mismo campo magnético y distinto fabricante del empleado para el entrenamiento del modelo). Como se puede observar en la Tabla 8, para la mayor parte de las métricas se obtiene un mejor resultado que el obtenido en el propio test interno. Por ejemplo, se ha producido una disminución de la tasa de falsos aciertos del 4% al 2%. Esto significa que pese a estar el modelo construido con imágenes de Philips, ante el mismo campo magnético, se observa una generalización del modelo al emplear imágenes de otro fabricante, Siemens en este caso.

Tabla 8. Métricas obtenidas en la comparativa de la máscara de referencia (manual) y la generada automáticamente por la CNN en el conjunto de datos de test externo (centro de Argentina – Siemens, 3T). Para cada métrica se ha calculado la mediana, media y desviación estándar.

Métrica	Mediana	Media±Std
DC	0,95	0,94±0,02
IoU	0,90	0,89±0,04
Sensib.	0,92	0,91±0,05
Especif.	1,00	1,00±0,00
Precisión	0,98	0,98±0,02
FDR	0,02	0,02±0,01
RVD	0,06	0,07±0,05
ASSD (mm)	1,06	1,10±0,60
MSD (mm)	25,70	27,32±9,67

A continuación, se realizó el mismo estudio para los casos del segundo centro que, en este caso, correspondían al centro de Chile con una RM de Philips de 1,5T (mismo fabricante, distinto campo magnético). Como se puede apreciar en la Tabla 9, al emplear este escáner se obtuvieron peores resultados comparando tanto con el test interno como con el test externo anterior. En este caso, se produjo una disminución del DC a un 86% y un aumento de la tasa de falsos aciertos al 7%.

La disminución en la precisión obtenida en este subconjunto de imágenes puede deberse a que éstas se adquirieron con una baja resolución en el plano (2,3 x 2,3 mm²) y un

Capítulo 4. Resultados y discusión

elevado ancho de corte (10 – 10,5 mm), lo cual agrava los efectos de volumen parcial y, por tanto, dificultan la diferenciación de los tejidos.

Sin embargo, a pesar de esto, como veremos a continuación, estos resultados son buenos para emplearse en práctica clínica para la cuantificación automática de PDFF y R2*

Tabla 9. Métricas obtenidas en la comparativa de la máscara de referencia (manual) y la generada automáticamente por la CNN en el conjunto de datos de test externo (centro de Chile – Philips, 1,5T). Para cada métrica se ha calculado la mediana, media y desviación estándar.

Métrica	Mediana	Media±Std
DC	0,86	0,87±0,06
IoU	0,76	0,77±0,10
Sensib.	0,82	0,83±0,11
Especif.	1,00	1,00±0,00
Precisión	0,93	0,92±0,04
FDR	0,07	0,08±0,04
RVD	0,08	0,12±0,11
ASSD (mm)	3,88	3,83±2,02
MSD (mm)	28,77	37,43±16,71

A continuación, se realizó el estudio de correlación de ambos subconjuntos de datos (Figura 57). En ambos casos, tras realizar la prueba Shapiro-Wilk, se obtuvo que los volúmenes seguían una distribución normal ($p > 0,05$), por lo tanto, se empleó el coeficiente de Pearson para el estudio de la correlación. Obteniendo una $r = 0,97$ para el escáner de Siemens, 3T y una $r = 0,87$ para el escáner de Philips, 1,5T ($p < 0,001$ en ambos casos). Además, con el análisis Bland-Altman se estudió la relación entre los dos volúmenes, obteniendo en ambos casos un sesgo bajo (Siemens 3T: 0,07L; Philips 1,5T: 0,11L) y límites de acuerdo estrechos (Siemens 3T: [-0,04 – 0,18]; Philips 1,5T: [-0,15 – 0,37]).

Capítulo 4. Resultados y discusión

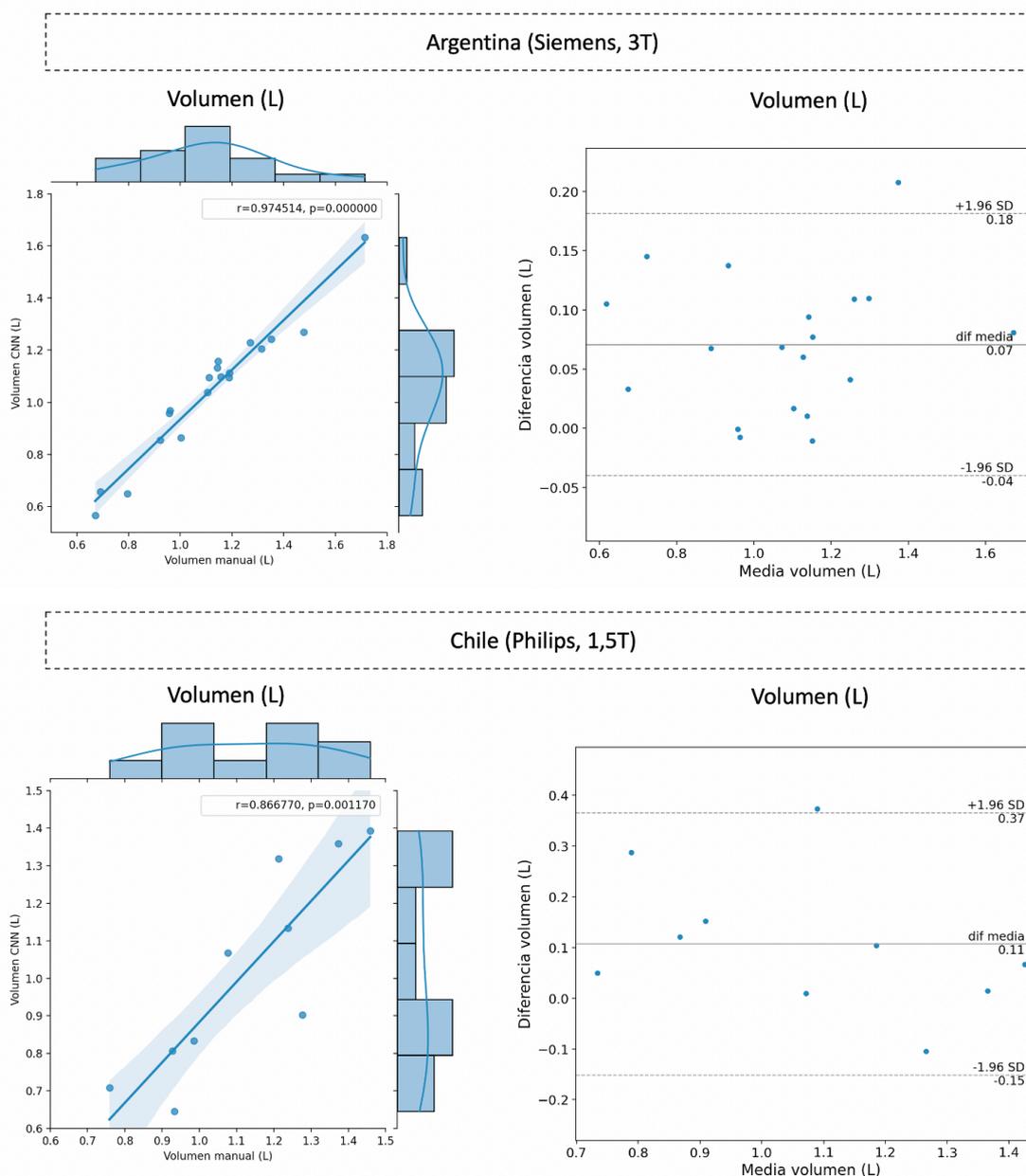


Figura 57. Análisis de correlación entre el volumen de referencia (segmentación manual) y el predicho por la CNN (segmentación automática) del conjunto de datos de test externo correspondiente al centro de Argentina (fila superior) y al centro de Chile (fila inferior). A la izquierda vemos el diagrama de pares entre ambos volúmenes; en los ejes superior y derecho se encuentran los diagramas de los volúmenes de la segmentación manual y predicha, respectivamente. A la derecha se muestra el diagrama Bland-Altman.

Finalmente, se realizó el estudio de correlación del PDFF y $R2^*$ cuantificados sobre la segmentación manual y la obtenida por la CNN para ambos subconjuntos (Figura 58). Mediante la prueba Shapiro-Wilk se obtuvo que, para el conjunto de datos de Argentina, la PDFF seguía una distribución normal ($p>0,05$), mientras que el $R2^*$ no lo hacía ($p<0,05$), por otro lado, para el conjunto de datos de Chile se obtuvo que la PDFF no seguía una distribución normal ($p<0,05$), mientras que el $R2^*$ sí que lo hacía ($p>0,05$). Por lo tanto, para el estudio de la correlación se empleó el coeficiente de Pearson para el

Capítulo 4. Resultados y discusión

análisis del $R2^*$ en el conjunto de datos de Argentina y del PDFFF en el de Chile, mientras que se empleó el coeficiente de Spearman para el PDFFF en los datos de Argentina y el $R2^*$ en los de Chile. Para todos los casos se obtuvo una correlación perfecta entre los valores cuantificados con la máscara de referencia y la obtenida automáticamente por la CNN ($r=1$, $p<0,001$). Por lo tanto, podemos decir, que, aunque la correlación volumétrica no fuera perfecta, los sesgos a nivel de cuantificación se redujeron. Al igual que en todos los casos anteriores, se realizó el análisis Bland-Altman, obteniendo sesgo bajo (Argentina: PDFFF: 0 %, $R2^*$: 0,16 s^{-1} ; Chile: PDFFF: -0,22 %, $R2^*$: -0,23 s^{-1}) y límites de acuerdo estrechos (Argentina: PDFFF: [-0,34 – 0,33] %, $R2^*$: [-1,25 – 1.57] s^{-1} ; Chile: PDFFF: [-0,57 – 0,13] %, $R2^*$: [-0,76 – 0,31] s^{-1}).

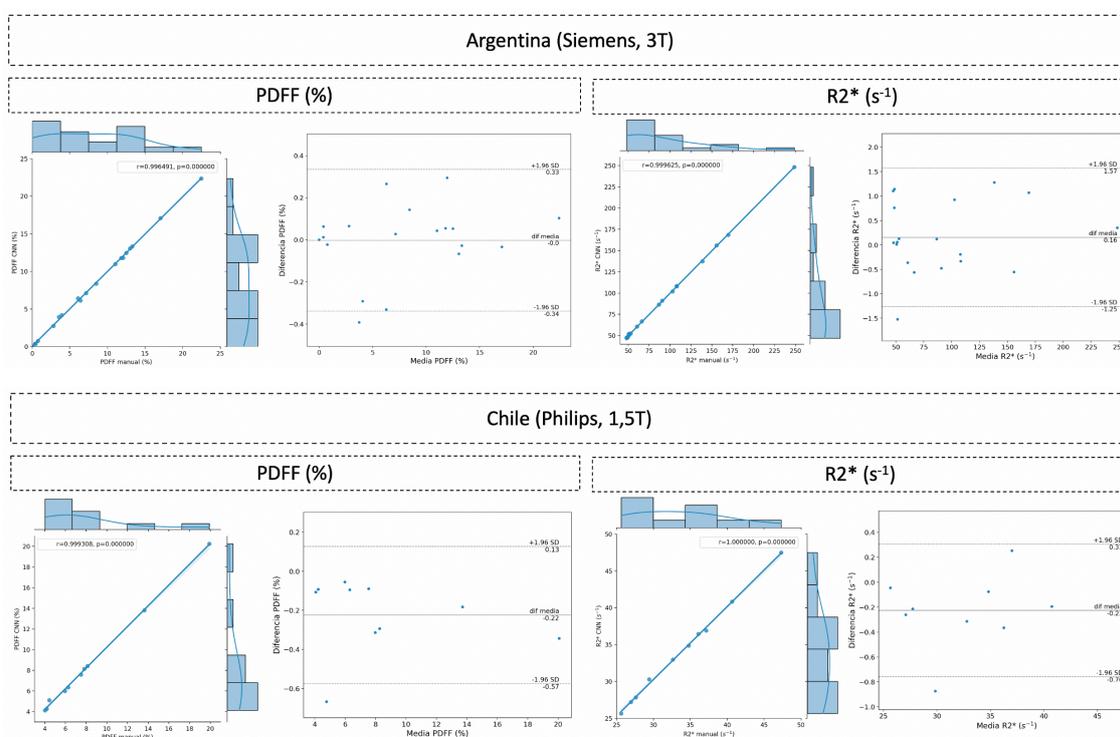


Figura 58. Análisis de correlación entre el PDFFF (izquierda) y $R2^*$ (derecha), en los conjuntos de datos de Argentina (fila superior) y Chile (fila inferior), cuantificados sobre el parénquima hepático empleando la segmentación de referencia (manual) y la predicha por la CNN. Para cada caso, a la izquierda vemos el diagrama de pares entre ambas máscaras; en los ejes superior y derecho se encuentran los diagramas de los biomarcadores cuantificados sobre la segmentación manual y predicha, respectivamente. A la derecha se muestra el diagrama Bland-Altman.

En cuanto al ER, en el conjunto de datos de Argentina, se obtuvo un valor medio \pm desviación de estándar de $3,24 \pm 3,93\%$ para la PDFFF y de $0,80 \pm 0,85\%$ para el $R2^*$, siendo la mediana de 1,99% y 0,53%, respectivamente. Por otro lado, para el conjunto de datos de Chile, la media \pm desviación estándar fue de $3,42 \pm 3,99\%$ para la PDFFF y de $0,86 \pm 0,76\%$ para el $R2^*$ mientras que la mediana tomó valores de 1,96% y 0,63%, respectivamente.

Capítulo 4. Resultados y discusión

En la Figura 59 se muestra el resultado obtenido al segmentar un caso de cada subconjunto de datos del set de validación externa. En ambos casos podemos ver cómo, a pesar de haber ciertas regiones hepáticas sin segmentar (falsos negativos), la mayoría de los vóxeles segmentados pertenecen al parénquima hepático, minimizando el sesgo introducido en la posterior caracterización del tejido hepático.

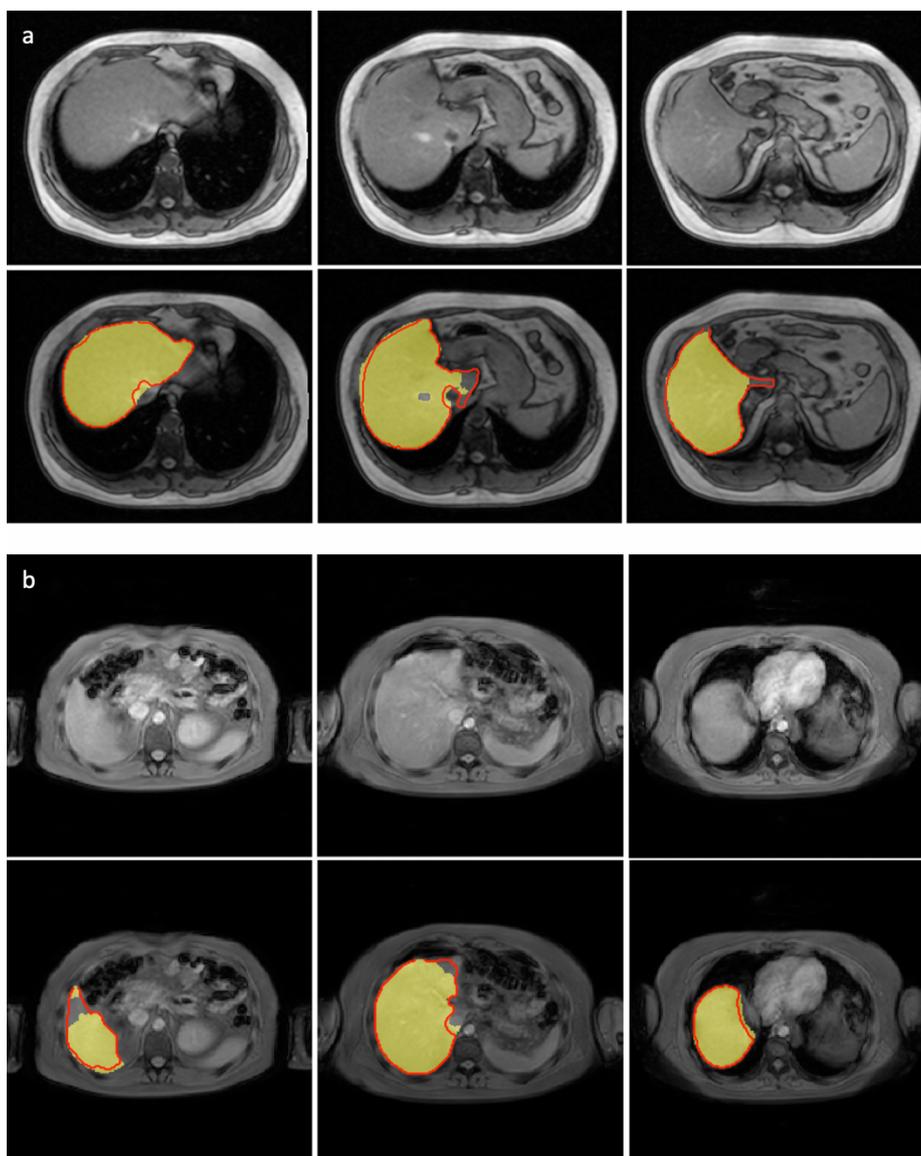


Figura 59. Segmentación obtenida en dos estudios pertenecientes al conjunto de datos de validación externa. En amarillo se representa la máscara predicha por la CNN y en rojo el contorno de la máscara de referencia generada manualmente. a) Estudio perteneciente al centro de Argentina con un escáner Siemens de 3T. b) Estudio perteneciente al centro de Chile con un escáner Philips de 1,5T.

4.1.2.3. Discusión

En la literatura podemos encontrar diferentes soluciones para la segmentación automática del hígado en diferentes secuencias de RM [109-112]. Sin embargo, solo hemos encontrado una que emplee la secuencia MECSE [110]. Este estudio emplea 300 casos para el entrenamiento del modelo y 50 para testear, provenientes de un único escáner de RM (GE 3T), obteniendo un DC de $0,93\pm 0,04$, ligeramente inferior al obtenido en el presente trabajo en el conjunto de datos de test ($n=46$ – DC= $0,94\pm 0,02$). A pesar de que en [110] también se realizó una validación externa ($n=16$) con escáneres de RM diferentes a los empleados para el desarrollo del modelo (Siemens 1,5 y 3T, GE 3T), no se ofrecen dichos resultados por escáner. Reportan un DC medio de $0,92\pm 0,05$, mientras que, en nuestra validación externa, se obtuvo un DC medio de $0,87\pm 0,06$ en el escáner de Philips de 1,5T y de $0,94\pm 0,02$ en el de Siemens de 3T.

Además, estudios previos, se centran en la maximización de las métricas de segmentación, sin centrarse en el uso de la máscara obtenida para la cuantificación de biomarcadores de imagen. El principal objetivo de este trabajo era obtener la máxima precisión posible en la segmentación hepática reduciendo, en la medida de lo posible, áreas con falsos positivos para evitar sesgos en la cuantificación de PDFF y R2*. De este modo, la máscara obtenida debía ser precisa pero conservativa, sin incluir estructuras vecinas como la vesícula biliar o grandes vasos sanguíneos, incluso si esto supone no incluir en la segmentación pequeñas áreas de la periferia hepática. Los resultados obtenidos en este trabajo muestran una elevada correlación y un error relativo bajo al comparar los valores de PDFF y R2* cuantificados a partir de la máscara manual y la predicha por la CNN.

Un estudio presentado en 2018 por Stocker et al. [113], muestra la diferencia obtenida al cuantificar el PDFF y R2* empleando una herramienta automática propia de la segmentación y cuantificación del hígado. Al comparar las gráficas Bland-Altman, en nuestro estudio obtuvimos límites de acuerdo más estrechos tanto en la cuantificación de PDFF como de R2*, mostrando un error menor entre las cuantificaciones obtenidas con la máscara manual y la automática.

Cabe destacar que la reproducibilidad del método desarrollado se estudió con un conjunto de datos adicional, proveniente de escáneres y centros diferentes a los empleados para el entrenamiento y validación del modelo. Se obtuvo un DC inferior ($0,87\pm 0,06$) en el conjunto de imágenes provenientes del escáner de Philips de 1,5T. Esto puede deberse,

por un lado, a la disminución en la relación señal-ruido en imágenes de RM adquiridas con menor campo magnético, o, por otro lado, al hecho de que estas imágenes se adquirieron con una resolución en el plano inferior ($2,3 \times 2,3 \text{ mm}^2$) y un ancho de corte mayor (10-10.5mm), agravando los efectos de volumen parcial y, por tanto, dificultando la diferenciación entre tejidos [114]. Para mejorar la generalización del modelo a este tipo de adquisiciones, una opción podría ser el re-entrenamiento y ajuste de la CNN con imágenes de menor campo magnético y menor resolución. Otra solución podría basarse en el uso de modelos basados en GANs [115], en concreto, modelos basados en la arquitectura Cycle-GAN [116] de forma que podamos trasladar imágenes de un contexto (imágenes de 1,5T) a otro contexto (imágenes de 3T). Sin embargo, en estos casos, al comparar los valores de PDFF y $R2^*$ obtenidos con la máscara manual y la obtenida automáticamente por la CNN, la elevada correlación entre ambos valores y el bajo error relativo en la cuantificación muestran su aplicabilidad en la práctica clínica. Además, la CNN entrenada mostró generalización a otros fabricantes, los resultados obtenidos en las imágenes adquiridas con el escáner de Siemens ($0,94 \pm 0,02$) fueron iguales a los obtenidos con Philips. También es importante resaltar que el entrenamiento, validación y test interno se realizaron con imágenes provenientes de centros de España y Portugal (Europa), mientras que el test externo se realizó con imágenes de centros en Argentina y Chile (América del Sur), demostrando la generalización del método desarrollado a nivel internacional.

4.1.3. Método de cuantificación

Una vez analizados todos los casos con la base de datos descrita en el apartado 3.1.1.1. *Cuantificación y validación clínica* se estudió cuál de todos los algoritmos evaluados es el que proporcionaba menor error en la cuantificación en un mayor número de casos. Para cada caso, se extrajo el error mínimo calculado y se analizó cuantos algoritmos habían obtenido un error igual a ese mínimo ± 1 unidad. Siendo 165 el total de casos analizados, se obtuvieron los resultados de la Tabla 10. En ésta se puede ver cómo, con el algoritmo BOBYQA se alcanza el error mínimo (± 1) en todos los casos, seguido muy de cerca del algoritmo NEWUOA (con límites) en los que no se alcanza ese error mínimo en un único caso. Por el contrario, podemos ver que, para muchos de los algoritmos evaluados, no se alcanza el error mínimo en ninguno de los casos, por lo que podemos decir que éstos no son los más indicados para la resolución de nuestro problema.

Capítulo 4. Resultados y discusión

Tabla 10. Resultados en la evaluación de los diferentes algoritmos que ofrece la librería NLOpt. En un total de 165 casos se ha evaluado en cuantos, cada algoritmo, alcanza el mínimo error (± 1). En verde se ha resaltado aquel que mejores resultados ofrece.

Algoritmo NLOpt	Nº de casos en los que se obtiene el error mínimo (± 1)
COBYLA	0
BOBYQA	165
NEWUOA (con límites)	164
PRAXIS	148
Nelder-Mead Simplex	159
Sbplx	102
MMA	0
CCSAQ	0
SLSQP	0
Low-storage BFGS	0
Preconditioned truncated Newton (con reinicio)	0
Preconditioned truncated Newton (sin reinicio)	0
Truncated Newton	0
Shifted limited-memory variable-metric – VAR1	0
Shifted limited-memory variable-metric – VAR2	0

De este modo, finalmente, se seleccionó el algoritmo BOBYQA para la estimación de los parámetros en la cuantificación de PDFFF y R2*.

Una vez seleccionado el algoritmo, se volvió a ejecutar el mismo en todos los casos, para seleccionar la tolerancia y número de iteraciones óptima. De este modo, de todas aquellas opciones que nos ofrecían los valores mínimos anteriores, se seleccionó aquella que menor carga computacional suponía (menor tolerancia y menor número de iteraciones). En la Tabla 11 se muestran las combinaciones de tolerancia y número de iteraciones en los que se obtiene el menor error ± 1 para los 165 casos. Por lo tanto, para garantizar la menor carga computacional se ha seleccionado una tolerancia de 10^{-5} y 900 iteraciones.

Capítulo 4. Resultados y discusión

Tabla 11. Combinación de valor de tolerancia y número de iteraciones con la que se obtiene el menor error ± 1 en los 165 casos que componen el set de validación.

Tolerancia	Número de iteraciones
10^{-5}	900
10^{-5}	1000
10^{-6}	900
10^{-6}	1000
10^{-7}	900
10^{-7}	1000
10^{-9}	900
10^{-8}	1000
10^{-9}	900
10^{-9}	1000
10^{-10}	900
10^{-10}	1000

4.1.3.1. Paralelización y tiempo de ejecución

En la Tabla 12 se muestran los tiempos de ejecución obtenidos en el análisis de la región hepática de 10 casos diferentes empleando entre 1 y 12 nodos de cálculo. El número de vóxeles analizados en estos 10 casos fueron 66.397 de media (mínimo: 48.141; máximo: 92.302). Se puede observar como el tiempo de cálculo se disminuyó considerablemente cuando se introdujo paralelización (7 minutos menos al emplear 2 nodos de cálculo). Conforme aumenta el número de nodos las diferencias no son tan significativas, siendo únicamente del orden de segundos a partir de 8 nodos.

Tabla 12. Tiempos de ejecución medios, mínimos y máximos obtenidos en el análisis hepático de 10 casos empleando entre 1 y 12 nodos de cálculo para el análisis.

N nodos	T medio (min)	T mínimo (min)	T máximo (min)
1	17,15	11,82	24,09
2	10,42	6,25	14,32
3	8,14	5,39	10,98
4	6,36	3,95	8,60
5	5,37	3,34	7,76
6	4,43	2,78	6,52

N nodos	T medio (min)	T mínimo (min)	T máximo (min)
7	3,86	2,46	5,41
8	3,23	2,06	4,48
9	2,93	1,89	4,24
10	2,70	1,69	3,82
11	2,46	1,55	3,55
12	2,25	1,44	3,08

Así, para la implementación final del módulo en entorno clínico, con el fin de no emplear todos los recursos de la máquina disponible y permitir el análisis de varios casos simultáneamente, se decidió emplear un máximo de 8 nodos.

4.1.4. Validación clínica

Una vez realizadas de manera independiente la validación del método de segmentación y cuantificación se procedió a realizar la validación clínica del algoritmo completo como se ha descrito en el apartado 3.5.1. *Cuantificación PDFF y R2* en secuencias MECSE de RM*. De este modo, se compararon los resultados obtenidos de la cuantificación de la RM con la información provenientes de la biopsia hepática (muestra histopatológica). Estos resultados se publicaron previamente en [90].

En la Tabla 13 encontramos los distintos valores de PDFF y R2* cuantificados en el parénquima hepático, segmentado automáticamente por la CNN, para los distintos grados de esteatosis y sobrecarga de hierro. Estos valores se obtuvieron a partir de los valores de mediana de cada caso independiente. Se puede ver como los valores de PDFF y R2* aumentaban conforme nos encontrábamos en estadios más avanzados. Se analizó si entre estos grupos existían diferencias estadísticamente significativas, para ello, dado que las distribuciones no seguían una distribución normal ($p < 0,05$ tanto para PDFF como R2* haciendo uso del test de normalidad Shapiro-Wilk), se empleó un test de Kruskal-Wallis. Se obtuvo un $p < 0,001$ tanto para el PDFF como para el R2*, por lo tanto, se demostró que existían diferencias estadísticamente significativas entre las medianas de los grupos. En la Tabla 13 podemos observar que para el PDFF no se obtuvo solapamiento entre los grupos si consideramos el rango intercuartílico (p_{25} - p_{75}). Sin embargo, para el R2* encontramos más solapamiento entre los grupos. Esto también puede deberse a la gran diferencia en el número de casos entre grupos, estando los grupos con sobrecarga de hierro infrarrepresentados. Sin embargo, podemos ver que la presencia o no de hierro

Capítulo 4. Resultados y discusión

(Fe0 vs. Fe1-4) tiene mejor diferenciación. Esto también lo podemos apreciar en la Figura 60.

Tabla 13. Valores de PDFF y R2* cuantificados en el parénquima hepático (segmentado por la CNN entrenada) para los distintos grados de esteatosis y sobrecarga de hierro.

Esteatosis – PDFF (%)					
Grado histológico	n	Media ± Std	Mediana	p25	p75
S0	70	6,84 ± 1,38	6,56	5,85	7,66
S1	33	9,95 ± 2,28	10,09	8,17	10,70
S2	33	13,23 ± 2,91	12,74	11,21	14,88
S3	28	18,39 ± 3,87	18,11	15,75	21,50
Sobrecarga de hierro – R2* (s ⁻¹)					
Grado histológico	n	Media ± Std	Mediana	p25	p75
Fe0	129	50,08 ± 11,18	48,75	40,98	55,10
Fe1	10	63,90 ± 14,98	60,81	53,74	69,75
Fe2	8	68,10 ± 16,24	69,24	54,90	73,83
Fe3	11	88,83 ± 45,44	77,19	59,06	103,25
Fe4	6	134,95 ± 82,26	113,31	72,96	162,03

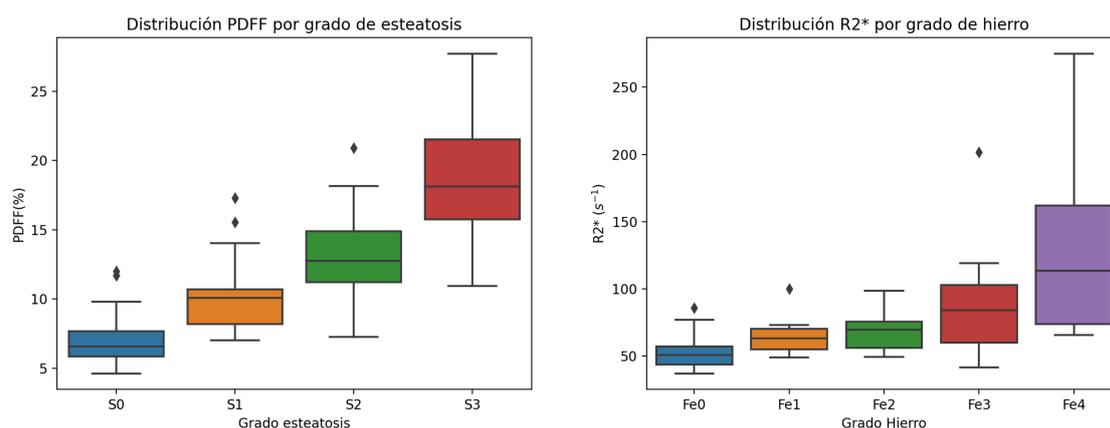


Figura 60. Distribución de PDFF (izquierda) y R2* (derecha) para cada grado de esteatosis y sobrecarga de hierro, respectivamente.

En la Tabla 14 se muestran los umbrales óptimos seleccionados para diferenciar cada uno de los diferentes grados de esteatosis y sobrecarga de hierro. Para esto, se obtuvieron los valores de área bajo la curva ROC (AUC, del inglés *Area Under the Curve*), sensibilidad

Capítulo 4. Resultados y discusión

y especificidad para cada grupo. Se puede ver como el método desarrollado es capaz de detectar esteatosis (S0 vs. S1-3) con un AUC del 97% y de detectar sobrecarga de hierro (Fe0 vs. Fe1-4) con una AUC del 83%.

Tabla 14. Umbrales seleccionados para diferenciar cada uno de los grupos de esteatosis (PDFF) y sobrecarga de hierro (R2*). Para cada uno se muestra el área bajo la curva ROC (AUC) junto con su intervalo de confianza al 95%, la sensibilidad y especificidad en la diferenciación clínica de cada grupo.

Esteatosis				
Grado histológico	Umbral PDFF (%)	AUC (95% IC)	Sensibilidad	Especificidad
S ≥ 1	7,8	0,97 (0,94 – 0,99)	0,95	0,84
S ≥ 2	10,8	0,96 (0,93 – 0,99)	0,88	0,91
S ≥ 3	14,1	0,96 (0,93 – 0,99)	0,89	0,91
Sobrecarga de hierro				
Grado histológico	Umbral R2* (s⁻¹)	AUC (95% IC)	Sensibilidad	Especificidad
Fe ≥ 1	55	0,83 (0,75 – 0,91)	0,80	0,74
Fe ≥ 2	56	0,84 (0,74 – 0,93)	0,84	0,71
Fe ≥ 3	64	0,85 (0,73 – 0,97)	0,83	0,81
Fe ≥ 4	69	0,94 (0,87 – 0,99)	0,86	0,81

Finalmente, se realizó un estudio de correlación entre la PDFF y R2* cuantificados y el porcentaje de grasa y hierro cuantificados, respectivamente, en la imagen de patología digital. Se empleó el coeficiente de Pearson para estudiar dicha correlación, obteniendo $r = 0,89$ ($p < 0,001$) para la grasa y $r = 0,65$ ($p < 0,002$) para el hierro, mostrando una correlación elevada entre el PDFF y el porcentaje de grasa y una correlación moderada entre el R2* y el porcentaje de hierro de la muestra histológica. Esto último puede deberse a la baja prevalencia de sobrecarga de hierro en la muestra empleada para este análisis.

4.1.4.1. Discusión

Los resultados obtenidos en este estudio demuestran que los valores cuantitativos de PDFF obtenidos de la segmentación automática del hígado son precisos para establecer el grado de esteatosis (AUC = 0,97) y muestran una elevada correlación con la fracción grasa obtenida de la muestra digital histológica ($r = 0,89$). En cuanto al R2*, los resultados

muestran buena capacidad de detección de hierro hepático ($AUC = 0,83$), mostrando una correlación moderada ($r = 0,65$) con la muestra histológica.

El método de segmentación automática se desarrolló con una población que mostraba sospecha de enfermedad hepática, pero valores normales de PDFF y $R2^*$ (mediana [rango] para el PDFF de 5,6 [0 – 28,4] % y de 47,77 [28,6 – 200,2] s^{-1} para el $R2^*$). Sin embargo, en la validación clínica, se empleó una población bien caracterizada formada por pacientes con enfermedad hepática crónica con valores más elevados de afectación hepática (PDFF medio: 11% y $R2^*$ medio: 60 s^{-1}). Por lo tanto, podemos demostrar la generalización del método de segmentación automática a una población con enfermedad hepática incluyendo un amplio espectro de lesiones hepáticas evaluadas a partir de la biopsia.

Para ser marcadores diagnósticos útiles, los parámetros derivados de la RM deben demostrar una fuerte correlación con el estándar de referencia histológico actualmente aceptado [117]. La correlación entre las características histológicas y la PDFF o el $R2^*$ obtenida de regiones de interés posicionadas de manera manual ya se ha demostrado en estudios anteriores [38,40,118,119]. Como valor añadido, este estudio confirma el rendimiento del método de segmentación hepática automática propuesto, demostrando correlación entre los valores cuantitativos y los valores obtenidos de la muestra histopatológica, sin embargo, con valores medios más elevados. Estos valores más altos pueden deberse a que se realiza un muestreo más representativo que incluye el hígado periférico subcapsular y áreas del ligamento falciforme y el hilio, que suelen tener mayores depósitos de grasa.

Por otro lado, como se ha comentado anteriormente, este estudio confirmó una alta correlación entre la PDFF y el área proporcional de grasa. En cambio, sólo se observó una correlación moderada para la correlación $R2^*$ y el área proporcional al hierro. Este hallazgo podría deberse a la escasa prevalencia y al bajo espectro de la sobrecarga de hierro y al menor rendimiento de $R2^*$ en los pacientes con EHGNA, atribuible a la influencia de la grasa [120]. El análisis digital patológico y el aprendizaje profundo de la segmentación de órganos por RM ofrecen un método robusto para evaluar los marcadores hepáticos. La segmentación automatizada reduce el tiempo de interpretación del radiólogo en la evaluación de grasa y hierro hepáticos.

4.2. Segmentación de grasa abdominal en imágenes de TC

En esta sección vamos a analizar los resultados obtenidos en el modelo entrenado para la segmentación automática de grasa abdominal en imágenes de TC. En primer lugar, vamos a evaluar el clasificador encargado de detectar los cortes transversales del TC pertenecientes a la región abdominal y que serán los que, posteriormente, se empleen para segmentar la grasa visceral y la grasa subcutánea cuya evaluación de los resultados también se incluye en este apartado.

4.2.1. Detección de la región abdominal

En este apartado se exponen los resultados del clasificador de imágenes de TC para la detección y diferenciación de los cortes pertenecientes a la región abdominal frente a los que se encuentran fuera de ella.

4.2.1.1. Validación del modelo de clasificación

En primer lugar, se hizo un ajuste de la tasa de aprendizaje de la CNN para validar, de manera interna, el correcto rendimiento de la arquitectura diseñada.

Inicialmente se evaluó el uso de una tasa de aprendizaje constante durante el proceso de entrenamiento. En la Figura 61 y en la Tabla 15 se encuentran los resultados obtenidos durante el proceso de validación cruzada para los distintos valores de tasa de aprendizaje evaluados. Vemos como con tasas de aprendizaje elevadas (1 y 0,1) el modelo no es capaz de converger, ofreciendo una precisión del 50% generando, como salida, siempre la misma clase. Por el contrario, con tasas de aprendizaje bajas ($<10^{-4}$) se puede ver como el modelo empezó a ajustarse y a generalizar sobre el conjunto de datos de validación, sin embargo, el proceso de entrenamiento es más lento y no llega a converger. Finalmente, con una tasa de aprendizaje intermedia se alcanzó la convergencia de los modelos, obteniéndose los mejores resultados con una tasa de aprendizaje de 10^{-3} (ID_{TA} : 4). Con esta tasa de aprendizaje se obtuvo una precisión media a lo largo de las 5 iteraciones del proceso de validación cruzada de 99,28% (Min: 98,80% - Máx: 99,58%).

Capítulo 4. Resultados y discusión

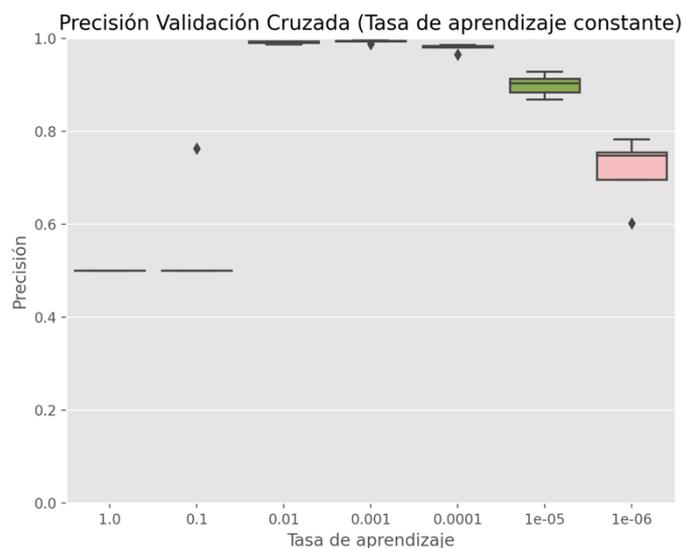


Figura 61. Clasificador de imágenes abdominales de TC: valores de precisión obtenidos durante el proceso de validación cruzada sobre el conjunto de datos de validación al emplear distintos valores de tasas de aprendizaje constante.

Tabla 15. Clasificador de imágenes abdominales de TC: valores de mediana, media, desviación estándar (Std), mínimo y máximo de la precisión obtenida a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos valores de tasa de aprendizaje constante. En negrita se resalta la tasa de aprendizaje constante que mejores resultados ofreció.

ID _{TA}	TA	Mediana	Media±Std	Mínimo	Máximo
1	1	0,50	0,50±0,00	0,50	0,50
2	10 ⁻¹	0,50	0,55±0,11	0,50	0,76
3	10 ⁻²	0,99	0,99±0,00	0,99	0,99
4	10⁻³	0,99	0,99±0,00	0,99	1,00
5	10 ⁻⁴	0,98	0,98±0,01	0,97	0,99
6	10 ⁻⁵	0,90	0,90±0,02	0,87	0,93
7	10 ⁻⁶	0,75	0,72±0,06	0,60	0,78

A continuación, se evaluó el rendimiento del modelo empleando una tasa de aprendizaje variable según un planificador. En la Figura 62 podemos observar la precisión obtenida durante el proceso de validación cruzada para cada uno de los planificadores evaluados. Se puede ver como algunos de ellos no llegan a ofrecer convergencia de la red (precisión de 0,5), mientras que otros dan muy buenos resultados (precisión alta y estable a lo largo de todas las iteraciones). En la Tabla 16 se muestran las métricas obtenidas para los planificadores que mostraron convergencia en el entrenamiento. El planificador que mejores resultados ofreció fue el lineal con una tasa de aprendizaje inicial de 10⁻² (ID_{TA}: 13) con una precisión media de 99,42% (Min: 99,10% - Máx: 99,56%) seguido muy de

Capítulo 4. Resultados y discusión

cerca por el planificador por saltos con una tasa de aprendizaje inicial de 10^{-2} , $\Delta_{\text{épocas}}$: 25 y tasa: 0,25 (ID_{TA}: 46).

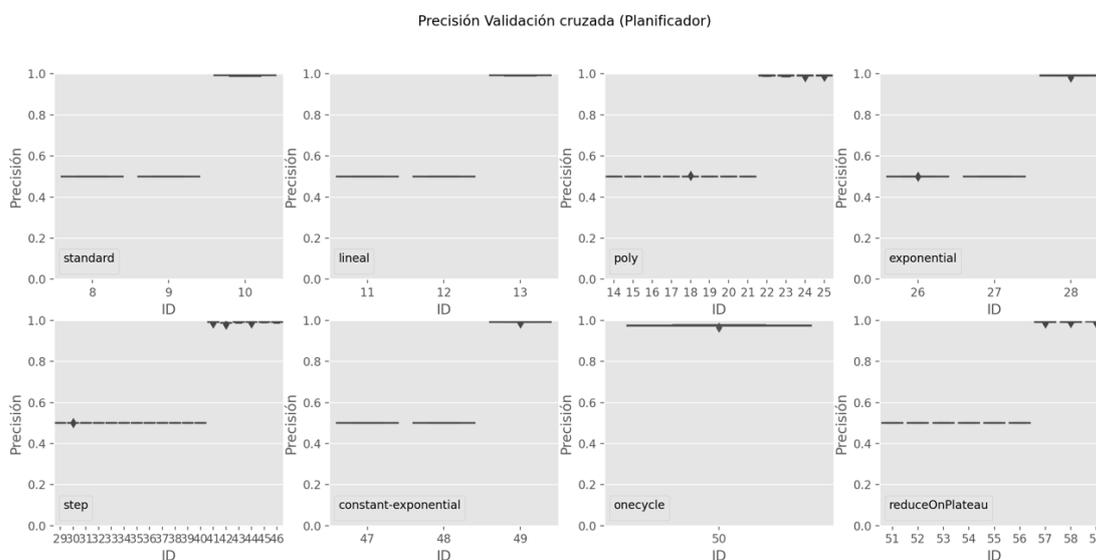


Figura 62. Modelo de clasificación de imágenes abdominales de TC: precisión obtenida a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos planificadores de tasa de aprendizaje con diferentes configuraciones de estos. El eje x representa el identificador del modelo entrenado.

Tabla 16. Clasificador de imágenes abdominales de TC: valores de mediana, media, desviación estándar (Std), mínimo y máximo de la precisión obtenida a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos planificadores de tasa de aprendizaje. Solo se muestran aquellos en los que hubo una convergencia en el entrenamiento. En negrita se resalta la tasa de aprendizaje constante que mejores resultados ofreció.

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo
10	Basado en el tiempo	10^{-2}	-	0,99	0,99±0,00	0,99	1,00
13	Lineal	10^{-2}	-	1,00	0,99±0,01	0,99	1,00
22	Polinomial	10^{-2}	Factor: 2	0,99	0,99±0,00	0,99	1,00
23			Factor: 3	0,99	0,99±0,00	0,99	1,00
24			Factor: 5	0,99	0,99±0,00	0,99	0,99
25			Factor: 7	0,99	0,99±0,00	0,99	0,99
28	Exponencial	10^{-2}	k: 0,1	0,99	0,99±0,00	0,98	1,00
41	Por saltos	10^{-2}	$\Delta_{\text{épocas}}$: 5 Tasa: 0,5	0,99	0,99±0,00	0,99	0,99
42			$\Delta_{\text{épocas}}$: 5 Tasa: 0,25	0,99	0,99±0,00	0,98	0,99
43			$\Delta_{\text{épocas}}$: 10 Tasa: 0,5	0,99	0,99±0,00	0,99	1,00
44			$\Delta_{\text{épocas}}$: 10	0,99	0,99±0,00	0,99	0,99

Capítulo 4. Resultados y discusión

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo
			Tasa: 0,25				
45			$\Delta_{\text{époc}}$: 25 Tasa: 0,5	0,99	0,99±0,00	0,99	1,00
46			$\Delta_{\text{époc}}$: 25 Tasa: 0,25	1,00	0,99±0,00	0,99	1,00
49	Constante y exponencial	10 ⁻²	$\Delta_{\text{époc}}$: 10 k: 0,1	0,99	0,99±0,00	0,99	1,00
50	De un ciclo	-	TA _{ini} :10 ⁻⁴ TA _{min} :10 ⁻⁵ TA _{max} :10 ⁻² $\Delta_{\text{époc},1}$:10 $\Delta_{\text{époc},2}$:10 k:0,8	0,97	0,97±0,00	0,97	0,98
57	Dinámica	10 ⁻²	$\Delta_{\text{époc}}$: 5	0,99	0,99±0,00	0,99	0,90
58			$\Delta_{\text{époc}}$: 10	0,99	0,99±0,00	0,99	1,00
59			$\Delta_{\text{époc}}$: 15	0,99	0,99±0,00	0,99	1,00

Finalmente, se evaluó el uso de un optimizador. En la Figura 63 se puede observar la precisión obtenida con cada uno de los optimizadores evaluados. De los planificadores evaluados, podemos ver que sólo converge con uno de ellos (Adadelta – ID_{TA}: 63), el cual ofrece una precisión media de 99,18% (Min: 99,03% - Máx: 99,59%).

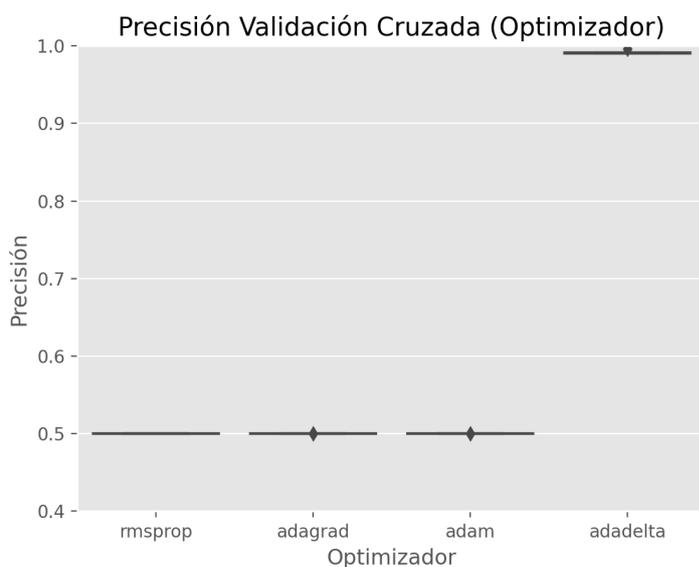


Figura 63. Modelo de clasificación de imágenes abdominales de TC: Precisión en el conjunto de datos de validación al emplear distintos optimizadores de la tasa de aprendizaje.

En todas aquellas configuraciones de tasa de aprendizaje donde no se consigue la convergencia de la red, si se hubieran evaluado otros valores de tasa de aprendizaje inicial, probablemente se hubiera logrado la convergencia, sin embargo, como ya se obtuvieron muy buenos resultados con aquellos modelos en los que sí se obtuvo convergencia ya no se probaron más configuraciones.

Finalmente, se seleccionó el uso de una tasa de aprendizaje variable según un planificador lineal (ID_{TA} : 13) ya que es el que mejor precisión ofreció durante el proceso de validación cruzada (Precisión media: 99,42%).

4.2.1.2. Test del modelo de clasificación

Para la evaluación de la capacidad de generalización del modelo desarrollado contábamos con 12 estudios de TC que suponían un total de 2.997 imágenes 2D (694 pertenecientes a la región abdominal y 2.303 no pertenecientes a la región abdominal).

Las métricas se evaluaron de manera independiente para cada estudio de TC considerando todos los cortes que forman la serie dada y, posteriormente, se calculó la media, desviación estándar y mediana del conjunto de test.

En primer lugar, se evaluó el valor del umbral óptimo para la diferenciación entre ambas regiones. Se evaluaron los valores 0,5 – 0,6 – 0,7 ya que para valores inferiores a 0,5 se observaron muchos falsos positivos y para valores superiores a 0,7 nos encontramos con un caso en el que no había ningún corte con una probabilidad superior a dicho umbral. En la Tabla 17 se muestran los valores de media y desviación estándar para cada una de las métricas definidas para la evaluación del algoritmo de clasificación considerando los tres valores de umbral seleccionados. En este caso lo que nos interesa es maximizar la especificidad, es decir, maximizar la correcta clasificación de cortes que no pertenecen a la región abdominal, para garantizar que no dejamos de segmentar cortes de interés. Esto se consigue tanto con el umbral de 0,6 como con el umbral de 0,7. Entre estos dos se ha seleccionado un umbral de 0,6 ya que ofrece una mayor sensibilidad y, por tanto, un mayor porcentaje de cortes correctamente clasificados como pertenecientes a la región abdominal.

Capítulo 4. Resultados y discusión

Tabla 17. Valores de media y desviación estándar para cada una de las métricas calculadas en la evaluación de la clasificación y localización de la región abdominal en imágenes de TC considerando diferentes valores de umbral (0,5 – 0,6 – 0,7) a la hora de diferenciar una clase de otra.

Umbral	Precisión	Exactitud	Sensib.	Especif.	F1 Score	AUC
0,5	0,91±0,09	0,86±0,17	0,93±0,11	0,90±0,16	0,88±0,10	0,91±0,08
0,6	0,94±0,04	0,91±0,12	0,89±0,13	0,96±0,06	0,89±0,08	0,92±0,06
0,7	0,94±0,04	0,92±0,12	0,87±0,17	0,96±0,05	0,88±0,12	0,92±0,08

Para todas las métricas del modelo seleccionado (umbral de 0,6) se ha obtenido una mediana superior al 90% (Tabla 18), lo cual supone un buen rendimiento de clasificación. Esto es de gran relevancia ya que esta clasificación define la región a segmentar posteriormente y, por lo tanto, un error en este paso puede conllevar errores en la segmentación de la grasa abdominal y la cuantificación de sus volúmenes. Realizando un análisis exploratorio de los resultados, se observó que, principalmente, los errores en clasificación se producían en las imágenes vecinas de los cortes que delimitaban la región abdominal, sin embargo, estos cortes son los que menor sesgo introducen en la posterior segmentación de la grasa abdominal. Por este motivo, se definió un intervalo de confianza de ± 5 mm, así, los cortes hasta 5 mm por encima y por debajo del corte inicial y final, respectivamente, que definen la región abdominal no se consideraron para calcular las métricas. La Tabla 18 muestra los resultados tras la exclusión de las imágenes pertenecientes al intervalo de confianza en comparación con las métricas obtenidas sin considerar dicho intervalo. Para todas las métricas podemos observar un aumento en los valores obtenidos, siendo, todas las medianas, superiores al 95%.

Tabla 18. Valores de mediana, media y desviación estándar para cada una de las métricas calculadas en la evaluación de la clasificación y localización de la región abdominal en imágenes de TC. Sin considerar (izquierda) y considerando (derecha) un intervalo de confianza de ± 5 mm alrededor de los límites obtenidos.

Métrica	Original		Con intervalo de confianza	
	Mediana	Media±Std	Mediana	Media±Std
Precisión	0,95	0,94±0,04	0,99	0,96±0,04
Exactitud	0,95	0,91±0,12	0,99	0,94±0,13
Sensib.	0,91	0,89±0,13	0,98	0,94±0,12
Especif.	0,98	0,96±0,06	1,00	0,97±0,06
F1 Score	0,92	0,89±0,08	0,97	0,92±0,09
AUC	0,93	0,92±0,08	0,98	0,95±0,06

En la Figura 64 se muestra el resultado de un caso de conjunto de datos de test en el que se obtuvo una precisión de 0,95 sin considerar el intervalo de confianza y 0,99 considerando el intervalo de confianza. En ésta se muestra el resultado medio obtenido por los cinco modelos que forman el ensamblado y la clasificación final tras el postprocesado para la definición final de la región abdominal comparado con la región de referencia definida manualmente.

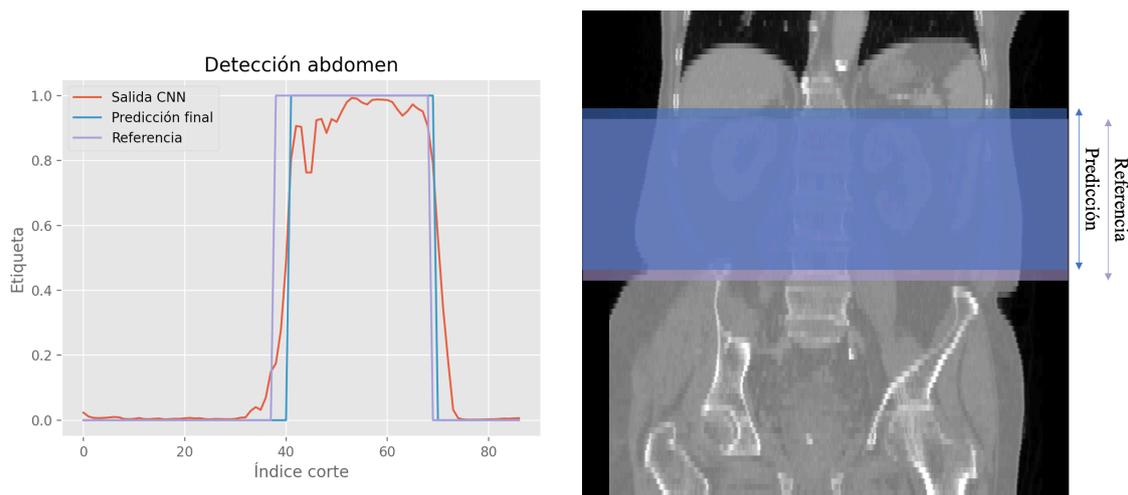


Figura 64. Resultado de un caso en la definición de la región abdominal. A la izquierda se representa una gráfica con los resultados predichos para cada uno de los cortes comparados con la región de referencia. Así, encontramos la probabilidad media para cada corte obtenida por los cinco modelos que forman el ensamblado final (línea roja); la predicción final tras el postprocesado (línea azul), donde 0 corresponde a los cortes que no pertenecen a la región abdominal y 1 a aquellos que sí pertenecen a la región abdominal; y la referencia definida de forma manual (línea morada). A la derecha vemos la correspondencia espacial de la región abdominal predicha (azul) y la de referencia (morado).

4.2.2. Segmentación de grasa visceral y subcutánea

En este apartado se exponen los resultados del modelo de segmentación automática de grasa visceral y subcutánea en imágenes de TC una vez definida la región a segmentar con el modelo de clasificación anterior.

4.2.2.1. Validación del modelo de segmentación

En primer lugar, se van a exponer los resultados obtenidos en el proceso de validación y ajuste de hiperparámetros.

4.2.2.1.1. Tasa de aprendizaje

A continuación, se realizó la evaluación del método de segmentación de grasa visceral y subcutánea en los cortes de la región abdominal, concretamente se evaluó el rendimiento de los diferentes modelos entrenados con las distintas variaciones de tasa de aprendizaje. Así, en primer lugar, se evaluó el uso de una tasa de aprendizaje constante. Para ello, se calculó el DC obtenido a lo largo de las cinco iteraciones del proceso de validación cruzada. Este DC tiene en cuenta tanto la grasa visceral como la subcutánea. La Figura 65 muestra la distribución del DC obtenida para cada una de las tasas de aprendizaje constantes evaluadas, mientras que la Tabla 19 muestra los valores concretos obtenidos en cada experimento. Se puede observar que para una tasa de aprendizaje constante de 1 se obtuvo poca generalización de los modelos ya que el rendimiento cambiaba considerablemente en función de los conjuntos de imágenes empleados para entrenar y validar. Con tasas de aprendizaje bajas ($<10^{-2}$) vemos que el modelo no llega a converger a valores de DC elevados como los obtenidos con tasas de aprendizaje de 0,1 o 0,01. Los mejores resultados se obtuvieron para una tasa de aprendizaje constante de 0,1 (ID_{TA}: 2), con un DC medio de 96,43% (DC min: 94,86% y DC máx: 97,48%)

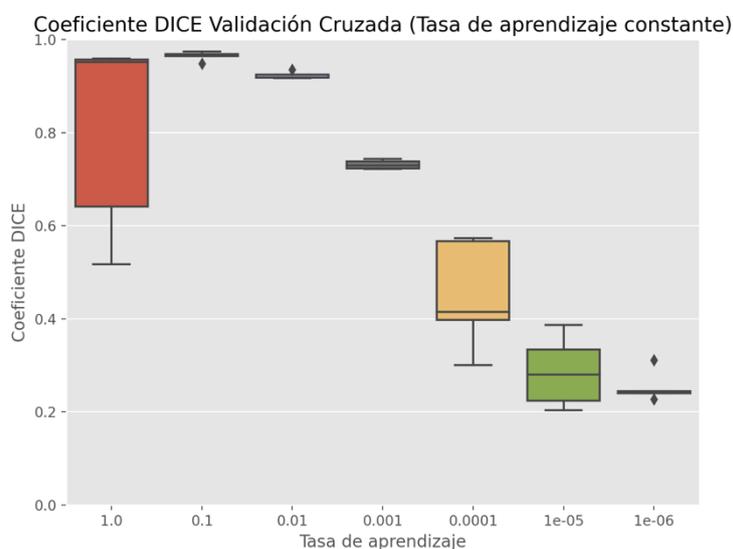


Figura 65. Modelo de segmentación de grasa visceral y subcutánea en imágenes de TC: distribución del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos valores de tasa de aprendizaje constante.

Capítulo 4. Resultados y discusión

Tabla 19. Modelo de segmentación de grasa visceral y subcutánea en imágenes de TC: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos valores de tasa de aprendizaje constante. En negrita se resalta la tasa de aprendizaje constante que mejores resultados ofreció.

ID _{TA}	TA	Mediana	Media±Std	Mínimo	Máximo
1	1	0,95	0,81±0,19	0,52	0,96
2	10⁻¹	0,96	0,96±0,01	0,95	0,97
3	10 ⁻²	0,92	0,92±0,01	0,92	0,94
4	10 ⁻³	0,73	0,73±0,01	0,72	0,74
5	10 ⁻⁴	0,41	0,45±0,11	0,30	0,57
6	10 ⁻⁵	0,28	0,29±0,07	0,20	0,39
7	10 ⁻⁶	0,24	0,25±0,03	0,23	0,32

Seguidamente se evaluó el rendimiento del modelo de segmentación automática al emplear diferentes planificadores de tasa de aprendizaje. En la Figura 66 se muestran las distribuciones de los DC obtenidos en el proceso de validación cruzada para cada uno de los planificadores evaluados, en esta figura las gráficas están agrupadas en función del tipo de planificador empleado. Por otro lado, la Tabla 20 muestra los valores concretos obtenidos para cada planificador. Se puede observar cómo, en general, los planificadores inicializados con una tasa de aprendizaje inicial de 1, la estabilidad entre las diferentes iteraciones del proceso de validación cruzada era baja, por lo que la generalización del modelo era baja. Sin embargo, en gran parte de los planificadores se obtuvo una gran estabilidad. Los mejores resultados se obtuvieron empleando un planificador dinámico (Δ_{epocas} : 15) con una tasa de aprendizaje inicial de 0,1 (ID_{TA}: 56), con la que se obtuvo una DC medio de 96,32% (DC min: 96,09%; DC máx: 96,51%).

En comparación con el mejor modelo empleando una tasa de aprendizaje constante, los DC medios obtenidos son similares, sin embargo, se observa una mayor estabilidad (menor diferencia entre el valor mínimo y máximo) al emplear el planificador.

Capítulo 4. Resultados y discusión

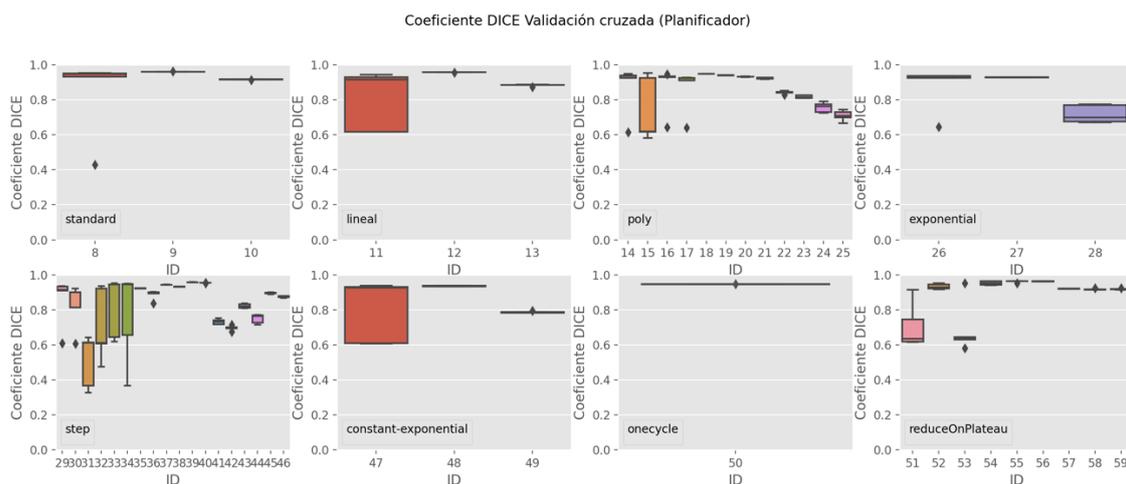


Figura 66. Modelo segmentación de grasa visceral y subcutánea en imágenes de TC: Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos planificadores de tasa de aprendizaje con distintas configuraciones de estos. El eje x representa el identificador del modelo entrenado.

Tabla 20. Modelo de segmentación de grasa visceral y subcutánea en imágenes de TC: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos planificadores de tasa de aprendizaje. En negrita se resalta el planificador que mejores resultados ofreció, además, este fue el que finalmente se seleccionó para el entrenamiento del modelo final.

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo
8	Basado en el tiempo	1	-	0,95	0,84±0,21	0,43	0,95
9		10 ⁻¹	-	0,96	0,96±0,00	0,96	0,96
10		10 ⁻²	-	0,92	0,92±0,00	0,91	0,92
11	Lineal	1	-	0,92	0,80±0,15	0,62	0,94
12		10 ⁻¹	-	0,96	0,96±0,00	0,96	0,96
13		10 ⁻²	-	0,89	0,88±0,01	0,87	0,89
14	Polinomial	1	Factor: 2	0,94	0,87±0,13	0,62	0,95
15			Factor: 3	0,63	0,74±0,16	0,58	0,95
16			Factor: 5	0,93	0,88±0,12	0,64	0,94
17			Factor: 7	0,92	0,87±0,11	0,64	0,93
18		10 ⁻¹	Factor: 2	0,95	0,95±0,00	0,95	0,95
19			Factor: 3	0,94	0,94±0,00	0,94	0,94
20			Factor: 5	0,93	0,93±0,00	0,93	0,93
21			Factor: 7	0,93	0,92±0,00	0,92	0,93
22		10 ⁻²	Factor: 2	0,85	0,84±0,01	0,83	0,85
23			Factor: 3	0,81	0,82±0,01	0,81	0,83
24			Factor: 5	0,76	0,76±0,02	0,73	0,79
25			Factor: 7	0,71	0,71±0,03	0,67	0,74
26	Exponencial	1	k: 0,1	0,93	0,87±0,12	0,64	0,94

Capítulo 4. Resultados y discusión

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo
27		10 ⁻¹		0,93	0,93±0,00	0,93	0,93
28		10 ⁻²		0,70	0,72±0,05	0,67	0,78
29	Por saltos	1	$\Delta_{\text{époc}}: 5$ Tasa: 0,5	0,91	0,86±0,13	0,61	0,94
30			$\Delta_{\text{époc}}: 5$ Tasa: 0,25	0,90	0,83±0,12	0,61	0,92
31			$\Delta_{\text{époc}}: 10$ Tasa: 0,5	0,61	0,51±0,14	0,33	0,64
32			$\Delta_{\text{époc}}: 10$ Tasa: 0,25	0,61	0,71±0,18	0,48	0,94
33			$\Delta_{\text{époc}}: 25$ Tasa: 0,5	0,94	0,82±0,15	0,62	0,95
34			$\Delta_{\text{époc}}: 25$ Tasa: 0,25	0,95	0,77±0,23	0,37	0,95
35		10 ⁻¹	$\Delta_{\text{époc}}: 5$ Tasa: 0,5	0,92	0,92±0,00	0,92	0,92
36			$\Delta_{\text{époc}}: 5$ Tasa: 0,25	0,90	0,89±0,02	0,84	0,90
37			$\Delta_{\text{époc}}: 10$ Tasa: 0,5	0,94	0,94±0,00	0,94	0,94
38			$\Delta_{\text{époc}}: 10$ Tasa: 0,25	0,93	0,93±0,00	0,93	0,93
39			$\Delta_{\text{époc}}: 25$ Tasa: 0,5	0,96	0,96±0,00	0,96	0,96
40			$\Delta_{\text{époc}}: 25$ Tasa: 0,25	0,95	0,95±0,00	0,95	0,95
41		10 ⁻²	$\Delta_{\text{époc}}: 5$ Tasa: 0,5	0,73	0,73±0,01	0,72	0,75
42			$\Delta_{\text{époc}}: 5$ Tasa: 0,25	0,70	0,70±0,01	0,68	0,71
43			$\Delta_{\text{époc}}: 10$ Tasa: 0,5	0,82	0,82±0,01	0,81	0,84
44			$\Delta_{\text{époc}}: 10$ Tasa: 0,25	0,76	0,75±0,02	0,71	0,77

Capítulo 4. Resultados y discusión

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo
45			$\Delta_{\acute{e}poc}$: 25 Tasa: 0,5	0,89	0,89±0,00	0,89	0,90
46			$\Delta_{\acute{e}poc}$: 25 Tasa: 0,25	0,87	0,87±0,00	0,87	0,88
47	Constante y exponencial	1	$\Delta_{\acute{e}poc}$: 10 k: 0,1	0,93	0,80±0,16	0,61	0,94
48		10^{-1}		0,94	0,94±0,00	0,93	0,94
49		10^{-2}		0,78	0,79±0,00	0,78	0,79
50	De un ciclo	-	TA _{ini} : 10^{-4} TA _{min} : 10^{-5} TA _{max} : 10^{-2} $\Delta_{\acute{e}poc,1}$:10 $\Delta_{\acute{e}poc,2}$:10 k:0,8	0,95	0,95±0,00	0,95	0,95
51	Dinámica	1	$\Delta_{\acute{e}poc}$: 5	0,63	0,71±0,11	0,62	0,91
52			$\Delta_{\acute{e}poc}$: 10	0,93	0,93±0,01	0,92	0,95
53			$\Delta_{\acute{e}poc}$: 15	0,63	0,69±0,13	0,58	0,95
54		10^{-1}	$\Delta_{\acute{e}poc}$: 5	0,95	0,95±0,01	0,94	0,96
55			$\Delta_{\acute{e}poc}$: 10	0,96	0,96±0,00	0,95	0,96
56			$\Delta_{\acute{e}poc}$: 15	0,96	0,96±0,00	0,96	0,97
57		10^{-2}	$\Delta_{\acute{e}poc}$: 5	0,92	0,92±0,00	0,92	0,92
58			$\Delta_{\acute{e}poc}$: 10	0,92	0,92±0,00	0,91	0,92
59			$\Delta_{\acute{e}poc}$: 15	0,92	0,92±0,00	0,92	0,92

Finalmente, se evaluó el uso de un optimizador. De igual modo que para la tasa de aprendizaje constante y el planificador, se evaluó el DC obtenido a lo largo del proceso de validación cruzada. La Figura 67 muestra la distribución de DC obtenida para cada uno de los optimizadores evaluados mientras que la Tabla 21 muestra las métricas obtenidas en cada caso. Los mejores resultados se obtuvieron al emplear un optimizador Adam, con un DC medio de 95,43% (DC min: 94,01%; DC máx: 96,80%).

Capítulo 4. Resultados y discusión

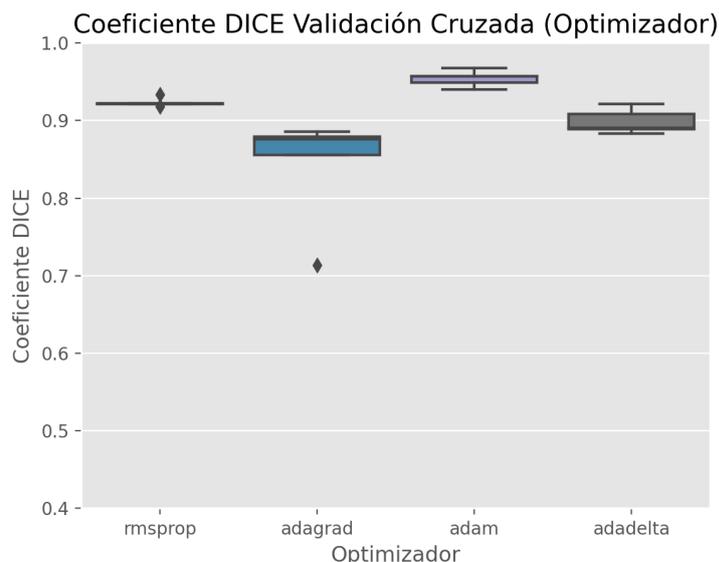


Figura 67. Modelo de segmentación de grasa visceral y subcutánea en imágenes de TC: Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos optimizadores de la tasa de aprendizaje.

Tabla 21. Modelo de segmentación de grasa visceral y subcutánea en imágenes de TC: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos optimizadores de tasa de aprendizaje. En negrita se resalta el optimizador que mejores resultados ofreció.

ID _{TA}	Optimizador	Mediana	Media±Std	Mínimo	Máximo
60	RMSprop	0,92	0,92±0,01	0,92	0,93
61	Adagrad	0,88	0,84±0,06	0,71	0,89
62	Adam	0,96	0,95±0,01	0,94	0,97
63	Adadelta	0,89	0,90±0,01	0,88	0,92

Por lo tanto, finalmente, se seleccionó una tasa de aprendizaje basada en un planificador dinámico con una tasa de aprendizaje inicial de 0,1 y una disminución de ésta un factor 0,1 cuando las pérdidas en validación se mantienen constantes 15 épocas (ID_{TA}: 56). Como se ha comentado anteriormente, con esta configuración, se obtuvo un DC medio de 96,32% (DC min: 96,09%; DC máx: 96,51%).

4.2.2.1.2. Función de pérdidas

Una vez definida la tasa de aprendizaje, se procedió a evaluar las diferentes funciones de pérdidas. Cabe recordar que para el cálculo de la función de pérdidas se empleó el valor medio de aplicar dicha función a cada una de las tres regiones de interés que forman la máscara (fondo, grasa visceral y grasa subcutánea). En la Figura 68 se muestra la distribución del DC obtenido a lo largo de las cinco iteraciones del proceso de validación

Capítulo 4. Resultados y discusión

cruzada para cada una de las funciones de pérdidas. Además, la Tabla 22 resume las métricas obtenidas en cada caso. Se puede observar que los peores resultados se produjeron al emplear la entropía cruzada como función de pérdidas (ID_{FP}: 1), esto puede deberse a que ésta es una función de pérdidas que no tiene en cuenta el solapamiento entre la máscara real y la predicha como en el resto de las funciones de pérdidas. Finalmente, tanto por ser el máximo valor de DC medio obtenido como por la estabilidad entre las iteraciones del proceso de validación cruzada, los mejores resultados se obtuvieron al emplear una función de pérdidas Focal Tversky con $\alpha = 0,5$ (ID_{FP}: 11) con la que se obtuvo un DC medio de 96,33% (DC min: 95,77%; DC máx: 96,80%).

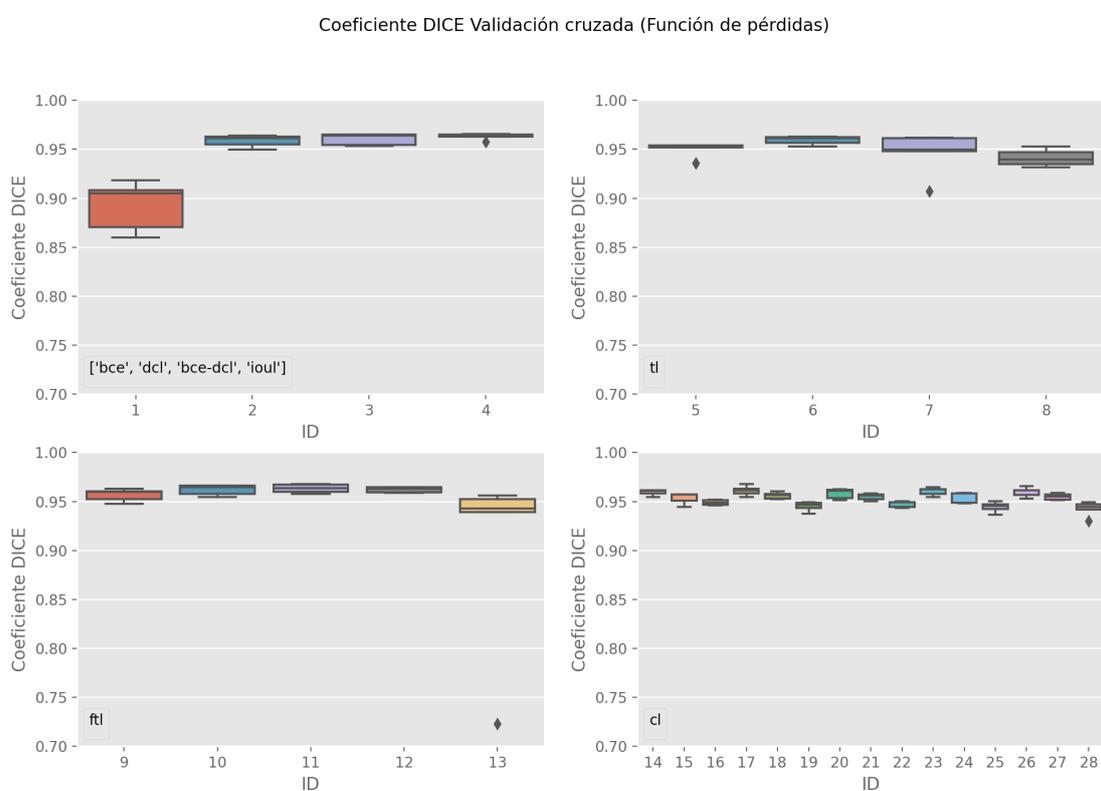


Figura 68. Modelo de segmentación de grasa visceral y subcutánea en imágenes de TC: Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintas funciones de pérdidas. El eje x representa el identificador del modelo entrenado.

Tabla 22. Modelo de segmentación de grasa visceral y subcutánea en imágenes de TC: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintas funciones de pérdidas. En negrita se resalta la función de pérdidas que mejores resultados ofreció y que se empleó para el entrenamiento del modelo final.

ID _{FP}	Función de pérdidas	Otros parámetros	Mediana	Media±Std	Mínimo	Máximo
1	BCE	-	0,91	0,89±0,02	0,86	0,92
2	DL	-	0,96	0,96±0,01	0,95	0,96
3	BCE_DL	-	0,96	0,96±0,01	0,95	0,97

Capítulo 4. Resultados y discusión

ID _{FP}	Función de pérdidas	Otros parámetros	Mediana	Media±Std	Mínimo	Máximo
4	IoUL	-	0,96	0,96±0,00	0,96	0,97
5	TL	$\alpha = 0,1$	0,95	0,95±0,01	0,94	0,95
6		$\alpha = 0,3$	0,96	0,96±0,00	0,95	0,96
7		$\alpha = 0,7$	0,95	0,95±0,02	0,91	0,96
8		$\alpha = 0,9$	0,94	0,94±0,01	0,93	0,95
9	FTL	$\alpha = 0,1$ $\gamma = 4/3$	0,96	0,96±0,01	0,95	0,96
10		$\alpha = 0,3$ $\gamma = 4/3$	0,96	0,96±0,00	0,95	0,97
11		$\alpha = 0,5$ $\gamma = 4/3$	0,96	0,96±0,00	0,96	0,97
12		$\alpha = 0,7$ $\gamma = 4/3$	0,96	0,96±0,00	0,96	0,96
13		$\alpha = 0,9$ $\gamma = 4/3$	0,94	0,90±0,09	0,72	0,96
14		CL	$\alpha = 0,3$ $\beta = 0,1$	0,96	0,96±0,00	0,95
15	$\alpha = 0,3$ $\beta = 0,3$		0,96	0,95±0,01	0,94	0,96
16	$\alpha = 0,3$ $\beta = 0,5$		0,95	0,95±0,00	0,95	0,95
17	$\alpha = 0,3$ $\beta = 0,7$		0,96	0,96±0,00	0,95	0,97
18	$\alpha = 0,3$ $\beta = 0,9$		0,96	0,96±0,00	0,95	0,96
19	$\alpha = 0,5$ $\beta = 0,1$		0,95	0,95±0,00	0,94	0,95
20	$\alpha = 0,5$ $\beta = 0,3$		0,96	0,96±0,00	0,95	0,96
21	$\alpha = 0,5$ $\beta = 0,5$		0,96	0,95±0,00	0,95	0,96
22	$\alpha = 0,5$ $\beta = 0,7$		0,94	0,95±0,00	0,94	0,95
23	$\alpha = 0,5$ $\beta = 0,9$		0,96	0,96±0,00	0,95	0,96

Capítulo 4. Resultados y discusión

ID _{FP}	Función de pérdidas	Otros parámetros	Mediana	Media±Std	Mínimo	Máximo
24		$\alpha = 0,7$ $\beta = 0,1$	0,96	0,95±0,00	0,95	0,96
25		$\alpha = 0,7$ $\beta = 0,3$	0,95	0,94±0,00	0,94	0,95
26		$\alpha = 0,7$ $\beta = 0,5$	0,96	0,96±0,00	0,95	0,97
27		$\alpha = 0,7$ $\beta = 0,7$	0,96	0,95±0,00	0,95	0,96
28		$\alpha = 0,7$ $\beta = 0,9$	0,94	0,94±0,01	0,93	0,95

Por lo tanto, para la segmentación automática de grasa visceral y abdominal en imágenes de TC finalmente se seleccionó una tasa de aprendizaje basada en un planificador dinámico con una tasa de aprendizaje inicial de 0,1 y una disminución de ésta un factor 0,1 cuando las pérdidas en validación se mantienen constantes 15 épocas; y una función de pérdidas Focal Tversky con $\alpha = 0,5$.

4.2.2.2. Test del modelo de segmentación

En la Tabla 23 se muestran las métricas obtenidas sobre el conjunto de datos de test (n=12). En este caso se extrajeron métricas de cada corte de manera individual ya que, en el conjunto de datos de test, había algunos casos en los que no todas las imágenes de la serie se habían segmentado manualmente y, por tanto, no se podía realizar la comparativa volumétrica. Del mismo modo, las métricas relacionadas con el volumen se sustituyeron por las correspondientes haciendo uso del área (e.g., RAD – Diferencia de Área Relativa). Podemos observar una elevada capacidad de generalización a imágenes independientes representado con un elevado grado de solapamiento entre las segmentaciones manuales y las generadas por el modelo diseñado, con una mediana de DC del 95% para la grasa visceral y del 97% para la subcutánea.

Capítulo 4. Resultados y discusión

Tabla 23. Métricas obtenidas en la comparativa de la máscara de referencia (manual) y la generada automáticamente por la CNN en el conjunto de datos de test tanto para la segmentación de la grasa visceral (tabla superior) como de la grasa subcutánea (tabla inferior). Para cada métrica se ha calculado la mediana, media y desviación estándar.

Grasa visceral		
Métrica	Mediana	Media±Std
DC	0,95	0,93±0,06
IoU	0,91	0,88±0,09
Sensib.	0,95	0,93±0,07
Especif.	1,00	1,00±0,00
Precisión	0,95	0,94±0,06
FDR	0,05	0,06±0,06
RAD	0,02	0,05±0,10
ASSD (mm)	0,46	0,85±1,59
MSD (mm)	20,58	26,58±18,02
Grasa subcutánea		
Métrica	Mediana	Media±Std
DC	0,97	0,96±0,03
IoU	0,94	0,92±0,05
Sensib.	0,97	0,97±0,03
Especif.	1,00	1,00±0,00
Precisión	0,96	0,95±0,04
FDR	0,04	0,05±0,04
RAD	0,01	0,02±0,03
ASSD (mm)	0,44	0,54±0,44
MSD (mm)	4,65	9,16±14,09

A continuación, se realizó un estudio de correlación entre las áreas de la grasa visceral y subcutánea calculadas a partir de la máscara de referencia y las calculadas a partir de la máscara generada por la CNN (Figura 69). La prueba de Shapiro-Wilk mostró que la distribución de áreas no seguía una distribución normal ($p < 0,05$), por lo tanto, se empleó el coeficiente de Spearman para el estudio de la correlación, obteniendo un valor de 0,99 ($p < 0,001$) para la grasa visceral y de 1,00 ($p < 0,001$) para la grasa subcutánea.

Por otro lado, se realizó el análisis Bland-Altman para comparar las diferencias de las áreas calculadas (Figura 69). Este análisis muestra un sesgo bajo (0,57 cm² para la grasa visceral y -1,6 cm² para la grasa subcutánea) y límites de acuerdo estrechos (-9,23-10,36 cm² para la grasa visceral y -10,22-7,02 cm² para la grasa subcutánea).

Capítulo 4. Resultados y discusión

Podemos ver que en únicamente 4 imágenes para la grasa visceral y 3 imágenes para la grasa subcutánea de las 144 que encontramos en el conjunto de imágenes de test se encontraban fuera del intervalo de confianza.

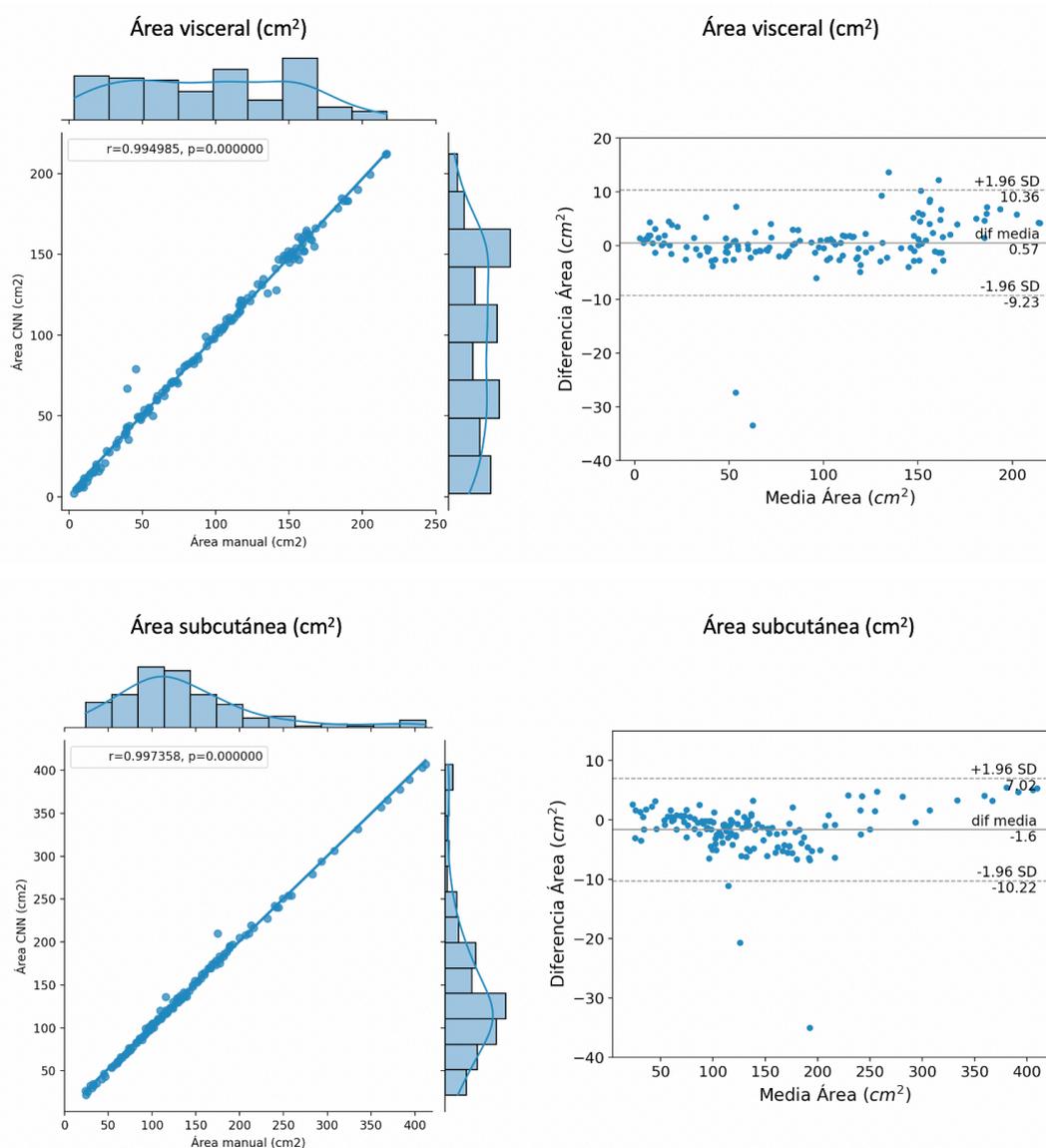


Figura 69. Análisis de correlación entre el área de la grasa visceral (superior) y subcutánea (inferior) calculados sobre la segmentación de referencia (manual) y la predicha por la CNN sobre imágenes de TC. A la izquierda vemos el diagrama de pares entre ambas áreas; en los ejes superior y derecho se encuentran los diagramas de las áreas de la segmentación manual y predicha, respectivamente. A la derecha se muestra el diagrama Bland-Altman.

Finalmente, en la Figura 70 podemos observar el funcionamiento del modelo entrenado en dos casos del conjunto de datos de test, cada uno con cantidades diferentes de grasa (Ejemplo 1: 114 cm² de grasa subcutánea y 47 cm² de grasa visceral a nivel de la vértebra L3; Ejemplo 2: 62 cm² de grasa subcutánea y 19 cm² de grasa visceral a nivel de la vértebra L3).

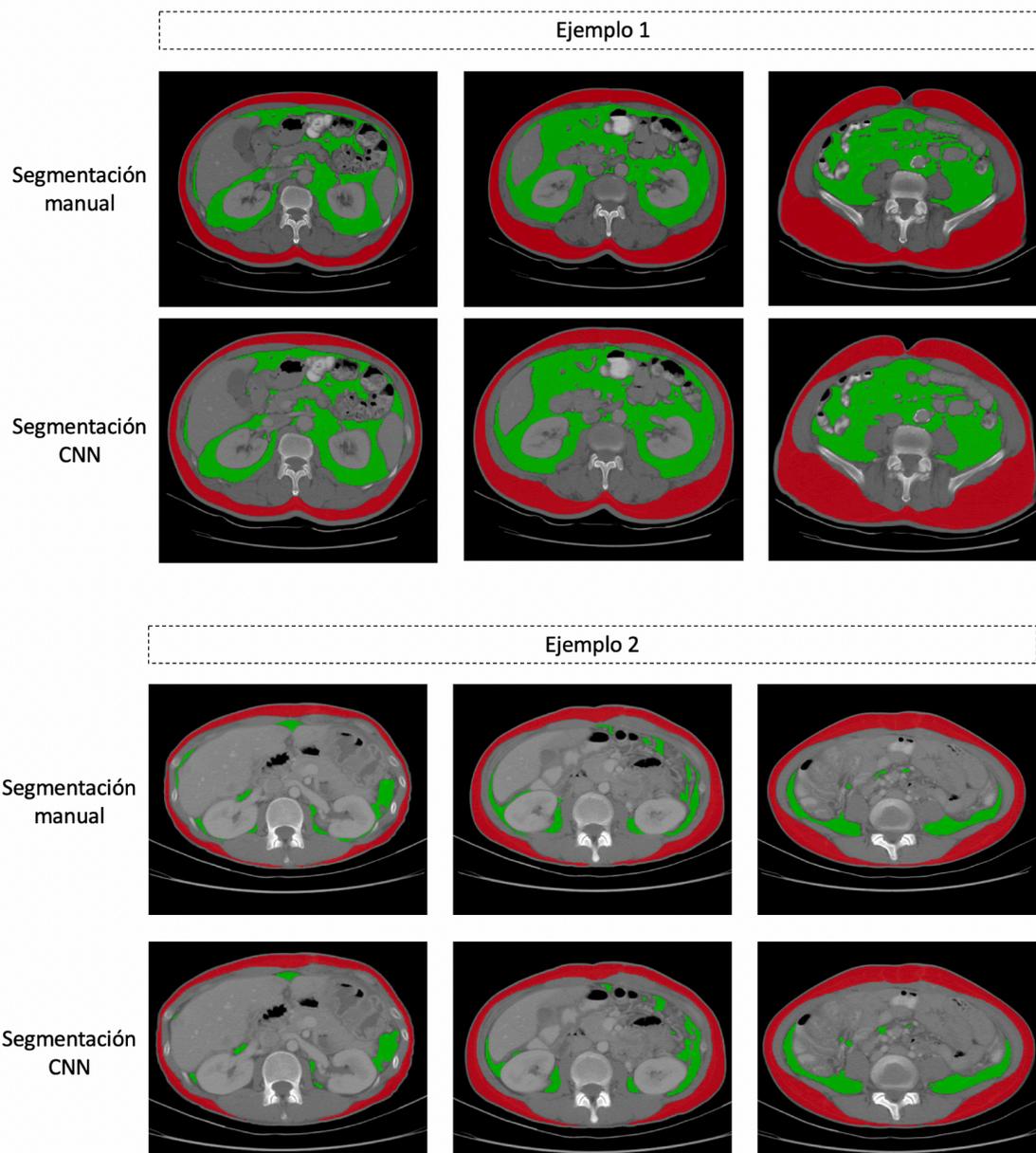


Figura 70. Resultados obtenidos mediante la segmentación automática de la grasa visceral (verde) y subcutánea (rojo) en dos casos del subconjunto de test, cada uno con cantidades diferentes de grasa. En ambos ejemplos podemos comparar la segmentación de referencia (fila superior) y la obtenida por la CNN (fila inferior) en tres cortes a diferentes niveles dentro de la región abdominal.

4.2.3. Discusión

Se ha desarrollado un método de segmentación automática de grasa visceral y subcutánea en toda la región abdominal empleando imágenes de TC. Para ello, se han implementado dos modelos basados en CNN, uno de clasificación, para realizar una localización de la región abdominal en un estudio de TC con cobertura anatómica arbitraria; y otro de segmentación, para la identificación de los vóxeles pertenecientes al tejido adiposo

visceral y subcutáneo, así como para su estudio volumétrico, de gran interés en el estudio de la composición corporal y el síndrome metabólico.

Tradicionalmente, la composición corporal se ha evaluado haciendo uso de un único corte bidimensional a nivel de la vértebra L3 [121, 122]. Sin embargo, en el estudio de la grasa puede haber diferencias debidas al sexo, edad o nivel corporal, por lo tanto, hay estudios que sugieren analizar la grasa en distintos niveles anatómicos [22]. Además, dado que el contenido del tracto gastrointestinal está en continuo movimiento, no hay garantía de que dos cortes a nivel de L3, adquiridos en dos instantes diferentes, incluyan la misma anatomía. Por lo que, una de las principales ventajas de este estudio es el análisis de la composición corporal volumétrica en toda la región abdominal, permitiendo ofrecer una visión global de la distribución de ambos tejidos adiposos.

Actualmente, se encuentran múltiples soluciones para la segmentación automática de la grasa visceral y subcutánea en un único corte [123, 124], siendo inferior la cantidad de trabajos en los que se realiza una segmentación volumétrica. Entre estos últimos encontramos el trabajo propuesto por Mat et al. [125] con un DC medio de 0,99 para la grasa subcutánea y de 0,96 para la grasa visceral, resultados similares a los obtenidos en el modelo propuesto ($0,93 \pm 0,06$ para la grasa visceral y $0,96 \pm 0,03$ para la grasa subcutánea). Sin embargo, estos resultados no son comparables directamente al tratarse de tamaños muestrales diferentes, en nuestro caso empleamos 12 casos para el testeo del modelo mientras que en [125] se emplearon 50. En el estudio de correlación entre las áreas segmentadas por el método automático y la segmentación manual se observó una elevada correlación (0,99 para la grasa visceral y 1,00 para la grasa subcutánea con $p < 0,001$), además, en el estudio Bland-Altman se observó un sesgo bajo en ambos casos ($0,57 \text{ cm}^2$ para la grasa visceral y $-1,6 \text{ cm}^2$ para la grasa subcutánea).

Una de las principales limitaciones de este trabajo es el tamaño de la base de datos anotada, contando con 59 casos etiquetados de los cuales 47 se emplearon para el entrenamiento del modelo. Por lo tanto, se espera que con una base de datos más grande se obtengan mejores resultados.

En imágenes de TC las intensidades se encuentran en un rango normalizado de UH, por lo tanto, encontramos menos variabilidad que en imágenes de RM que genera niveles de intensidad en un rango dinámico y, por tanto, más variable que en una imagen de TC. Sin embargo, podemos encontrar variabilidad por otros factores como la corriente del tubo, el kernel de reconstrucción o la resolución y cobertura de la imagen, entre otros. Así, para garantizar la generalización del modelo a todos estos factores es necesario evaluar el

Capítulo 4. Resultados y discusión

rendimiento del modelo ante ellos. También, se puede emplear métodos de armonización basados en CNN que permitan reducir dicha variabilidad [126, 127].

Si bien es cierto que la imagen de TC está basada en una adquisición que emite radiación ionizante, es de gran interés en el estudio de la composición corporal en pacientes oncológicos ya que este estudio de imagen se realiza de manera rutinaria. Procesos metabólicos como la sarcopenia o la obesidad sarcopénica son comunes en pacientes oncológicos y es de gran importancia identificarlos ya que afectan a la eficacia y toxicidad de la quimioterapia [23-25]. Por lo tanto, mediante la segmentación automática de la grasa visceral y subcutánea, junto con información de la masa muscular del paciente, se puede identificar pacientes que sufren este tipo de procesos metabólicos.

4.3. Segmentación hepática y cuantificación en secuencias de perfusión

En esta sección se van a analizar los resultados obtenidos en el modelo entrenado para la segmentación automática hepática en imágenes de RM de perfusión, así como en la validación del método de cuantificación implementado para la extracción de descriptores de la curva de concentración de contraste.

4.3.1. Validación del modelo de segmentación hepática

En este apartado se exponen los resultados obtenidos en el proceso de validación y ajuste de hiperparámetros del modelo de segmentación automática para cada una de las orientaciones.

4.3.1.1. Orientación transversal

En primer lugar, se hizo un ajuste de los hiperparámetros de la CNN para el modelo entrenado con imágenes orientadas en la dirección transversal.

4.3.1.1.1. Tasa de aprendizaje

Se comprobó el efecto de emplear una tasa de aprendizaje constante durante el proceso de entrenamiento. En la Figura 71 podemos observar los resultados obtenidos durante el proceso de validación cruzada al emplear distintos valores. Además, la Tabla 24 muestra las métricas obtenidas en cada caso. Podemos observar que a partir de un valor de 10^{-3} se produce una disminución del rendimiento de la red ya que los modelos no llegan a converger en el número de épocas establecido. Los mejores resultados empleando una tasa de aprendizaje constante se obtuvieron al emplear un valor de 0,1 (ID_{TA} : 2), siendo el valor medio de DC a lo largo de las 5 iteraciones de 97,23% (DC min: 94,13% - DC máx: 98,23%).

Capítulo 4. Resultados y discusión

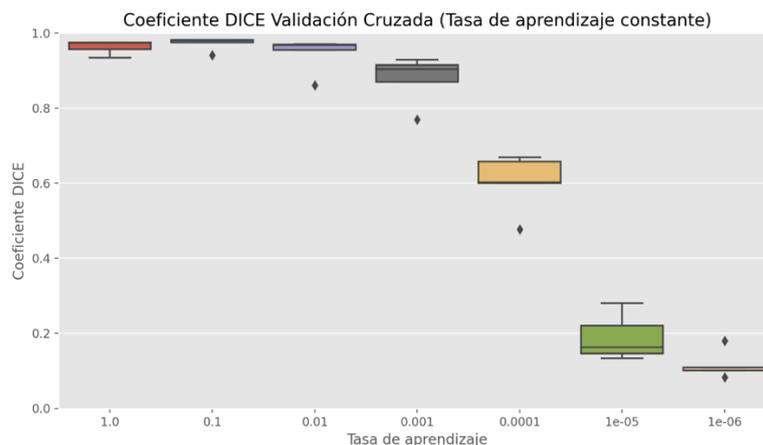


Figura 71. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación transversal: Coeficiente Dice obtenido en el proceso de validación cruzada sobre el conjunto de datos de validación al emplear distintos valores de tasa de aprendizaje constante.

Tabla 24. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación transversal: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos valores de tasa de aprendizaje constante. En negrita se resalta la tasa de aprendizaje constante que mejores resultados ofreció.

ID _{TA}	TA	Mediana	Media±Std	Mínimo	Máximo
1	1	0,97	0,96±0,02	0,93	0,98
2	10⁻¹	0,98	0,97±0,02	0,94	0,98
3	10 ⁻²	0,97	0,94±0,04	0,86	0,97
4	10 ⁻³	0,90	0,88±0,06	0,77	0,93
5	10 ⁻⁴	0,60	0,60±0,07	0,48	0,67
6	10 ⁻⁵	0,16	0,19±0,05	0,13	0,28
7	10 ⁻⁶	0,11	0,12±0,03	0,08	0,18

A continuación, se evaluó el rendimiento empleando un planificador de la tasa de aprendizaje. En la Figura 72 se puede observar el DC durante el proceso de validación cruzada empleando diferentes planificadores y diferentes configuraciones de estos. Además, la Tabla 25 ofrece las métricas que ofrecieron dichos planificadores. En general, se puede observar menor estabilidad en todas sus configuraciones en los planificadores basado en el tiempo, lineal y polinomial. Los mejores resultados se obtuvieron empleando un planificador dinámico con una tasa de aprendizaje inicial de 0,1 y reduciéndose esta un factor 0,1 tras 5 épocas con los mismos resultados (ID_{TA}: 54). Empleando esta configuración se obtuvo un DC medio de 98,20% (DC min: 97,87% - DC máx: 98,42%).

Capítulo 4. Resultados y discusión

Se puede observar, por lo tanto, que en este caso se obtuvieron mejores resultados empleando una tasa de aprendizaje variable que una constante, ofreciendo, además de un DC medio superior, mayor estabilidad en el entrenamiento.

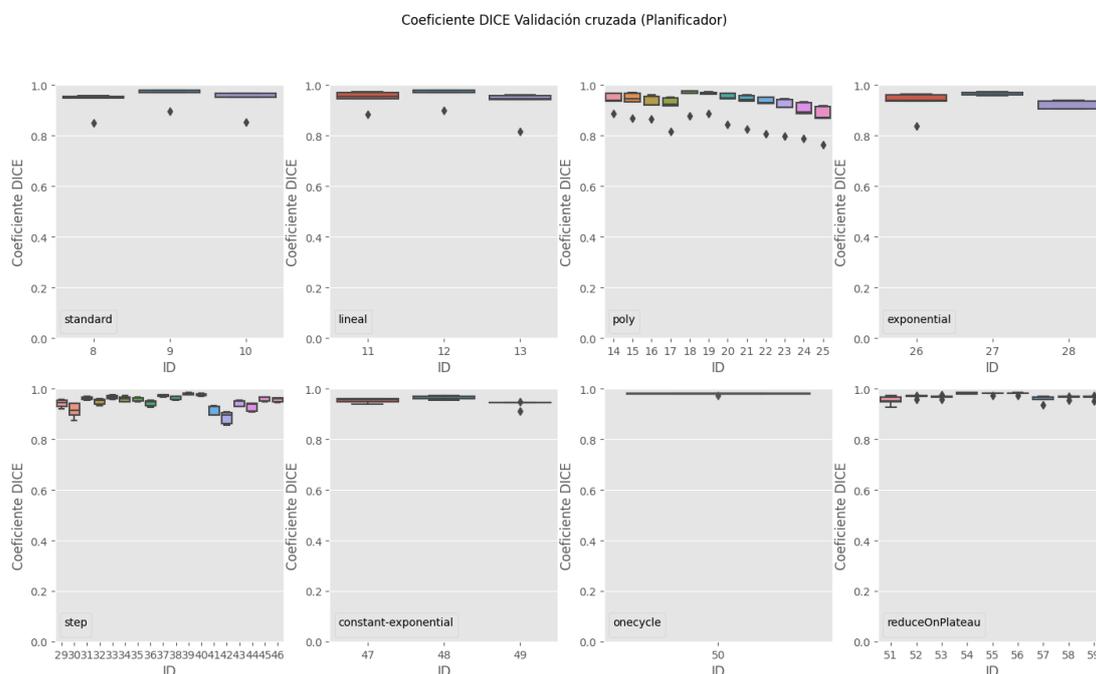


Figura 72. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación transversal: Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos planificadores de tasa de aprendizaje con distintas configuraciones de estos. El eje x representa el identificador del modelo entrenado.

Tabla 25. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación transversal: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos planificadores de tasa de aprendizaje. En negrita se resalta el planificador que mejores resultados ofreció siendo este, además, el que se empleó para el entrenamiento del modelo final.

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo
8	Basado en el tiempo	1	-	0,95	0,93±0,04	0,85	0,96
9		10 ⁻¹	-	0,97	0,96±0,03	0,90	0,98
10		10 ⁻²	-	0,95	0,94±0,04	0,85	0,97
11	Lineal	1	-	0,95	0,95±0,03	0,88	0,97
12		10 ⁻¹	-	0,97	0,96±0,03	0,90	0,98
13		10 ⁻²	-	0,95	0,92±0,05	0,82	0,96
14	Polinomial	1	Factor: 2	0,94	0,94±0,03	0,89	0,97
15			Factor: 3	0,95	0,94±0,04	0,87	0,97
16			Factor: 5	0,92	0,92±0,03	0,87	0,96
17			Factor: 7	0,93	0,91±0,05	0,82	0,95
18		10 ⁻¹	Factor: 2	0,97	0,95±0,04	0,88	0,98

Capítulo 4. Resultados y discusión

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo	
19			Factor: 3	0,96	0,95±0,03	0,89	0,97	
20			Factor: 5	0,95	0,93±0,05	0,84	0,97	
21			Factor: 7	0,94	0,92±0,05	0,83	0,96	
22			10 ⁻²	Factor: 2	0,93	0,91±0,05	0,81	0,95
23				Factor: 3	0,91	0,90±0,05	0,80	0,95
24				Factor: 5	0,89	0,89±0,05	0,79	0,93
25				Factor: 7	0,87	0,87±0,06	0,77	0,92
26	Exponencial	1	k: 0,1	0,94	0,93±0,05	0,84	0,97	
27		10 ⁻¹		0,97	0,97±0,01	0,96	0,97	
28		10 ⁻²		0,94	0,93±0,02	0,90	0,94	
29	Por saltos	1	Δ _{époc} : 5 Tasa: 0,5	0,94	0,94±0,01	0,92	0,96	
30			Δ _{époc} : 5 Tasa: 0,25	0,91	0,91±0,03	0,88	0,94	
31			Δ _{époc} : 10 Tasa: 0,5	0,96	0,96±0,01	0,95	0,97	
32			Δ _{époc} : 10 Tasa: 0,25	0,95	0,95±0,01	0,93	0,96	
33			Δ _{époc} : 25 Tasa: 0,5	0,97	0,97±0,01	0,96	0,98	
34			Δ _{époc} : 25 Tasa: 0,25	0,96	0,96±0,01	0,95	0,97	
35		10 ⁻¹	Δ _{époc} : 5 Tasa: 0,5	0,96	0,96±0,01	0,95	0,97	
36			Δ _{époc} : 5 Tasa: 0,25	0,95	0,94±0,01	0,93	0,95	
37			Δ _{époc} : 10 Tasa: 0,5	0,98	0,97±0,01	0,97	0,98	
38			Δ _{époc} : 10 Tasa: 0,25	0,97	0,96±0,01	0,95	0,97	
39			Δ _{époc} : 25 Tasa: 0,5	0,98	0,98±0,00	0,98	0,98	
40			Δ _{époc} : 25 Tasa: 0,25	0,98	0,98±0,00	0,97	0,98	

Capítulo 4. Resultados y discusión

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo
41		10 ⁻²	$\Delta_{\acute{e}poc}$: 5 Tasa: 0,5	0,93	0,92±0,02	0,90	0,93
42			$\Delta_{\acute{e}poc}$: 5 Tasa: 0,25	0,89	0,88±0,02	0,85	0,91
43			$\Delta_{\acute{e}poc}$: 10 Tasa: 0,5	0,95	0,94±0,01	0,93	0,95
44			$\Delta_{\acute{e}poc}$: 10 Tasa: 0,25	0,94	0,93±0,02	0,91	0,94
45			$\Delta_{\acute{e}poc}$: 25 Tasa: 0,5	0,97	0,96±0,01	0,95	0,97
46			$\Delta_{\acute{e}poc}$: 25 Tasa: 0,25	0,96	0,96±0,01	0,94	0,96
47	Constante y exponencial	1	$\Delta_{\acute{e}poc}$: 10 k: 0,1	0,96	0,95±0,01	0,94	0,96
48		10 ⁻¹		0,97	0,97±0,01	0,96	0,97
49		10 ⁻²		0,94	0,94±0,01	0,91	0,95
50	De un ciclo	-	TA _{ini} :10 ⁻⁴ TA _{min} :10 ⁻⁵ TA _{max} :10 ⁻² $\Delta_{\acute{e}poc,1}$:10 $\Delta_{\acute{e}poc,2}$:10 k:0,8	0,98	0,98±0,00	0,97	0,98
51	Dinámica	1	$\Delta_{\acute{e}poc}$: 5	0,95	0,95±0,02	0,93	0,97
52			$\Delta_{\acute{e}poc}$: 10	0,97	0,97±0,01	0,96	0,97
53			$\Delta_{\acute{e}poc}$: 15	0,97	0,97±0,01	0,96	0,98
54		10 ⁻¹	$\Delta_{\acute{e}poc}$: 5	0,98	0,98±0,00	0,98	0,98
55			$\Delta_{\acute{e}poc}$: 10	0,98	0,98±0,00	0,97	0,98
56			$\Delta_{\acute{e}poc}$: 15	0,98	0,98±0,00	0,97	0,98
57		10 ⁻²	$\Delta_{\acute{e}poc}$: 5	0,97	0,96±0,01	0,94	0,97
58			$\Delta_{\acute{e}poc}$: 10	0,97	0,97±0,01	0,95	0,97
59			$\Delta_{\acute{e}poc}$: 15	0,97	0,97±0,01	0,95	0,97

Finalmente, se evaluó el uso de diferentes optimizadores. En la Figura 73 se representa el DC en el conjunto de datos de validación durante el proceso de validación cruzada al emplear diferentes optimizadores. Además, en la Tabla 26 se ofrecen las métricas

Capítulo 4. Resultados y discusión

obtenidas en cada caso. Se puede observar que los mejores resultados se obtuvieron con el optimizador Adadelta (ID_{TA} : 63), con un valor medio de DC de 95,21% (DC min: 88,12% - DC máx: 97,59%), sin embargo, ésta no ofrece suficiente estabilidad en el entrenamiento.

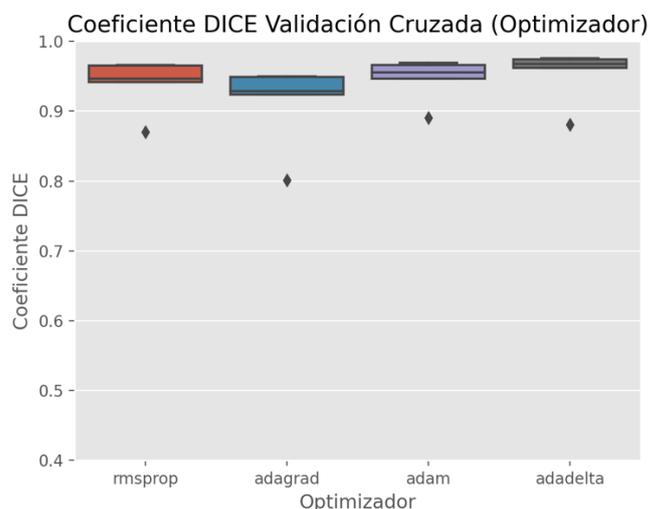


Figura 73. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación transversal: Coeficiente Dice en el conjunto de datos de validación al emplear distintos optimizadores de la tasa de aprendizaje.

Tabla 26. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación transversal: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos optimizadores de tasa de aprendizaje. En negrita se resalta el optimizador que mejores resultados ofreció.

ID_{TA}	Optimizador	Mediana	Media±Std	Mínimo	Máximo
60	RMSprop	0,95	0,94±0,04	0,87	0,97
61	Adagrad	0,93	0,91±0,06	0,80	0,95
62	Adam	0,96	0,95±0,03	0,89	0,97
63	Adadelta	0,97	0,95±0,04	0,88	0,98

Por lo tanto, tanto por el elevado valor de DC que se obtuvo sobre el conjunto de datos de validación (DC medio: 98,20%) como por la estabilidad observada a lo largo de las 5 iteraciones del proceso de validación cruzada (DC min: 97,87% y DC máx: 98,42%), para la resolución de este problema se seleccionó una tasa de aprendizaje variable empleando un planificador dinámico ($\Delta_{\text{épocas}}$: 5) y una tasa de aprendizaje inicial de 0,1 (ID_{TA} : 54).

4.3.1.1.2. Función de pérdidas

Una vez seleccionada la tasa de aprendizaje óptima se abordó el ajuste de la función de pérdidas. En la Figura 74 podemos observar los resultados obtenidos en el subconjunto

Capítulo 4. Resultados y discusión

de datos de validación durante el proceso de validación cruzada, mientras que en la Tabla 27 encontramos las métricas correspondientes. Con algunas de las funciones de pérdidas, como es el caso de la entropía cruzada (ID_{FP}: 1) o la función de pérdidas Focal Tversky con $\alpha = 0,9$ (ID_{FP}: 13), podemos observar que se obtuvo una menor estabilidad durante la validación cruzada, sin embargo, en la mayoría se obtuvo una elevada estabilidad. Los mejores resultados se obtienen con una función de pérdidas Focal Tversky con $\alpha = 0,5$ (ID_{FP}: 11) (DC medio: 98,62%; DC min: 96,68%; DC máx: 99,46%).

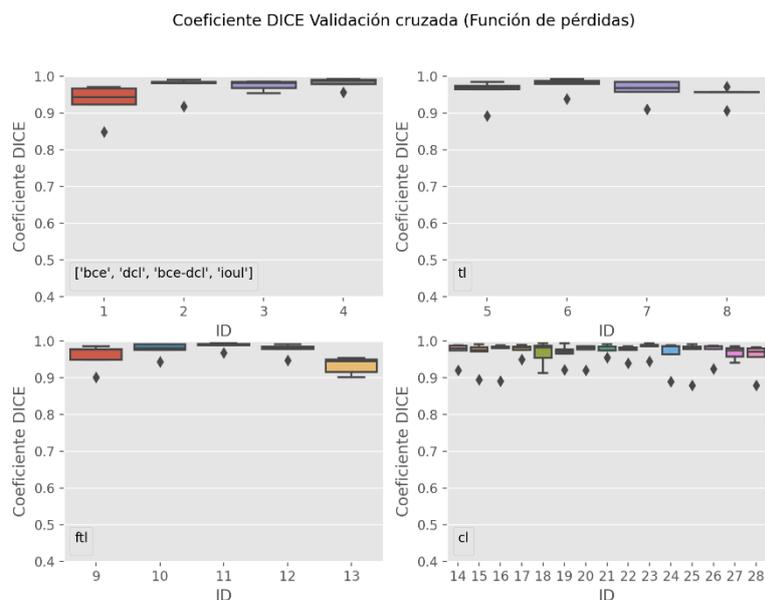


Figura 74. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación transversal: Coeficiente Dice obtenido a lo largo del proceso de validación cruzada sobre el conjunto de datos de validación al emplear distintas funciones de pérdidas. El eje x representa el identificador del modelo entrenado.

Tabla 27. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación transversal: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintas funciones de pérdidas. En negrita se resalta la función de pérdidas que mejores resultados ofreció y que se empleó para el entrenamiento del modelo final.

ID _{FP}	Función de pérdidas	Otros parámetros	Mediana	Media±Std	Mínimo	Máximo
1	BCE	-	0,94	0,93±0,04	0,85	0,97
2	DL	-	0,98	0,97±0,03	0,92	0,99
3	BCE_DL	-	0,98	0,97±0,01	0,95	0,98
4	IoUL	-	0,99	0,98±0,01	0,96	0,99
5	TL	$\alpha = 0,1$	0,97	0,96±0,03	0,89	0,98
6		$\alpha = 0,3$	0,98	0,98±0,02	0,94	0,99
7		$\alpha = 0,7$	0,97	0,96±0,03	0,91	0,98
8		$\alpha = 0,9$	0,96	0,95±0,02	0,91	0,97
9	FTL	$\alpha = 0,1$	0,98	0,96±0,03	0,90	0,99

Capítulo 4. Resultados y discusión

ID _{FP}	Función de pérdidas	Otros parámetros	Mediana	Media±Std	Mínimo	Máximo
		$\gamma = 4/3$				
10		$\alpha = 0,3$ $\gamma = 4/3$	0,98	0,98±0,02	0,94	0,99
11		$\alpha = 0,5$ $\gamma = 4/3$	0,99	0,99±0,01	0,97	0,99
12		$\alpha = 0,7$ $\gamma = 4/3$	0,98	0,98±0,02	0,95	0,99
13		$\alpha = 0,9$ $\gamma = 4/3$	0,94	0,93±0,02	0,90	0,95
14	CL	$\alpha = 0,3$ $\beta = 0,1$	0,98	0,97±0,03	0,92	0,99
15		$\alpha = 0,3$ $\beta = 0,3$	0,97	0,96±0,03	0,89	0,99
16		$\alpha = 0,3$ $\beta = 0,5$	0,98	0,96±0,04	0,89	0,99
17		$\alpha = 0,3$ $\beta = 0,7$	0,98	0,98±0,01	0,95	0,99
18		$\alpha = 0,3$ $\beta = 0,9$	0,98	0,97±0,03	0,91	0,99
19		$\alpha = 0,5$ $\beta = 0,1$	0,97	0,97±0,02	0,92	0,99
20		$\alpha = 0,5$ $\beta = 0,3$	0,98	0,97±0,03	0,92	0,99
21		$\alpha = 0,5$ $\beta = 0,5$	0,97	0,98±0,01	0,95	0,99
22		$\alpha = 0,5$ $\beta = 0,7$	0,98	0,97±0,02	0,94	0,99
23		$\alpha = 0,5$ $\beta = 0,9$	0,99	0,98±0,02	0,94	0,99
24		$\alpha = 0,7$ $\beta = 0,1$	0,99	0,96±0,04	0,89	0,99
25		$\alpha = 0,7$ $\beta = 0,3$	0,98	0,96±0,04	0,88	0,99
26		$\alpha = 0,7$ $\beta = 0,5$	0,97	0,97±0,02	0,92	0,99

Capítulo 4. Resultados y discusión

ID _{FP}	Función de pérdidas	Otros parámetros	Mediana	Media±Std	Mínimo	Máximo
27		$\alpha = 0,7$ $\beta = 0,7$	0,97	0,97±0,02	0,94	0,98
28		$\alpha = 0,7$ $\beta = 0,9$	0,97	0,95±0,04	0,88	0,98

4.3.1.2. Orientación coronal

A continuación, se siguió el mismo procedimiento que para la orientación transversal, pero en este caso se reorientaron las imágenes a la orientación coronal y se siguió el mismo procedimiento.

4.3.1.2.1. Tasa de aprendizaje

En primer lugar, se estudió el uso de una tasa de aprendizaje constante. En la Figura 75 se pueden observar los resultados obtenidos con cada uno de los valores analizados, además, la Tabla 28 muestra las métricas obtenidas en cada caso. Los mejores resultados se obtuvieron empleando una tasa de aprendizaje de 1 (ID_{TA}: 1), con un valor medio de DC a lo largo de las cinco iteraciones de 96,86% (DC min: 95,45% - DC máx: 97,59%).

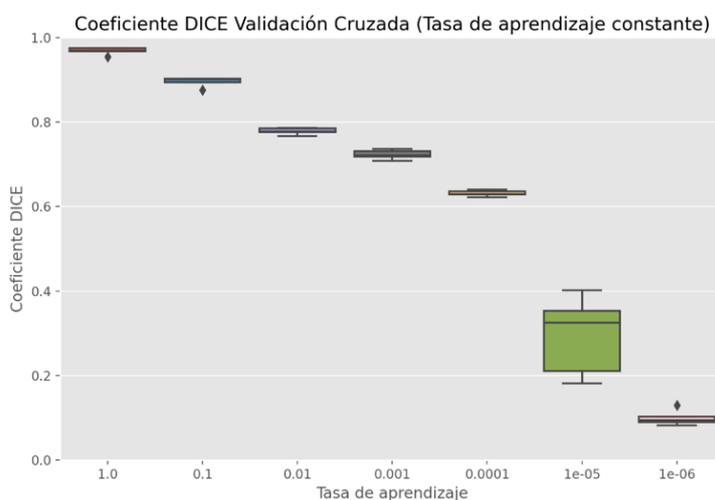


Figura 75. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación coronal: Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos valores de tasa de aprendizaje constante.

Capítulo 4. Resultados y discusión

Tabla 28. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación coronal: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos valores de tasa de aprendizaje constante. En negrita se resalta la tasa de aprendizaje constante que mejores resultados ofreció.

ID _{TA}	TA	Mediana	Media±Std	Mínimo	Máximo
1	1	0,97	0,97±0,01	0,95	0,98
2	10 ⁻¹	0,90	0,90±0,01	0,88	0,90
3	10 ⁻²	0,78	0,78±0,01	0,77	0,79
4	10 ⁻³	0,72	0,72±0,01	0,71	0,74
5	10 ⁻⁴	0,63	0,63±0,01	0,62	0,64
6	10 ⁻⁵	0,32	0,29±0,08	0,18	0,40
7	10 ⁻⁶	0,09	0,10±0,02	0,08	0,13

A continuación, se realizó la evaluación de emplear un planificador de tasa de aprendizaje. En la Figura 76 se puede observar el DC durante el proceso de validación cruzada al emplear los diferentes planificadores definidos. Además, la Tabla 29 ofrece las métricas obtenidas para cada planificador. Al igual que ocurría con la orientación transversal, se observó un peor rendimiento en los modelos con tasa de aprendizaje inicial de 1 basados en los planificadores basado en el tiempo, lineal y polinomial, ocurriendo además en el exponencial en este caso. Los mejores resultados se obtuvieron al emplear un planificador dinámico con tasa de aprendizaje inicial de 1 y reduciéndose esta un factor 0,1 tras 10 épocas con mismos resultados (ID_{TA}: 52). Esta configuración nos dio un valor medio de DC de 96,25% (DC min: 94,81% - DC máx: 97,57%). Siendo ligeramente inferior a los resultados obtenidos con una tasa de aprendizaje constante de 1.

Capítulo 4. Resultados y discusión

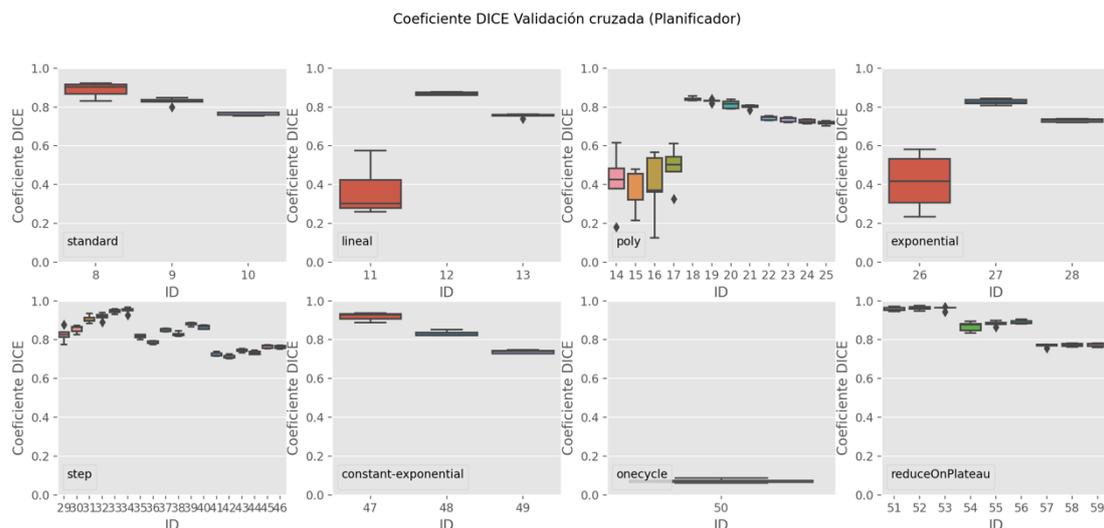


Figura 76. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación coronal: Coeficiente Dice obtenidos a lo largo del proceso de validación cruzada sobre el conjunto de datos de validación al emplear distintos planificadores de tasa de aprendizaje con distintas configuraciones de estos. El eje x representa el identificador del modelo entrenado.

Tabla 29. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación coronal: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos planificadores de tasa de aprendizaje. En negrita se resalta el planificador que mejores resultados ofreció.

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo
8	Basado en el tiempo	1	-	0,90	0,89±0,03	0,83	0,92
9		10 ⁻¹	-	0,83	0,83±0,02	0,80	0,85
10		10 ⁻²	-	0,77	0,77±0,01	0,75	0,77
11	Lineal	1	-	0,30	0,37±0,12	0,26	0,58
12		10 ⁻¹	-	0,87	0,87±0,01	0,86	0,88
13		10 ⁻²	-	0,76	0,76±0,01	0,74	0,76
14	Polinomial	1	Factor: 2	0,43	0,42±0,14	0,18	0,62
15			Factor: 3	0,46	0,39±0,10	0,22	0,48
16			Factor: 5	0,37	0,39±0,16	0,12	0,57
17			Factor: 7	0,50	0,49±0,10	0,33	0,61
18		10 ⁻¹	Factor: 2	0,84	0,84±0,01	0,83	0,86
19			Factor: 3	0,83	0,83±0,01	0,82	0,84
20			Factor: 5	0,82	0,81±0,02	0,79	0,84
21			Factor: 7	0,81	0,80±0,01	0,78	0,81
22		10 ⁻²	Factor: 2	0,75	0,74±0,01	0,73	0,75
23			Factor: 3	0,74	0,74±0,01	0,72	0,75
24			Factor: 5	0,73	0,73±0,01	0,71	0,74
25			Factor: 7	0,72	0,72±0,01	0,70	0,73

Capítulo 4. Resultados y discusión

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo	
26	Exponencial	1	k: 0,1	0,42	0,41±0,13	0,23	0,58	
27		10 ⁻¹		0,83	0,83±0,01	0,81	0,84	
28		10 ⁻²		0,73	0,73±0,01	0,72	0,74	
29	Por saltos	1	Δ _{époc} : 5 Tasa: 0,5	0,83	0,83±0,03	0,78	0,88	
30			Δ _{époc} : 5 Tasa: 0,25	0,86	0,85±0,02	0,83	0,87	
31			Δ _{époc} : 10 Tasa: 0,5	0,90	0,91±0,02	0,88	0,93	
32			Δ _{époc} : 10 Tasa: 0,25	0,92	0,92±0,02	0,89	0,94	
33			Δ _{époc} : 25 Tasa: 0,5	0,95	0,95±0,01	0,93	0,96	
34			Δ _{époc} : 25 Tasa: 0,25	0,95	0,95±0,01	0,93	0,97	
35			10 ⁻¹	Δ _{époc} : 5 Tasa: 0,5	0,82	0,82±0,01	0,80	0,83
36				Δ _{époc} : 5 Tasa: 0,25	0,79	0,79±0,01	0,78	0,80
37		Δ _{époc} : 10 Tasa: 0,5		0,85	0,85±0,01	0,84	0,86	
38		Δ _{époc} : 10 Tasa: 0,25		0,83	0,83±0,01	0,82	0,85	
39		Δ _{époc} : 25 Tasa: 0,5		0,88	0,88±0,01	0,87	0,89	
40		Δ _{époc} : 25 Tasa: 0,25		0,87	0,87±0,01	0,85	0,88	
41		10 ⁻²		Δ _{époc} : 5 Tasa: 0,5	0,72	0,72±0,01	0,72	0,74
42				Δ _{époc} : 5 Tasa: 0,25	0,71	0,71±0,01	0,70	0,73
43			Δ _{époc} : 10 Tasa: 0,5	0,74	0,74±0,01	0,73	0,75	
44			Δ _{époc} : 10	0,73	0,73±0,01	0,72	0,74	

Capítulo 4. Resultados y discusión

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo
			Tasa: 0,25				
45			$\Delta_{\acute{e}poc}$: 25 Tasa: 0,5	0,77	0,76±0,01	0,75	0,77
46			$\Delta_{\acute{e}poc}$: 25 Tasa: 0,25	0,76	0,76±0,01	0,75	0,77
47	Constante y exponencial	1	$\Delta_{\acute{e}poc}$: 10 k: 0,1	0,93	0,92±0,02	0,89	0,94
48		10 ⁻¹		0,84	0,83±0,01	0,82	0,85
49		10 ⁻²		0,74	0,74±0,01	0,73	0,75
50	De un ciclo	-	TA _{ini} :10 ⁻⁴ TA _{min} :10 ⁻⁵ TA _{max} :10 ⁻² $\Delta_{\acute{e}poc,1}$:10 $\Delta_{\acute{e}poc,2}$:10 k:0,8	0,07	0,07±0,01	0,06	0,09
51	Dinámica	1	$\Delta_{\acute{e}poc}$: 5	0,95	0,96±0,01	0,95	0,97
52			$\Delta_{\acute{e}poc}$: 10	0,96	0,96±0,01	0,95	0,98
53			$\Delta_{\acute{e}poc}$: 15	0,96	0,96±0,01	0,94	0,97
54		10 ⁻¹	$\Delta_{\acute{e}poc}$: 5	0,88	0,87±0,02	0,84	0,89
55			$\Delta_{\acute{e}poc}$: 10	0,89	0,88±0,01	0,87	0,90
56			$\Delta_{\acute{e}poc}$: 15	0,89	0,89±0,01	0,88	0,91
57		10 ⁻²	$\Delta_{\acute{e}poc}$: 5	0,77	0,77±0,01	0,76	0,78
58			$\Delta_{\acute{e}poc}$: 10	0,78	0,77±0,01	0,76	0,78
59			$\Delta_{\acute{e}poc}$: 15	0,78	0,77±0,01	0,76	0,78

Finalmente, se evaluó el uso de un optimizador. La Figura 77 muestra el DC obtenido durante el proceso de validación cruzada para cada uno de los optimizadores evaluados, mientras que la Tabla 30 muestra las métricas obtenidas en cada caso. Los mejores resultados se obtuvieron con el optimizador Adam, (ID_{TA}: 62) siendo el DC medio obtenido de 97,19% (DC min: 96,41% - DC máx: 97,92%).

Capítulo 4. Resultados y discusión

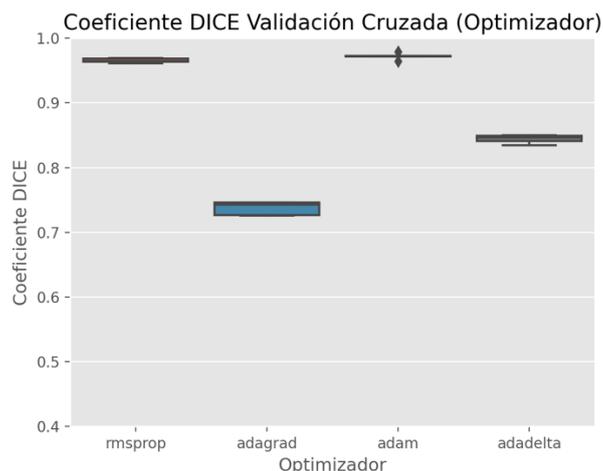


Figura 77. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación coronal: Coeficiente Dice en el conjunto de datos de validación al emplear distintos optimizadores de tasa de aprendizaje.

Tabla 30. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación coronal: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos optimizadores de tasa de aprendizaje. En negrita se resalta el optimizador que mejores resultados ofreció siendo, además, la que se empleó para el entrenamiento del modelo final.

ID _{TA}	Optimizador	Mediana	Media±Std	Mínimo	Máximo
60	RMSprop	0,96	0,97±0,00	0,96	0,97
61	Adagrad	0,74	0,74±0,01	0,73	0,75
62	Adam	0,97	0,97±0,00	0,96	0,98
63	Adadelta	0,85	0,84±0,01	0,83	0,85

Por lo tanto, tanto por el elevado valor de DC que se obtuvo sobre el conjunto de datos de validación (DC medio: 97,19%) como por la estabilidad observada a lo largo de las 5 iteraciones del proceso de validación cruzada (DC min: 96,41% y DC máx: 97,92%), para la resolución de este problema se seleccionó una tasa de aprendizaje variable empleando el optimizador Adam.

4.3.1.2.2. Función de pérdidas

A continuación, una vez seleccionada la tasa de aprendizaje, se realizaron los experimentos necesarios para la selección de la función de pérdidas. La Figura 78 muestra los resultados del proceso de validación cruzada para cada una de las funciones de pérdidas evaluadas. Además, la Tabla 31 ofrece las métricas obtenidas en cada caso. Se puede observar, que con la entropía cruzada binaria (ID_{FP}: 1) y la entropía cruzada binaria con coeficiente Dice (ID_{FP}: 3), la estabilidad en el entrenamiento no fue buena, en varias de las validaciones se obtuvieron valores de DC menores de 0,5. Con el resto de las

Capítulo 4. Resultados y discusión

funciones de pérdidas se obtuvieron buenos resultados tanto considerando el valor de DC como la estabilidad entre las interacciones. Finalmente, los mejores resultados se alcanzaron con la función de pérdidas Combo con $\alpha = 0,5$ y $\beta = 0,3$ (ID_{FP}: 20) con DC medio: 98,04%, DC min: 97,54% y DC máx: 98,52%.

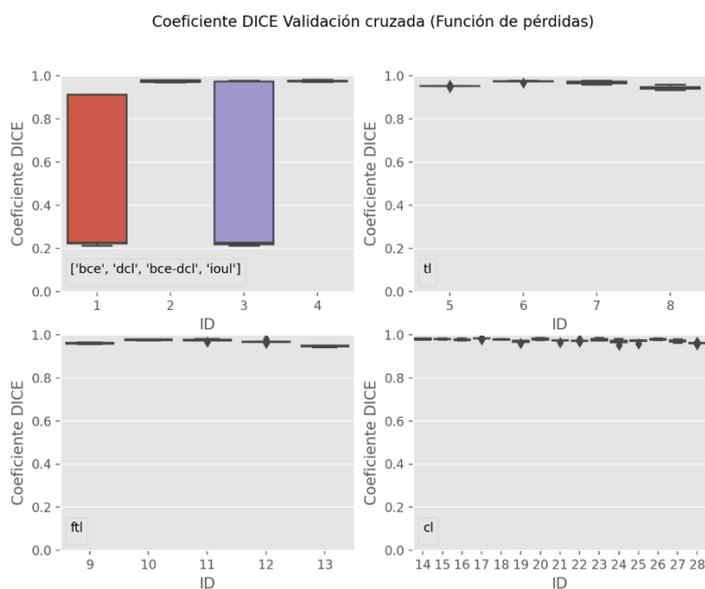


Figura 78. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación coronal: Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintas funciones de pérdidas. El eje x representa el identificador del modelo entrenado.

Tabla 31. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación coronal: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintas funciones de pérdidas. En negrita se resalta la función de pérdidas que mejores resultados ofreció y que se empleó para el entrenamiento del modelo final.

ID _{FP}	Función de pérdidas	Otros parámetros	Mediana	Media±Std	Mínimo	Máximo
1	BCE	-	0,23	0,50±0,34	0,21	0,91
2	DL	-	0,97	0,98±0,00	0,97	0,98
3	BCE_DL	-	0,23	0,52±0,37	0,21	0,98
4	IoUL	-	0,97	0,98±0,00	0,97	0,98
5	TL	$\alpha = 0,1$	0,95	0,95±0,00	0,95	0,96
6		$\alpha = 0,3$	0,98	0,97±0,00	0,97	0,98
7		$\alpha = 0,7$	0,97	0,97±0,01	0,96	0,98
8		$\alpha = 0,9$	0,94	0,95±0,01	0,94	0,96
9	FTL	$\alpha = 0,1$ $\gamma = 4/3$	0,96	0,96±0,00	0,96	0,96
10		$\alpha = 0,3$ $\gamma = 4/3$	0,98	0,98±0,00	0,97	0,98
11		$\alpha = 0,5$	0,98	0,97±0,00	0,97	0,98

Capítulo 4. Resultados y discusión

ID _{FP}	Función de pérdidas	Otros parámetros	Mediana	Media±Std	Mínimo	Máximo
		$\gamma = 4/3$				
12		$\alpha = 0,7$ $\gamma = 4/3$	0,97	0,97±0,01	0,96	0,98
13		$\alpha = 0,9$ $\gamma = 4/3$	0,95	0,95±0,00	0,94	0,95
14	CL	$\alpha = 0,3$ $\beta = 0,1$	0,98	0,98±0,00	0,98	0,98
15		$\alpha = 0,3$ $\beta = 0,3$	0,98	0,98±0,00	0,98	0,98
16		$\alpha = 0,3$ $\beta = 0,5$	0,98	0,98±0,00	0,97	0,98
17		$\alpha = 0,3$ $\beta = 0,7$	0,98	0,98±0,00	0,98	0,98
18		$\alpha = 0,3$ $\beta = 0,9$	0,98	0,98±0,00	0,98	0,98
19		$\alpha = 0,5$ $\beta = 0,1$	0,97	0,97±0,01	0,96	0,97
20		$\alpha = 0,5$ $\beta = 0,3$	0,98	0,98±0,00	0,98	0,99
21		$\alpha = 0,5$ $\beta = 0,5$	0,97	0,97±0,00	0,96	0,98
22		$\alpha = 0,5$ $\beta = 0,7$	0,97	0,97±0,00	0,97	0,98
23		$\alpha = 0,5$ $\beta = 0,9$	0,97	0,98±0,00	0,97	0,98
24		$\alpha = 0,7$ $\beta = 0,1$	0,97	0,97±0,01	0,95	0,98
25		$\alpha = 0,7$ $\beta = 0,3$	0,97	0,97±0,01	0,96	0,97
26		$\alpha = 0,7$ $\beta = 0,5$	0,98	0,98±0,00	0,97	0,98
27		$\alpha = 0,7$ $\beta = 0,7$	0,97	0,97±0,01	0,96	0,98
28		$\alpha = 0,7$ $\beta = 0,9$	0,96	0,96±0,00	0,95	0,97

Capítulo 4. Resultados y discusión

4.3.1.3. Orientación sagital

Finalmente, se realizó el mismo procedimiento que en las dos orientaciones anteriores para el desarrollo de un modelo entrenado con imágenes en la orientación sagital.

4.3.1.3.1. Tasa de aprendizaje

En primer lugar, se evaluó el rendimiento del modelo empleando una tasa de aprendizaje constante. En la Figura 79 se puede observar que, los resultados son significativamente mejores al emplear una tasa de aprendizaje constante de 1 ($ID_{TA}: 1$), con un valor medio de DC a lo largo de las cinco iteraciones de 94,32% (DC min: 93,18% - DC máx: 95,68%). Esto se aprecia también al observar las métricas obtenidas para cada tasa de aprendizaje (Tabla 32).

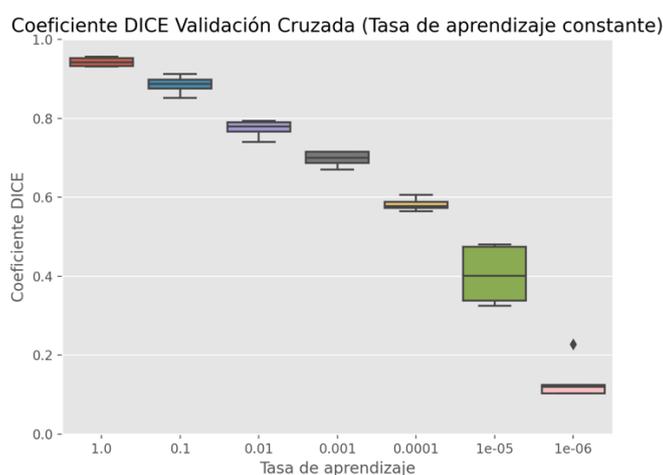


Figura 79. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación sagital: Coeficiente Dice a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos valores de tasa de aprendizaje constante durante el proceso de entrenamiento.

Tabla 32. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación sagital: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos valores de tasa de aprendizaje constante. En negrita se resalta la tasa de aprendizaje constante que mejores resultados ofreció siendo esta, además, la que se empleó para el entrenamiento del modelo final.

ID_{TA}	TA	Mediana	Media±Std	Mínimo	Máximo
1	1	0,94	0,94±0,01	0,93	0,96
2	10^{-1}	0,89	0,89±0,02	0,85	0,91
3	10^{-2}	0,78	0,77±0,02	0,74	0,79
4	10^{-3}	0,70	0,70±0,02	0,67	0,72
5	10^{-4}	0,58	0,58±0,01	0,56	0,61
6	10^{-5}	0,40	0,40±0,07	0,32	0,48

Capítulo 4. Resultados y discusión

7	10^{-6}	0,12	$0,14 \pm 0,05$	0,10	0,23
---	-----------	------	-----------------	------	------

A continuación, se evaluó el rendimiento de los modelos al emplear los diferentes planificadores de tasa de aprendizaje. En este caso, como se observa en la Figura 80 y en la Tabla 33, al igual que para la orientación coronal, los mejores resultados se obtuvieron al emplear un planificador dinámico con tasa de aprendizaje inicial de 1 y reduciéndose esta un factor 0,1 tras 10 épocas con mismos resultados (ID_{TA} : 52). Esta configuración nos dio un valor medio de DC de 94,10% (DC min: 92,28% - DC máx: 95,24%), ligeramente inferior al DC obtenido con una tasa de aprendizaje constante de 1. Además, se mantiene el descenso en el rendimiento y estabilidad al emplear una tasa de aprendizaje inicial de 1 con planificador línea, polinomial y exponencial.

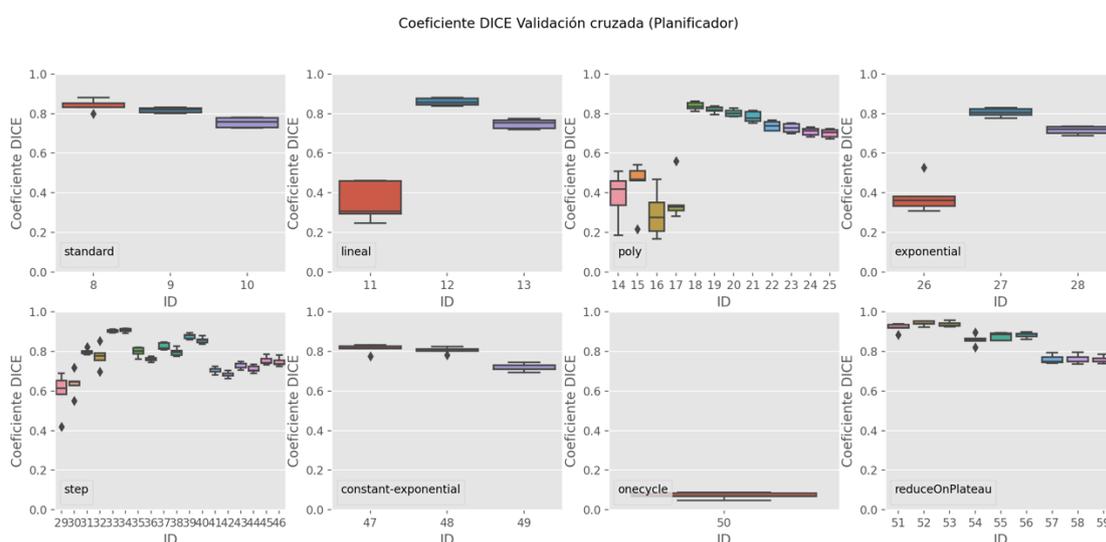


Figura 80. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación sagital: Coeficiente Dice a lo largo del proceso de validación cruzada sobre el conjunto de datos de validación al emplear distintos planificadores de tasa de aprendizaje con diferentes configuraciones de éstos. El eje x representa el identificador del modelo entrenado.

Tabla 33. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación sagital: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos planificadores de tasa de aprendizaje. En negrita se resalta el planificador que mejores resultados ofreció.

ID_{TA}	Planificador	TA inicial	Otros	Mediana	Media \pm Std	Mínimo	Máximo
8	Basado en el tiempo	1	-	0,85	$0,84 \pm 0,03$	0,80	0,88
9		10^{-1}	-	0,82	$0,82 \pm 0,01$	0,80	0,83
10		10^{-2}	-	0,76	$0,76 \pm 0,02$	0,73	0,78
11	Lineal	1	-	0,31	$0,35 \pm 0,09$	0,25	0,46
12		10^{-1}	-	0,86	$0,86 \pm 0,02$	0,84	0,88

Capítulo 4. Resultados y discusión

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo	
13		10 ⁻²	-	0,75	0,75±0,02	0,72	0,77	
14	Polinomial	1	Factor: 2	0,42	0,38±0,11	0,19	0,51	
15			Factor: 3	0,47	0,44±0,12	0,22	0,54	
16			Factor: 5	0,28	0,29±0,11	0,17	0,47	
17			Factor: 7	0,33	0,36±0,01	0,28	0,56	
18		10 ⁻¹	Factor: 2	0,84	0,84±0,02	0,81	0,86	
19			Factor: 3	0,82	0,82±0,02	0,80	0,84	
20			Factor: 5	0,80	0,80±0,02	0,78	0,83	
21			Factor: 7	0,78	0,78±0,02	0,75	0,82	
22		10 ⁻²	Factor: 2	0,74	0,74±0,02	0,71	0,77	
23			Factor: 3	0,73	0,73±0,02	0,70	0,75	
24			Factor: 5	0,71	0,71±0,02	0,68	0,73	
25			Factor: 7	0,71	0,70±0,02	0,67	0,72	
26		Exponencial	1	k: 0,1	0,36	0,38±0,08	0,31	0,53
27	10 ⁻¹		0,81		0,81±0,02	0,78	0,83	
28	10 ⁻²		0,72		0,72±0,02	0,69	0,73	
29	Por saltos	1	Δ _{époc} : 5 Tasa: 0,5	0,61	0,59±0,09	0,42	0,69	
30			Δ _{époc} : 5 Tasa: 0,25	0,65	0,64±0,05	0,55	0,72	
31			Δ _{époc} : 10 Tasa: 0,5	0,79	0,80±0,01	0,78	0,82	
32			Δ _{époc} : 10 Tasa: 0,25	0,78	0,77±0,05	0,70	0,85	
33			Δ _{époc} : 25 Tasa: 0,5	0,90	0,90±0,01	0,89	0,91	
34			Δ _{époc} : 25 Tasa: 0,25	0,91	0,91±0,01	0,89	0,92	
35			10 ⁻¹	Δ _{époc} : 5 Tasa: 0,5	0,80	0,80±0,02	0,76	0,82
36				Δ _{époc} : 5 Tasa: 0,25	0,76	0,76±0,01	0,75	0,78
37				Δ _{époc} : 10 Tasa: 0,5	0,82	0,82±0,02	0,81	0,85
38				Δ _{époc} : 10	0,79	0,80±0,02	0,78	0,83

Capítulo 4. Resultados y discusión

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo
			Tasa: 0,25				
39			$\Delta_{\acute{e}poc}$: 25 Tasa: 0,5	0,88	0,88±0,01	0,86	0,89
40			$\Delta_{\acute{e}poc}$: 25 Tasa: 0,25	0,86	0,86±0,02	0,84	0,88
41		10 ⁻²	$\Delta_{\acute{e}poc}$: 5 Tasa: 0,5	0,70	0,70±0,01	0,68	0,72
42			$\Delta_{\acute{e}poc}$: 5 Tasa: 0,25	0,68	0,68±0,01	0,66	0,70
43			$\Delta_{\acute{e}poc}$: 10 Tasa: 0,5	0,72	0,73±0,02	0,70	0,75
44			$\Delta_{\acute{e}poc}$: 10 Tasa: 0,25	0,70	0,71±0,02	0,69	0,73
45			$\Delta_{\acute{e}poc}$: 25 Tasa: 0,5	0,74	0,75±0,02	0,73	0,79
46			$\Delta_{\acute{e}poc}$: 25 Tasa: 0,25	0,74	0,75±0,02	0,72	0,76
47			Constante y exponencial	1	$\Delta_{\acute{e}poc}$: 10 k: 0,1	0,82	0,81±0,02
48	10 ⁻¹	0,81		0,81±0,01		0,78	0,82
49	10 ⁻²	0,71		0,72±0,02		0,69	0,74
50	De un ciclo	-	TA _{ini} :10 ⁻⁴ TA _{min} :10 ⁻⁵ TA _{max} :10 ⁻² $\Delta_{\acute{e}poc,1}$:10 $\Delta_{\acute{e}poc,2}$:10 k:0,8	0,07	0,07±0,01	0,05	0,09
51	Dinámica	1	$\Delta_{\acute{e}poc}$: 5	0,92	0,92±0,02	0,88	0,94
52			$\Delta_{\acute{e}poc}$: 10	0,94	0,94±0,01	0,92	0,95
53			$\Delta_{\acute{e}poc}$: 15	0,93	0,94±0,01	0,92	0,96
54		10 ⁻¹	$\Delta_{\acute{e}poc}$: 5	0,86	0,86±0,02	0,82	0,90
55			$\Delta_{\acute{e}poc}$: 10	0,89	0,88±0,02	0,85	0,89
56			$\Delta_{\acute{e}poc}$: 15	0,89	0,88±0,01	0,86	0,90
57		10 ⁻²	$\Delta_{\acute{e}poc}$: 5	0,75	0,76±0,02	0,74	0,79
58			$\Delta_{\acute{e}poc}$: 10	0,75	0,76±0,02	0,74	0,80

Capítulo 4. Resultados y discusión

ID _{TA}	Planificador	TA inicial	Otros	Mediana	Media±Std	Mínimo	Máximo
59			$\Delta_{\text{époc}}: 15$	0,75	0,76±0,02	0,74	0,79

Finalmente, se evaluó el rendimiento empleando un optimizador. La Figura 81 muestra el DC obtenido durante el proceso de validación cruzada para cada uno de los optimizadores evaluados, así como la Tabla 34 que muestra las métricas obtenidas en cada caso. Los mejores resultados se obtuvieron con el optimizador Adam, (ID_{TA}: 62) siendo el DC medio obtenido de 94,29% (DC min: 92,40% - DC máx: 95,71%), también ligeramente inferior al que se obtuvo con la tasa de aprendizaje constante.

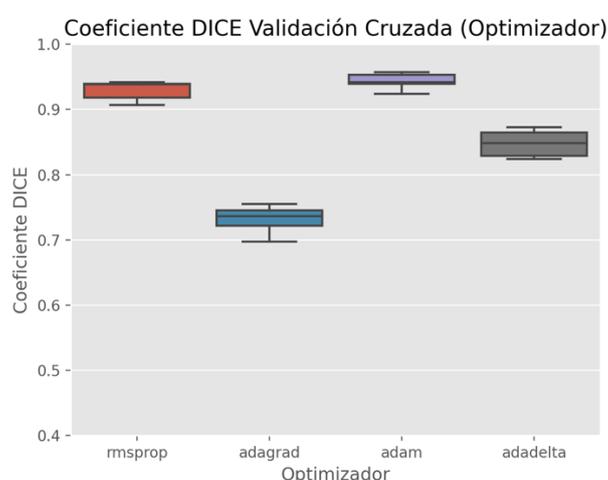


Figura 81. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación sagital: Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos optimizadores de la tasa de aprendizaje.

Tabla 34. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación sagital: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintos optimizadores de tasa de aprendizaje. En negrita se resalta el optimizador que mejores resultados ofreció.

ID _{TA}	Optimizador	Mediana	Media±Std	Mínimo	Máximo
60	RMSprop	0,94	0,93±0,01	0,91	0,94
61	Adagrad	0,74	0,73±0,02	0,70	0,75
62	Adam	0,94	0,94±0,01	0,92	0,96
63	Adadelta	0,85	0,85±0,02	0,82	0,87

Tanto por el elevado valor de DC que se obtuvo sobre el conjunto de datos de validación (DC medio: 94,32%) como por la estabilidad observada a lo largo de las 5 iteraciones del proceso de validación cruzada (DC min: 93,18% y DC máx: 95,68%), para la resolución de este problema se seleccionó una tasa de aprendizaje constante de 1 (ID_{TA}: 1).

Capítulo 4. Resultados y discusión

4.3.1.3.2. Función de pérdidas

A continuación, una vez seleccionada de tasa de aprendizaje, se realizaron los experimentos necesarios para la selección de la función de pérdidas. La Figura 82 muestra los resultados del proceso de validación cruzada para cada una de las funciones de pérdidas evaluadas. Además, la Tabla 35 ofrece las métricas obtenidas en cada caso. Al igual que ocurrió con el modelo en orientación coronal, se obtuvo que con la entropía cruzada binaria ($ID_{FP}: 1$) y la entropía cruzada binaria con coeficiente Dice ($ID_{FP}: 3$), la estabilidad en el entrenamiento no era buena. Esto mismo ocurrió con una función de pérdidas Tversky con $\alpha = 0,7$ ($ID_{FP}: 7$) y $\alpha = 0,9$ ($ID_{FP}: 8$). Con el resto de las funciones de pérdidas se obtuvieron buenos resultados tanto considerando el valor de DC como la estabilidad entre las interacciones. Finalmente, los mejores resultados se lograron con la función de pérdidas Combo con $\alpha = 0,3$ y $\beta = 0,1$ ($ID_{FP}: 14$) con DC medio: 96,51%, DC min: 95,87% y DC máx: 97,41%.

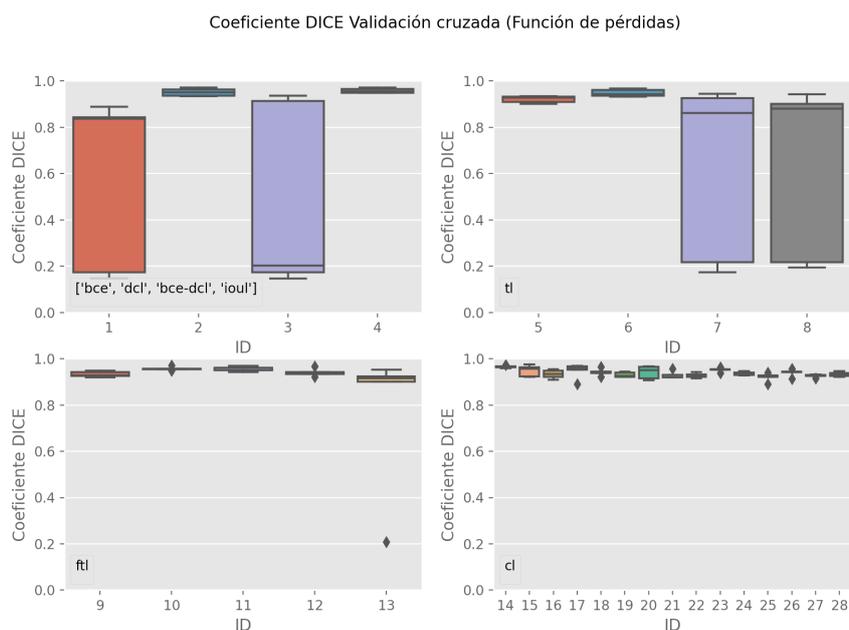


Figura 82. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación sagital: Coeficiente Dice en el conjunto de datos de validación durante el proceso de validación cruzada al emplear distintas funciones de pérdidas en la optimización del modelo durante el proceso de entrenamiento. El eje x representa el identificador del modelo entrenado.

Capítulo 4. Resultados y discusión

Tabla 35. Modelo de segmentación hepática en secuencias de RM de perfusión con imágenes en orientación sagital: valores de mediana, media, desviación estándar (Std), mínimo y máximo del Coeficiente Dice obtenido a lo largo del proceso de validación cruzada en el conjunto de datos de validación al emplear distintas funciones de pérdidas. En negrita se resalta la función de pérdidas que mejores resultados ofreció y que se empleó para el entrenamiento del modelo final.

ID _{FP}	Función de pérdidas	Otros parámetros	Mediana	Media±Std	Mínimo	Máximo
1	BCE	-	0,84	0,58±0,34	0,15	0,89
2	DL	-	0,95	0,95±0,02	0,93	0,97
3	BCE_DL	-	0,20	0,47±0,37	0,15	0,94
4	IoUL	-	0,95	0,96±0,01	0,95	0,97
5	TL	$\alpha = 0,1$	0,93	0,92±0,01	0,90	0,93
6		$\alpha = 0,3$	0,94	0,95±0,01	0,93	0,97
7		$\alpha = 0,7$	0,86	0,62±0,35	0,17	0,95
8		$\alpha = 0,9$	0,88	0,63±0,35	0,19	0,94
9	FTL	$\alpha = 0,1$ $\gamma = 4/3$	0,93	0,93±0,01	0,92	0,95
10		$\alpha = 0,3$ $\gamma = 4/3$	0,96	0,96±0,01	0,95	0,97
11		$\alpha = 0,5$ $\gamma = 4/3$	0,95	0,95±0,01	0,94	0,97
12		$\alpha = 0,7$ $\gamma = 4/3$	0,94	0,94±0,02	0,92	0,97
13		$\alpha = 0,9$ $\gamma = 4/3$	0,91	0,78±0,29	0,21	0,95
14	CL	$\alpha = 0,3$ $\beta = 0,1$	0,96	0,97±0,01	0,96	0,97
15		$\alpha = 0,3$ $\beta = 0,3$	0,96	0,95±0,02	0,92	0,98
16		$\alpha = 0,3$ $\beta = 0,5$	0,93	0,93±0,02	0,91	0,95
17		$\alpha = 0,3$ $\beta = 0,7$	0,96	0,95±0,03	0,89	0,97
18		$\alpha = 0,3$ $\beta = 0,9$	0,94	0,94±0,01	0,92	0,97
19		$\alpha = 0,5$ $\beta = 0,1$	0,93	0,93±0,01	0,92	0,94
20		$\alpha = 0,5$ $\beta = 0,3$	0,95	0,94±0,02	0,91	0,97

Capítulo 4. Resultados y discusión

ID _{FP}	Función de pérdidas	Otros parámetros	Mediana	Media±Std	Mínimo	Máximo
21		$\alpha = 0,5$ $\beta = 0,5$	0,93	0,93±0,01	0,92	0,96
22		$\alpha = 0,5$ $\beta = 0,7$	0,93	0,93±0,01	0,92	0,94
23		$\alpha = 0,5$ $\beta = 0,9$	0,95	0,95±0,01	0,94	0,96
24		$\alpha = 0,7$ $\beta = 0,1$	0,94	0,94±0,01	0,93	0,95
25		$\alpha = 0,7$ $\beta = 0,3$	0,92	0,92±0,02	0,89	0,94
26		$\alpha = 0,7$ $\beta = 0,5$	0,94	0,94±0,01	0,91	0,96
27		$\alpha = 0,7$ $\beta = 0,7$	0,93	0,93±0,01	0,92	0,93
28		$\alpha = 0,7$ $\beta = 0,9$	0,93	0,93±0,01	0,92	0,95

4.3.1.4. Resumen de los modelos seleccionados

Para todos los casos podemos ver la importancia del ajuste óptimo de los hiperparámetros. Podemos ver como el DC puede aumentar significativamente con el correcto ajuste de la función de pérdidas.

Las configuraciones finalmente seleccionadas para cada uno de los modelos fueron:

- Orientación transversal: tasa de aprendizaje variable empleando un planificador dinámico ($\Delta_{\text{épocas}}: 5$) y una tasa de aprendizaje inicial de 0,1; función de pérdidas Focal Tversky con $\alpha = 0,5$ (DC medio: 98,62%; DC min: 96,68%; DC máx: 99,46%).
- Orientación coronal: tasa de aprendizaje variable empleando el optimizador Adam; función de pérdidas Combo con $\alpha = 0,5$ y $\beta = 0,3$ (DC medio: 98,04%, DC min: 97,54% y DC máx: 98,52%).
- Orientación sagital: tasa de aprendizaje constante de 1; función de pérdidas Combo con $\alpha = 0,3$ y $\beta = 0,1$ (DC medio: 96,51%, DC min: 95,87% y DC máx: 97,41%).

4.3.2. Test del modelo de segmentación hepática

Una vez seleccionados los hiperparámetros, se entrenó el modelo final. Para ello, se mantuvieron las particiones de la validación cruzada y se entrenaron cinco modelos independientes para construir un ensamblado de modelos. En la Tabla 36 se encuentra la comparativa de las métricas obtenidas para cada una de las orientaciones independientes y la combinación de todas ellas (multiplanar) sobre el conjunto de datos de test (n=33).

Tabla 36. Resultados obtenidos en el modelo final formado por cinco modelos independientes entrenados con las diferentes particiones del proceso de entropía cruzada. Se muestran cada una de las métricas definidas para cada una de las orientaciones (transversal, coronal y sagital), así como la combinación de todas ellas (multiplanar). En negrita se destaca la aproximación que mejores resultados ofrece para cada métrica.

Métrica	Transversal		Coronal		Sagital		Multiplanar	
	Mediana	Media±Std	Mediana	Media±Std	Mediana	Media±Std	Mediana	Media±Std
DC	0,97	0,96±0,02	0,95	0,95±0,02	0,95	0,94±0,03	0,97	0,96±0,02
IoU	0,93	0,94±0,03	0,91	0,91±0,03	0,90	0,89±0,05	0,94	0,93±0,03
Sensib.	0,97	0,95±0,04	0,95	0,94±0,03	0,93	0,91±0,05	0,96	0,95±0,03
Especif.	1,00	1,00±0,00	1,00	1,00±0,00	1,00	1,00±0,00	1,00	1,00±0,00
Precisión	0,97	0,97±0,01	0,96	0,96±0,02	0,97	0,96±0,01	0,98	0,97±0,01
FDR	0,03	0,03±0,01	0,04	0,04±0,02	0,03	0,03±0,01	0,02	0,03±0,01
RVD	0,01	0,03±0,04	0,02	0,03±0,03	0,04	0,05±0,05	0,01	0,02±0,03
ASSD (mm)	0,89	1,03±0,57	1,19	1,48±0,83	1,46	1,59±0,64	0,83	0,97±0,44
MSD (mm)	21,18	23,18±10,24	25,90	30,97±16,81	24,50	28,13±10,43	20,59	22,96±8,27

Como se puede observar en la Tabla 36 las mejores métricas se obtuvieron para la aproximación multiplanar. Este modelo, a excepción de la sensibilidad donde los resultados son ligeramente superiores en el modelo entrenado en la orientación transversal, ofrece los mejores resultados para el resto de las métricas consideradas.

En la Figura 83 encontramos un ejemplo en el que se muestran las ventajas de emplear una aproximación multiplanar. Se puede ver cómo hay una región en el lóbulo hepático izquierdo que el modelo transversal no segmenta satisfactoriamente, sin embargo, al emplear la combinación de los tres modelos, se obtiene una segmentación final más precisa.

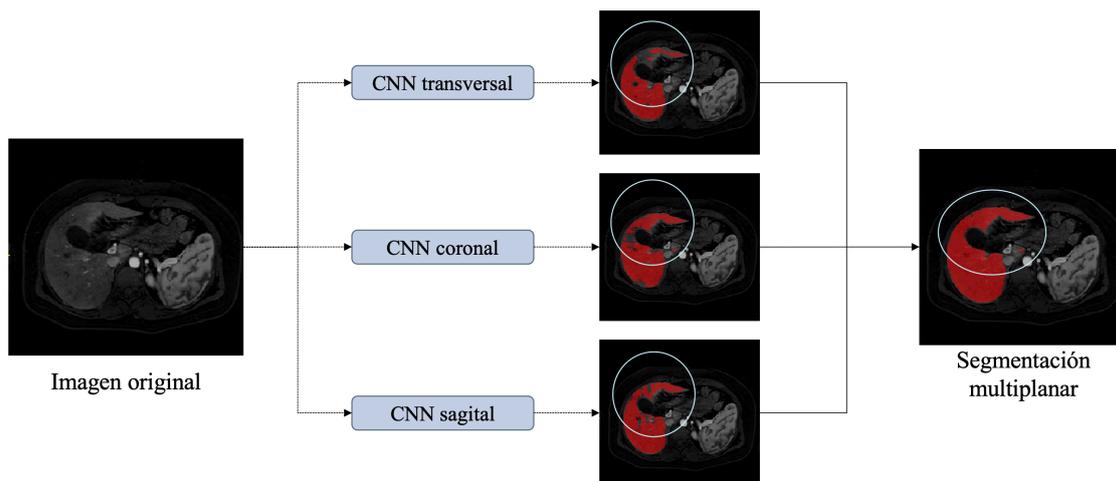


Figura 83. Ejemplo de mejora en la segmentación final obtenida mediante el uso de una aproximación multiplanar donde se combina la información proveniente de los modelos entrenados en las tres orientaciones (transversal, coronal y sagital). Destacado con un círculo se observa una región en el lóbulo hepático izquierdo en la que no se segmenta correctamente el parénquima hepático.

Del modelo final seleccionado, se analizó la correlación entre los volúmenes hepáticos medidos con la máscara de referencia y la predicha por la CNN (Figura 84). La prueba de Shapiro-Wilk mostró que la distribución de volúmenes seguía una distribución normal ($p > 0,05$), por lo tanto, se empleó el coeficiente de Pearson para el estudio de la correlación, obteniendo un valor de 0,99 ($p < 0,001$). Esto también se puede apreciar de manera visual observando los histogramas de ambos volúmenes en la Figura 84. Por otro lado, se realizó el análisis Bland-Altman para comparar las diferencias entre ambos volúmenes (Figura 84). Este análisis muestra un sesgo bajo (0,03 L) y límites de acuerdo estrechos (-0,07 - 0,13 L).

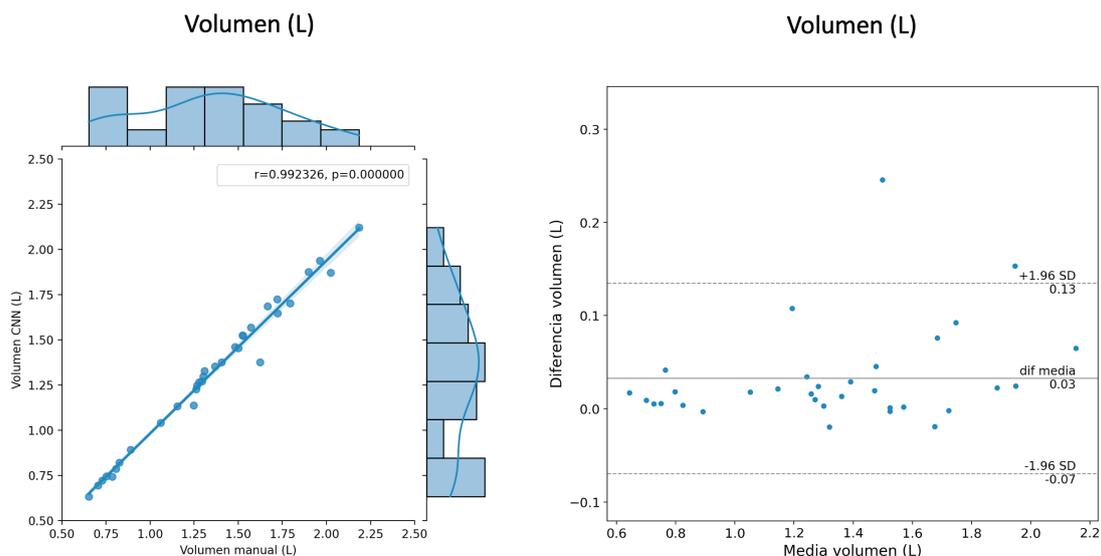


Figura 84. Análisis de correlación entre el volumen hepático calculado sobre la segmentación de referencia (manual) y la predicha por la CNN sobre una imagen de RM de perfusión. A la izquierda vemos el diagrama de pares entre ambos volúmenes; en los ejes superior y derecho se encuentran los diagramas de los volúmenes de la segmentación manual y predicha, respectivamente. A la derecha se muestra el diagrama Bland-Altman.

Al igual que se realizó para la cuantificación de PDFF y $R2^*$ en secuencias MECSE de RM, se evaluó la correlación obtenida en los parámetros obtenidos del análisis de perfusión semicuantitativo al emplear la máscara realizada manualmente o la obtenida por la CNN. Tras realizar la prueba de Shapiro-Wilk, se vio que la distribución de $IAUC_{60}$ no seguía una distribución normal, por lo que se empleó el coeficiente de Spearman para el estudio de la correlación, mientras que el resto de los biomarcadores (pendiente inicial y pico de contraste) seguían una distribución normal, por lo que, para éstos, se hizo uso del coeficiente de Pearson.

La Figura 85 muestra la relación para el $IAUC_{60}$, la pendiente inicial y el pico de contraste entre los valores cuantificados sobre la máscara generada manualmente y los cuantificados sobre la máscara generada por la CNN. Para todos ellos, se puede observar un coeficiente de correlación elevado ($r = 1$; $p < 0,001$), también reflejado en la similitud entre los histogramas de ambas distribuciones para cada caso. Por otro lado, el análisis de Bland-Altman muestra un sesgo nulo para cada caso y límites de acuerdo muy cercanos a cero.

Capítulo 4. Resultados y discusión

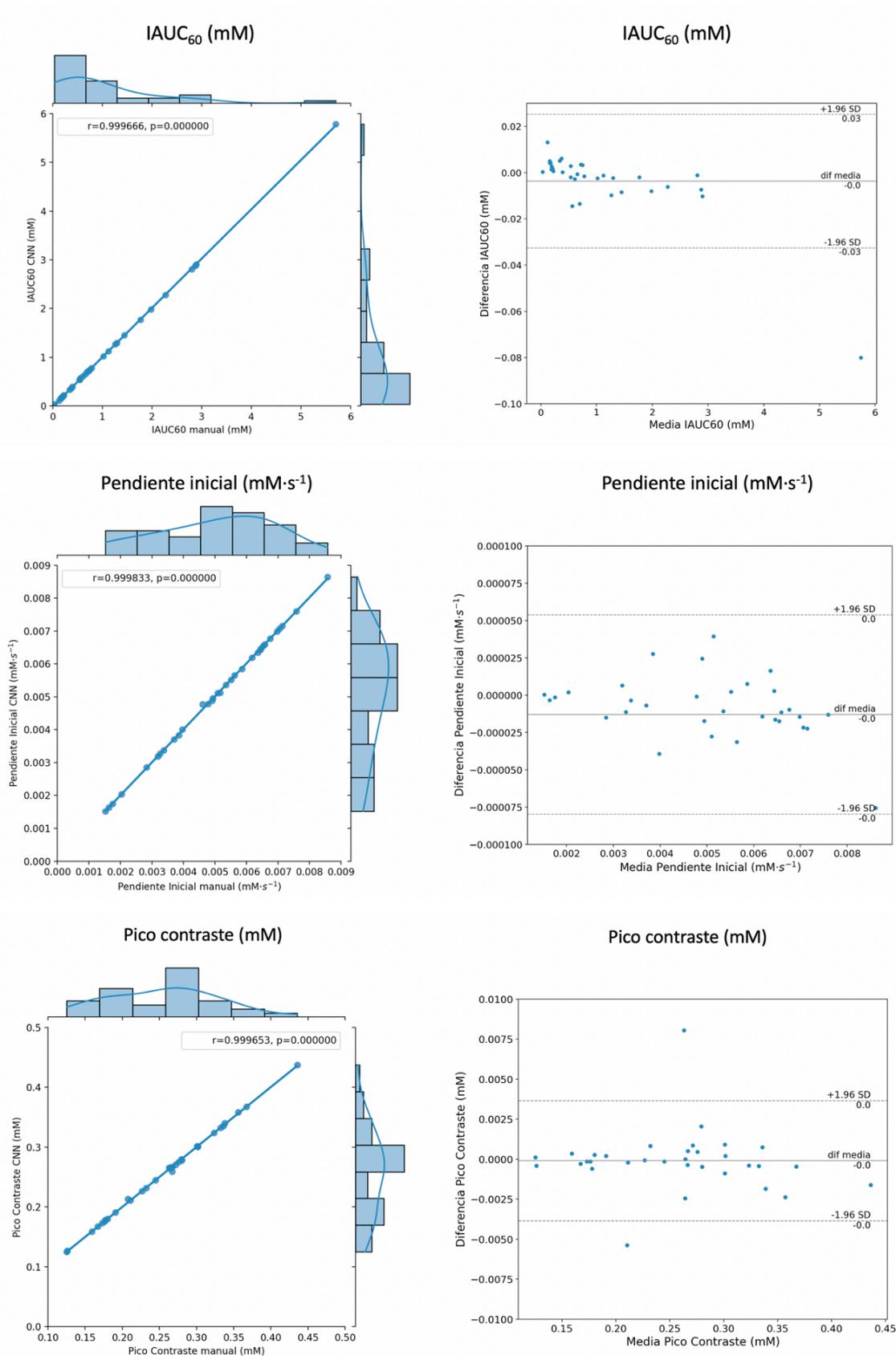


Figura 85. Análisis de correlación entre el IAU₆₀ (superior), pendiente inicial (medio) y pico de contraste (inferior) calculado sobre la segmentación de referencia (manual) y la predicha por la CNN sobre una imagen de RM de perfusión. A la izquierda vemos el diagrama de pares entre ambas máscaras; en los ejes superior y derecho se encuentran los diagramas de los biomarcadores cuantificados sobre la segmentación manual y predicha, respectivamente. A la derecha se muestra el diagrama Bland-Altman.

Capítulo 4. Resultados y discusión

Finalmente, se evaluó el Error Relativo (ER) cometido al emplear la segmentación automática en lugar de la manual (Tabla 37). Para ello se hizo uso de la mediana de cada uno de los biomarcadores. Podemos observar que, en todos los casos, la mediana del ER se encuentra por debajo del 1%.

Tabla 37. ER obtenido en la cuantificación de los biomarcadores de perfusión haciendo uso de la máscara obtenida con la segmentación automática comparado que los obtenidos a través de la segmentación manual.

Biomarcador	Mediana (%)	Media±Std (%)
IAUC₆₀	0,47	1,13±1,86
Pendiente inicial	0,23	0,41±0,63
Pico de contraste	0,17	0,40±0,65

Por lo tanto, la metodología diseñada nos permite implementar un algoritmo automático para la cuantificación de los descriptores de perfusión hepáticos en secuencias de RM de perfusión abdominales.

4.3.3. Discusión

Se ha desarrollado un método de segmentación automática hepática a partir de imágenes de RM de perfusión. En este caso, aprovechado que estamos trabajando con imágenes con mejor resolución entre cortes, se ha abordado la resolución del problema mediante la implementación de una solución formada por tres modelos entrenados en las tres direcciones ortogonales principales: transversal, coronal y sagital. Esto ha demostrado ofrecer mejores resultados que emplear un único modelo entrenado en cualquiera de las tres orientaciones.

Actualmente, encontramos diferentes desarrollos centrados en la segmentación hepática en secuencias de RM de perfusión [128,129]. En la solución propuesta encontramos resultados (DC: 0,96±0,02) comparables con los presentados en otros trabajos (DC medio de 0,95 en [128] y 0,95±0.02 en [129]). En este trabajo, además encontramos un buen rendimiento del modelo al considerar casos de dos fabricantes diferentes (Siemens y GE) demostrando, por lo tanto, la capacidad de generalización a otros fabricantes.

Este trabajo, se puede considerar el primero en calcular el error cometido al realizar la cuantificación de biomarcadores de imagen como el IAUC₆₀, la pendiente inicial o el pico de contraste, demostrando una elevada correlación ($r>0,99$) para todos los casos al comparar los valores de mediana obtenidos con la segmentación manual de referencia y la obtenida automáticamente por la CNN, con un ER inferior al 0,5% para todos los casos.

Capítulo 4. Resultados y discusión

Obtener un buen rendimiento en la segmentación del tejido hepático es de gran importancia al poder establecer la región de referencia para la posterior segmentación de lesiones en dicho órgano. De esto modo, se permite reducir la región a analizar por el modelo desarrollado para la detección/segmentación de lesiones a únicamente la región hepática.

Además, este método de segmentación tiene un potencial por sí mismo al permitir cuantificar el volumen hepático al ser éste un biomarcador de interés en el estudio de determinadas enfermedades, como por ejemplo en planificación quirúrgica de resección de lesiones hepáticas o en enfermedades hematológicas para estudiar la magnitud de la hepatomegalia. En este sentido, el método desarrollado presenta una correlación elevada ($r > 0,99$) entre el volumen predicho y el obtenido con la segmentación manual con una diferencia media de 0,03L.

4.4. Conclusiones generales

Durante la consecución de los distintos casos de uso y el entrenamiento y ajuste de hiperparámetros de cada uno de los modelos de segmentación se han observado una serie de aspectos que se exponen a continuación:

- En general, durante el ajuste de la tasa de aprendizaje, se ha observado que la tasa de aprendizaje constante que ofrecía mejores resultados en cada modelo es, a su vez, la que ofrecía los mejores resultados al ser empleada como tasa de aprendizaje inicial en los distintos planificadores. Esto es de gran interés ya que, en caso de validarse dicho efecto en modelos adicionales, podría acelerar el proceso de entrenamiento ya que, en la evaluación de los planificadores, sólo se evaluarían las configuraciones en las que se empleen como tasa de aprendizaje inicial la que mejor rendimiento haya mostrado en la evaluación de la tasa de aprendizaje constante.
- En varios de los modelos se ha observado una mayor inestabilidad de los entrenamientos al emplear un planificador lineal, polinomial, exponencial y algunas de las configuraciones del planificador por saltos al emplear una tasa de aprendizaje inicial de 1.
- Dada una tasa de aprendizaje inicial, al emplear un planificador polinomial, en todos los casos de uso se han obtenido los mejores resultados al emplear un orden 2, seguido de orden 3, orden 5 y, finalmente, con los peores resultados, orden 7.
- En cuanto al planificador por saltos, en todos los casos de uso se cumple que los mejores resultados se obtuvieron con un $\Delta_{\text{épocas}}: 25$, seguido de un $\Delta_{\text{épocas}}: 10$ y, finalmente, un $\Delta_{\text{épocas}}: 0,5$. En cuanto a la tasa, se han observado mejores resultados al emplear una tasa de 0,5 que con una tasa de 0,25.
- El planificador de un ciclo no consiguió que el modelo converja en 2 de los modelos entrenados.
- El planificador dinámico ofreció resultados similares en todas sus configuraciones.
- Como conclusión en cuanto a los planificadores, se ha observado que, en todos los casos de uso entrenados para resolver un problema de segmentación, el planificador dinámico es el que mejores resultados ha ofrecido. Por lo tanto, si se desea acelerar el proceso de entrenamiento y no extender el proceso de experimentación, se recomienda evaluar dicho planificador, en cualquiera de sus

Capítulo 4. Resultados y discusión

configuraciones, con la tasa de aprendizaje constante que mejores resultados haya ofrecido.

- En cuanto a los optimizadores, no se ha observado ningún patrón que nos permita concluir cual de todos los evaluados ofrece mejores resultados, sin embargo, se ha observado que el optimizador Adagrad, en todos los casos de uso, ha ofrecido los peores resultados.
- En cuanto a la función de pérdidas, en algunos entrenamientos, se ha observado un comportamiento inestable al emplear BCE, BCE_DL y TL con α de 0,7 y 0,9.
- Al emplear una función de pérdidas FTL, se ha observado que con un $\alpha = 0,9$ se obtuvieron los peores resultados en todos los casos.
- Al emplear una función de pérdidas CL, se ha observado que con un $\alpha = 0,3$, se obtuvieron los mejores resultados en todos los casos, sin embargo, el valor de β ha variado más.
- En los modelos entrenados, se han observado los mejores resultados al emplear funciones de pérdidas basadas en FTL o CL.
- Como conclusión general en cuanto a las funciones de pérdidas, se recomienda emplear las funciones de pérdidas FTL o CL. Para reducir el número de experimentos a realizar, al evaluar la FTL se podría excluir la evaluación de los modelos que emplean un $\alpha = 0,9$, por otro lado, al evaluar la CL, se podrían evaluar únicamente aquellos que empleen un $\alpha = 0,3$.

Capítulo 5. Implementación en práctica clínica

En este capítulo se muestran los pasos y desarrollos realizados para la utilización de los diferentes algoritmos obtenidos en un entorno clínico. Desde el desarrollo de un entorno de anotación manual para el etiquetado y edición de las segmentaciones necesarias, hasta la puesta en producción de los métodos de cuantificación automática y la generación de los resultados en formatos estándar.

5.1. Plataforma de análisis de imagen

Todos los desarrollos realizados en la presente tesis doctoral tienen aplicación directa en la práctica clínica. La plataforma Quibim Precision, desarrollada por la empresa Quibim SL (Valencia, España) en la cual se ha desarrollado la presente tesis doctoral, ofrece una plataforma web que permite la centralización, visualización y anotación de estudios de imagen y, además, el análisis de estos mediante un conjunto de algoritmos basados en procesamiento de imagen, modelado matemático e IA. Estos algoritmos permiten la extracción de biomarcadores de imagen asociados a diferentes patologías y áreas anatómicas.

Entre otras cosas, Quibim Precision cuenta con un visor de imágenes DICOM cuya principal aplicación es la anotación de los estudios de imagen mediante la delineación de regiones de interés que se emplearán, a posteriori, para la extracción de los biomarcadores de imagen sobre la región delineada. En los métodos de segmentación automática, como los desarrollados en la presente tesis doctoral, esta herramienta adquiere especial relevancia para, en primer lugar, la visualización de la máscara generada de manera automática y, en segundo lugar, su corrección en caso de que sea necesario. Esto permite que el usuario final tenga el control sobre la región analizada y eliminar la brecha de inseguridad que existe hoy en día en el uso de los métodos basados en IA en los que el clínico no puede participar. Además, la corrección de estas máscaras por parte de los expertos permite aumentar el tamaño de la base de datos etiquetada y, por tanto, mediante el reentrenamiento de los modelos con los nuevos casos, la mejora de su rendimiento.

5.1.1. Desarrollo de un entorno de anotación

Dentro del visor de imágenes DICOM de la plataforma de Quibim Precision encontramos diferentes herramientas de anotación (círculo, rectángulo, polígonos irregulares, crecimiento de regiones, etc.), sin embargo, lo que realmente facilita la labor de anotación

de diferentes regiones es la herramienta del pincel, es decir, poder seleccionar vóxel a vóxel la región de interés.

Por este motivo, el visor se enriqueció con dicha funcionalidad en el contexto de la presente tesis doctoral. Se trata de una aplicación web *zero footprint* basada principalmente en JavaScript como lenguaje de programación. Otros lenguajes de programación empleados son CoffeScript, JADE y Less para la implementación de los controles necesarios en el código principal de la plataforma.

Para llevar a cabo la total implementación del pincel en la plataforma se incorporaron las siguientes funcionalidades:

- Se incluyó un botón que activaba la herramienta del pincel de forma que, tras pulsar en dicho botón, se mostraba un cursor con forma circular que simulaba la brocha del pincel y que se movía con el movimiento del ratón. El tamaño de dicha brocha se podía cambiar tanto haciendo uso del teclado a través de las teclas “-“ y “+“ para disminuir y aumentar el tamaño, respectivamente; como a través de un cuadro de diálogo donde se podía introducir tanto el nuevo tamaño como cambiar el tamaño con un control deslizante.

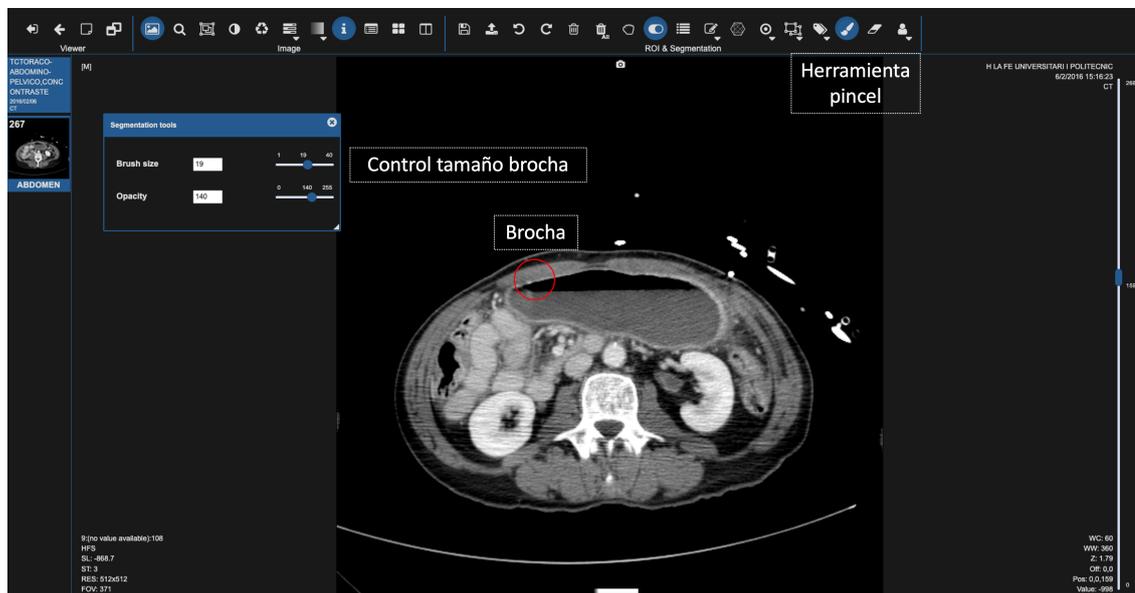


Figura 86. Visor de imágenes DICOM – Herramienta del pincel. Control de la brocha de dibujo.

- Al pulsar con el botón izquierdo del ratón sobre una región, la región dentro del área definida por la brocha se coloreaba del color correspondiente a la etiqueta activa. Mientras se mantenga el botón del ratón activo se continuaba con la segmentación de las regiones por las que desplazaba el ratón.

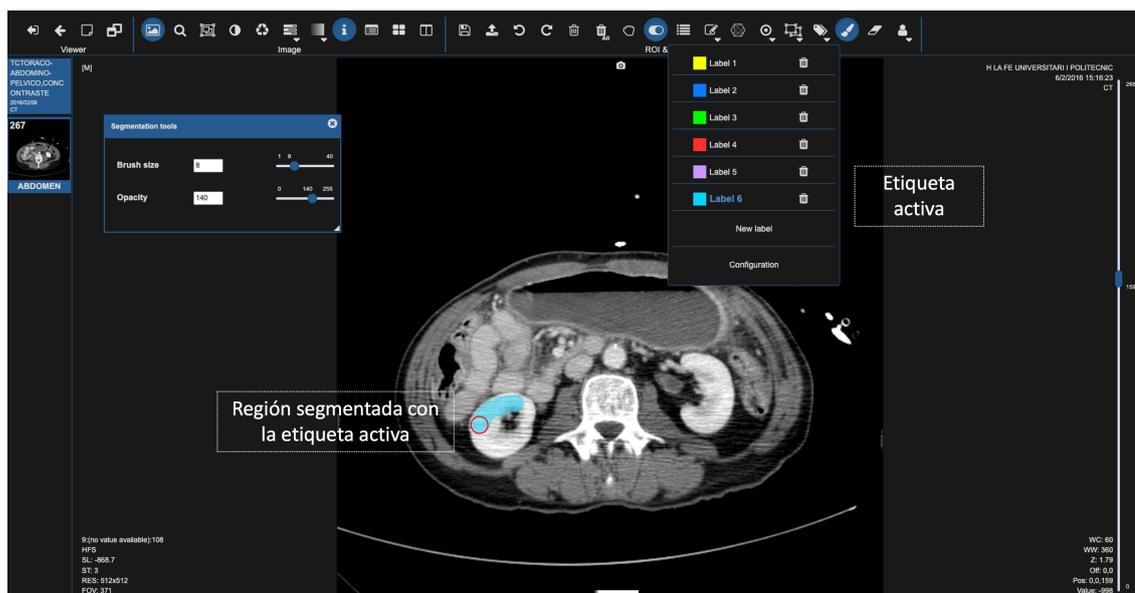


Figura 87. Visor de imágenes DICOM – Herramienta del pincel. Anotación de una región con una etiqueta específica.

- En cambio, si se pulsaba con el botón derecho del ratón, todos los píxeles con región segmentada que se encontraban dentro del área definida por la brocha se borraban. Este mismo efecto se conseguía activando la herramienta de la goma, para la cual se incluyó otro botón.
- Se implementó la posibilidad de ajustar la opacidad de la máscara a través del teclado con las teclas “A” y “D” para disminuir y aumentar la opacidad, respectivamente, y con la tecla “S” para ocultar/mostrar la máscara. Además, se podía hacer uso del cuadro de diálogo de control para, o bien definir la opacidad con un valor entre 0 y 255 (donde 0 es mínima opacidad), o bien emplear un control deslizante para ajustar dicha opacidad.
- Se incorporó la posibilidad de deshacer/rehacer los cambios realizados en la última interacción con el ratón.
- Se implementó un menú de configuración que, para cada etiqueta, permitía:
 - Cambiar el color.
 - Cambiar el nombre.
 - Definir un umbral con valor mínimo y/o máximo, éste limitaba el rango de intensidad sobre el que el pincel actuaba, descartando de la segmentación todos aquellos píxeles que se encontraban fuera del rango establecido. Esta funcionalidad es de especial interés en la anotación de estudios de TC donde las UH de cada tejido son conocidas a priori.

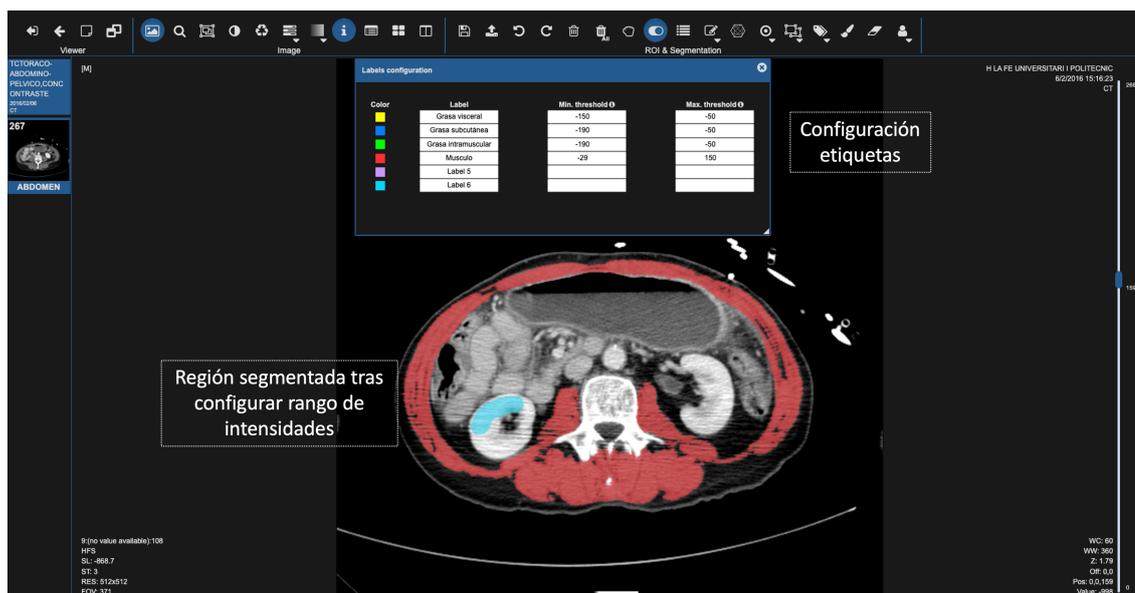


Figura 88. Visor de imágenes DICOM – Herramienta del pincel. Configuración de las etiquetas y definición de límites de intensidad sobre los que dibujar.

Finalmente, una vez se realizaba la anotación, dicha información se almacenaba en la base de datos de la plataforma.

5.1.2. Integración de un módulo de análisis

Para la integración de un módulo de análisis en Quibim Precision es necesario construir la imagen Docker correspondiente, que será lo que se ejecute cuando un usuario desee aplicar un análisis sobre un estudio de imagen.

Docker⁴ es una plataforma de software que simplifica el proceso de creación, ejecución, gestión y distribución de aplicaciones mediante la virtualización del sistema operativo en el que se instala y ejecuta.

Las imágenes Docker se crean gracias a procesos de integración y despliegue continuos (CI/CD del inglés *Continuous Integration/Continuous Delivery*) que siguen el siguiente flujo:

- Todos los módulos de análisis se encuentran en repositorios independientes en Bitbucket⁵, un servicio de alojamiento de código fuente basado en Git, propiedad de Atlassian.
- Cada repositorio se encuentra conectado con Jenkins⁶, una herramienta de código abierto que permite la automatización de flujos de CI/CD. Mediante

⁴ <https://www.docker.com/>

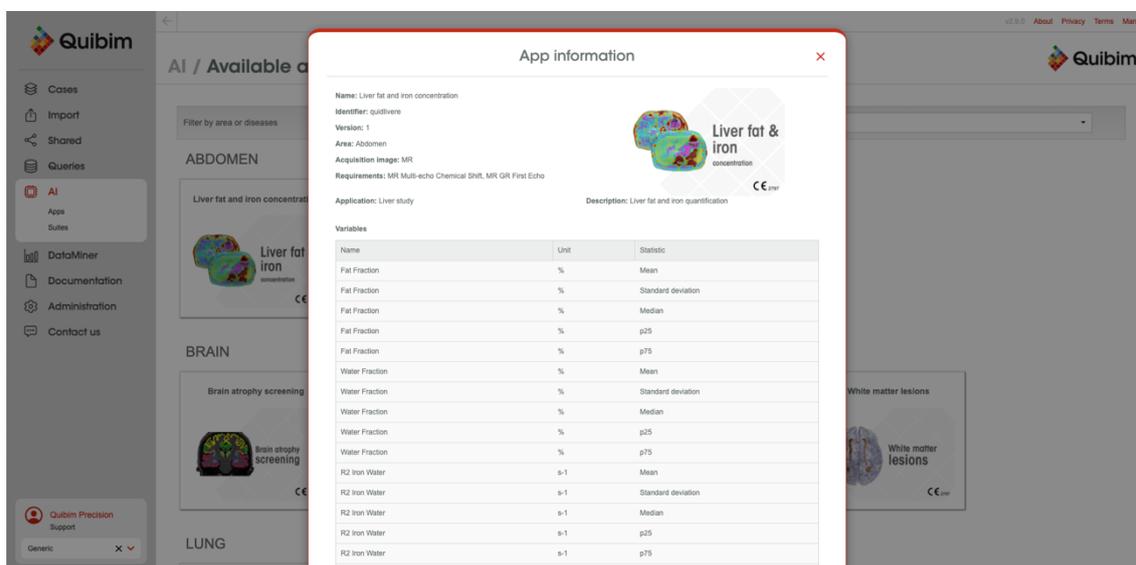
⁵ <https://bitbucket.org/>

⁶ <https://www.jenkins.io/>

Capítulo 5. Implementación en práctica clínica

un archivo llamado *Jenkinsfile* se definen los pasos que se desean ejecutar en este proceso. En todos los módulos integrados se encuentran las etapas correspondientes a (1) la descarga de los pesos de los modelos entrenados, al tratarse de archivos pesados no es recomendable subirlos al repositorio de Git, por este motivo, se almacenan en Microsoft Azure y se descargan mediante acceso SAS (*Single Access Signature*); (2) la realización de testeos unitarios, que se ejecutan para garantizar la retrocompatibilidad del código; (3) la creación de la imagen Docker correspondiente y su subida al registro de imágenes Docker de Quibim, para la creación de dicha imagen Docker, se incluye, dentro del repositorio, un archivo *Dockerfile* que define cada uno de los comandos necesarios a ejecutar para disponer de un entorno con todas las dependencias necesarias para la correcta ejecución del módulo.

Una vez la imagen Docker se encuentra disponible en el registro de imágenes, se da de alta el módulo de análisis en la plataforma de Quibim Precision. Esto se puede realizar a través de la propia interfaz de la plataforma.



The screenshot shows the Quibim Precision platform interface. A modal window titled 'App information' is open, displaying details for the 'Liver fat and iron concentration' application. The modal includes a logo for 'Liver fat & iron concentration' and a list of variables with their units and statistics.

Name	Unit	Statistic
Fat Fraction	%	Mean
Fat Fraction	%	Standard deviation
Fat Fraction	%	Median
Fat Fraction	%	p25
Fat Fraction	%	p75
Water Fraction	%	Mean
Water Fraction	%	Standard deviation
Water Fraction	%	Median
Water Fraction	%	p25
Water Fraction	%	p75
R2 Iron Water	s-1	Mean
R2 Iron Water	s-1	Standard deviation
R2 Iron Water	s-1	Median
R2 Iron Water	s-1	p25
R2 Iron Water	s-1	p75

Figura 89. Módulo de análisis para la segmentación automática del hígado y posterior cuantificación de la PDF y R2* en secuencia MECSE de RM integrado en la plataforma de Quibim Precision.

Finalmente, los módulos de análisis generan un informe que resumen la información principal obtenida del análisis. Estos informes se diseñan con plantillas en formato JADE⁷ que, posteriormente en la plataforma, se renderizan en formato PDF. En el Anexo I

⁷ <https://jade-lang.com/>

podemos encontrar un ejemplo del informe generado para el módulo de análisis de segmentación y cuantificación hepática en secuencias MECSE.

5.1.3. Generación de los archivos segmentación para su visualización

El visor trabaja con un formato interno basado en archivos JSON para el almacenamiento de las máscaras de segmentación. De este modo, los distintos módulos de análisis desarrollados se han adaptado para que almacenen la máscara en este formato interno permitiendo su visualización en el visor.

En la Figura 90 podemos ver un ejemplo de cómo el usuario observa la máscara generada por un modelo de IA en el visor, en este caso, de la máscara generada por el modelo de segmentación automática de hígado en secuencias de RM de perfusión. Como se puede observar, se genera una etiqueta asociada a la región de interés (hígado en el caso de la Figura 90) y, empleando la herramienta del pincel, descrita anteriormente, se puede editar en caso de que haya algún error de segmentación permitiendo, además, el relanzado del análisis con la nueva segmentación en caso de que el usuario lo desee.

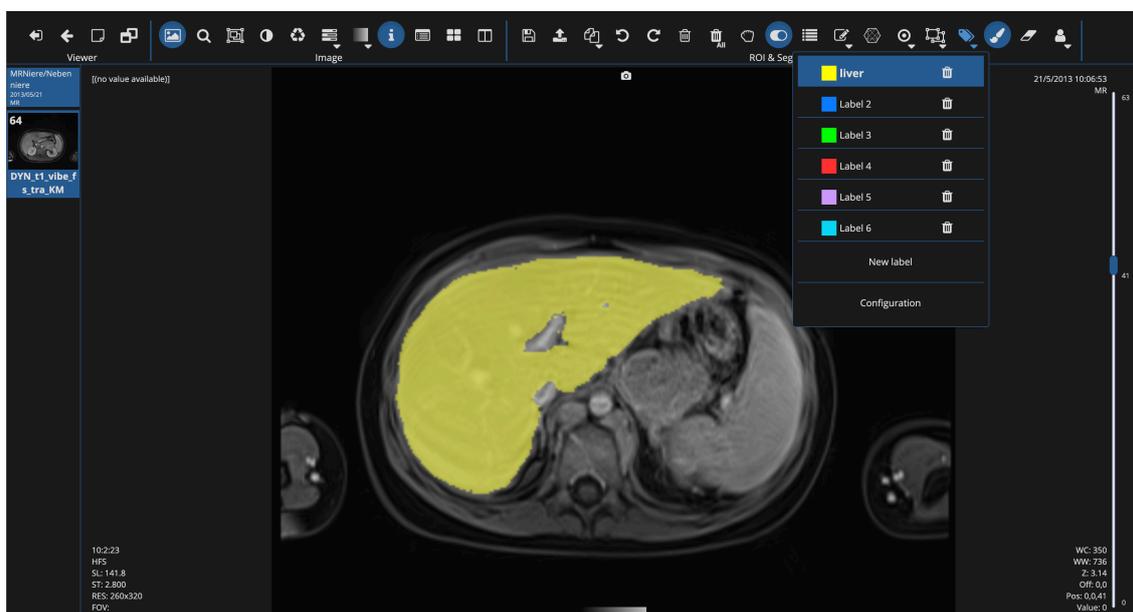


Figura 90. Visualización en el visor de imágenes DICOM de la máscara obtenida en un ejemplo de segmentación hepática sobre una serie de RM de perfusión. Se puede observar la etiqueta “liver” (hígado) dentro de la lista de ROIs disponibles.

5.2. Estándares de segmentación

En este apartado se exponen los diferentes estándares disponibles en cuanto a archivos para el almacenamiento de máscaras de segmentación en el ámbito de la imagen médica radiológica.

El estándar DICOM (del inglés, *Digital Imaging and Communications in Medicine*), propuesto por la NEMA (*National Electrical Manufacturers Association*), es el estándar principal, tanto para la comunicación como para el almacenamiento de las imágenes médicas radiológicas. En la presente tesis doctoral nos centramos en el estándar DICOM en cuanto a formato de almacenamiento de archivos se refiere.

Las imágenes DICOM están formadas por dos partes: la cabecera, que contiene una serie de metadatos relacionados con el protocolo de adquisición empleado, información del paciente, información del centro médico, entre otros; y la imagen, donde se encuentran los valores de cada uno de los píxeles/vóxeles que forman la propia imagen. Para el almacenamiento de máscaras de segmentación surgen las extensiones DICOM RT Structure⁸ (DICOM RT Struct) y, posteriormente, DICOM Segmentation⁹ (DICOM Seg). El primero surge ante la necesidad de definir un estándar para el almacenamiento de los contorneos realizados para la planificación de la radioterapia que va a recibir un paciente. La principal diferencia entre ambos formatos es la forma en la que se almacenan las segmentaciones. El formato DICOM Seg permite almacenar la información exacta de la segmentación realizada, esto es, se almacena la información píxel a píxel y, por tanto, no hay pérdida de información. Sin embargo, el formato RT Struct almacena, únicamente, determinados puntos de los contornos de las regiones delineadas, de este modo, previo al almacenamiento de la máscara de segmentación, es necesario realizar una conversión de máscara binaria a contornos, por lo que puede sufrir de pérdida de información en la conversión.

5.2.1. DICOM RT Structure

La extensión DICOM RT Struct está formada por una serie de metadatos, similares a los que encontramos en cualquier archivo DICOM, y a los que se añade una sección concreta donde se almacena la información de cada ROI con sus contornos asociados. En la Figura

⁸ http://dicom.nema.org/dicom/2013/output/chtml/part03/sect_A.19.html

⁹ http://dicom.nema.org/medical/Dicom/2018d/output/chtml/part03/sect_C.8.20.html

91 podemos observar cómo se relaciona los contornos almacenados en dicho archivo con la serie, el estudio y el paciente original. De este modo, en un archivo de este tipo encontraremos los identificadores únicos del paciente, estudio y serie concreta sobre la que se ha realizado la segmentación.

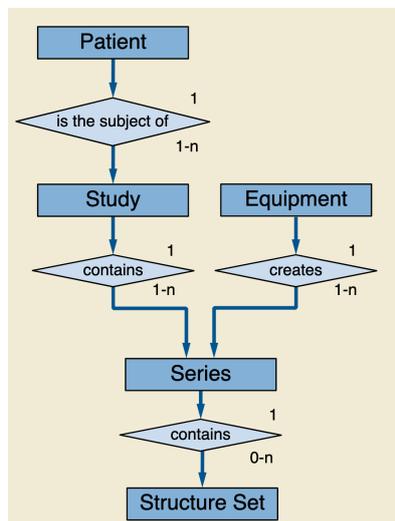


Figura 91. Relación entre el paciente, estudio, series y archivos de segmentación (Fuente imagen: DICOM NEMA).

Los archivos DICOM RT Struct se pueden identificar por la presencia, en la cabecera, del valor 'RTSTRUCT' en el campo relacionado con la modalidad (0008,0060) y por la presencia del campo *ROI Contour Sequence* (3006,0039) que va a contener la información correspondiente a cada uno de los contornos delineados y que permitirán la reconstrucción de la máscara de segmentación.

Por lo tanto, para poder almacenar la máscara de segmentación en este formato es necesario convertir las máscaras binarias a contornos. Además, las coordenadas no se almacenan como índices de los vóxeles dentro de la imagen sino como coordenadas físicas en el espacio de coordenadas del paciente. Estas coordenadas se calculan teniendo en cuenta los campos de la cabecera *Image Position Patient* (0020,0032), que contiene las coordenadas (x,y,z) del vóxel superior izquierdo (centro del primer vóxel transmitido) de la imagen en mm; *Image Orientation Patient* (0020,0037), que indica la dirección de los cosenos de la primera fila y columna con respecto al paciente; y *Pixel Spacing* (0028,0030), que define la distancia física en el paciente entre dos píxeles, en mm, tanto de las filas como de las columnas. Toda esta información permite construir lo que se conoce como matriz afín y que permitirá convertir los índices de un vóxel dado a coordenadas físicas. La matriz afín se construye de acuerdo con la Ecuación 64.

$$Matriz\ afín = \begin{bmatrix} r_x\Delta_x & c_x\Delta_y & \frac{x_n - x_1}{n - 1} & x_1 \\ r_y\Delta_x & c_y\Delta_y & \frac{y_n - y_1}{n - 1} & y_1 \\ r_z\Delta_x & c_z\Delta_y & \frac{z_n - z_1}{n - 1} & z_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (64)$$

Donde r_x, r_y, r_z son los cosenos de la primera fila; c_x, c_y, c_z son los cosenos de la primera columna; Δ_x e Δ_y es el tamaño del píxel a nivel de columna y fila, respectivamente; x_1, y_1, z_1 son las coordenadas físicas del vóxel superior izquierdo del primer corte; x_n, y_n, z_n son las coordenadas físicas del vóxel superior izquierdo del último corte; y n es el número total de cortes que tiene la imagen.

Una vez convertida la máscara a contornos se almacenan en el objeto DICOM.

En la presente tesis doctoral, se ha diseñado e implementado un algoritmo que es capaz de convertir las máscaras de segmentación generadas a formato DICOM RT Struct, empleando, principalmente, la librería *pydicom*¹⁰. En el Anexo II encontramos una lista con todos los campos de la cabecera DICOM necesario para implementar dicho objeto, así como el valor asignado a cada uno de ellos.

5.2.2. DICOM Segmentation

El formato DICOM Seg se diseñó para el almacenamiento específico de máscaras de segmentación. Su relación con el paciente, estudios y series de imagen es la misma que para el objeto DICOM RT Struct (Figura 91). Para identificar este tipo de imágenes también podemos emplear el metadato asociado a la modalidad (0008, 0060) que toma el valor ‘SEG’. El formato DICOM Seg, se aproxima más al modo de almacenamiento de una imagen ya que cuenta tanto con una serie de metadatos y la matriz que compone la máscara de segmentación. En este caso, en series volumétricas, toda la matriz se almacena en un único archivo DICOM. A diferencia del formato RT Struct, no es necesario aplicarle una transformación a la máscara, se puede almacenar como matriz directamente, por lo tanto, se reducen, a su vez, las pérdidas que se pueden generar en la conversión de las máscaras a contornos. Al igual que para el formato DICOM RT Struct se ha desarrollado un algoritmo que permite la conversión de las máscaras generadas a un formato DICOM Seg haciendo uso de la librería *pydicom*. En el Anexo II podemos

¹⁰ <https://pydicom.github.io/>

encontrar una lista con los campos de la cabecera DICOM necesarios para construir un objeto DICOM Seg así como los valores asignados a cada uno de ellos.

5.2.3. Otros formatos

Otros formatos comúnmente empleados para el almacenamiento de máscaras de segmentación son el formato Analyze, NIFTI (*Neuroimaging Informatics Technology Initiative*), MetaImage y NRRD (*Nearly Raw Raster Data*). Estos formatos se emplean, mayoritariamente, en el ámbito de la investigación, ya que simplifican, en gran medida, la lectura y escritura de éstos.

Dentro del algoritmo de conversión de máscaras a formatos estándares que se desarrolló, también se incorporó la posibilidad de convertir las máscaras a formato NIFTI.

5.2.3.1. Neuroimaging Informatics Technology Initiative (NiftI)

Este formato surge ante la necesidad de disponer de un formato estándar de imagen en el campo de la neuroimagen. La principal ventaja frente a DICOM es la agrupación de una imagen 3D (o 4D) en un único archivo en lugar de tener los distintos cortes de la imagen en archivos independientes. Esto facilita, por tanto, la lectura y escritura de imágenes 3D. Al igual que DICOM, dispone de una cabecera con una serie de metadatos necesarios para la correcta representación de la imagen (orientación de la imagen con respecto al paciente, tamaño del píxel/vóxel, tamaño de la imagen, etc.), sin embargo, se pierden gran parte de los metadatos de los que se dispone en la cabecera DICOM, lo que se convierte en la principal desventaja de este formato. Sin embargo, para el almacenamiento de máscaras de segmentación, dado que no se necesitan metadatos adicionales, se convierte en un formato interesante de considerar ya que permite el almacenamiento directo de la máscara binaria generada (sin pérdidas).

Para el almacenamiento de la máscara de segmentación se diseñó el algoritmo con el objetivo de que el archivo NIFTI obtenido se superpusiera sobre la serie DICOM original. Para la construcción de este archivo se hizo uso de la librería *nibabel*¹¹. Para el almacenamiento de la máscara, se necesitan dos componentes principales: la matriz correspondiente a la propia imagen y la matriz afín en el sistema de coordenadas adecuado.

¹¹ <https://nipy.org/nibabel/>

La Figura 92 muestra la diferencia principal en el sistema de coordenadas empleado por cada uno de los formatos. DICOM emplea una orientación LPS (del inglés, *Left-Posterior-Superior*) mientras que NIfTI emplea una orientación RAS (del inglés, *Right-Anterior-Superior*). Por lo tanto, podemos ver, que la dirección positiva de los ejes x y entre ambos formatos es la opuesta.

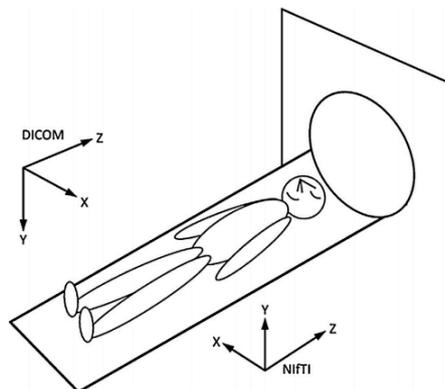


Figura 92. Sistema de coordenadas del paciente empleado por DICOM y NIfTI [130]. Se puede observar cómo los ejes x e y van en direcciones opuestas.

De este modo, para la construcción de la matriz afín empleamos la misma ecuación que para archivo DICOM, definida anteriormente, cambiando los signos correspondientes a los ejes x e y (Ecuación 65).

$$\text{Matriz afín NIfTI} = \begin{bmatrix} -r_x \Delta_x & -c_x \Delta_y & -\frac{x_n - x_1}{n - 1} & -x_1 \\ -r_y \Delta_x & -c_y \Delta_y & -\frac{y_n - y_1}{n - 1} & -y_1 \\ r_z \Delta_x & c_z \Delta_y & \frac{z_n - z_1}{n - 1} & z_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (65)$$

5.2.4. Servicio API REST

Para poder realizar la conversión de cualquier máscara de segmentación a formato DICOM RT Struct, DICOM Seg o NIfTI se implementó un servicio API (Interfaz de programación de aplicaciones, del inglés *Application Programming Interface*) REST (Transferencia de estado representacional, del inglés, *Representational state transfer*) en Python haciendo uso de la librería FastAPI¹². Esto permitía a la plataforma Quibim Precision la conversión de las máscaras generadas a un formato estándar.

¹² <https://fastapi.tiangolo.com/>

La API REST estaba formada principalmente por un *endpoint* (`/convert/url`) (Figura 93) que, dada una máscara en formato JSON (formato interno del visor), devolvía como *stream* de datos el archivo correspondiente.



Figura 93. Documentación correspondiente al endpoint implementado en el servicio API REST para la conversión de máscaras entre formatos estándar.

Para interactuar con dicho servicio se integró dentro del visor de Quibim Precision una vista para importar y/o exportar segmentaciones en los formatos descritos (Figura 94). A través del back-end de la plataforma se hacían llamadas a la API REST enviando, en el cuerpo de la petición, la información requerida (Figura 93):

- *input*: formato de origen (al importar es uno de los siguientes: “nifti”, “rtstruct”, o “dicomseg”, al exportar se emplea el formato “qrjson” ya que se refiere al formato interno del visor.).
- *output*: formato de salida (al importar se emplea el formato “qrjson” como formato de salida, mientras que al exportar se emplea uno de los siguientes: “nifti”, “rtstruct” o “dicomseg”).
- *rois*: URL al archivo que contiene la máscara. Al importar es la URL al archivo NIFTI, DICOM RT Struct o DICOM Seg subido y al exportar se trata de la URL al archivo JSON que genera el visor.
- *origin*: vector con la URL a los archivos DICOM de la serie. Para reducir número de peticiones y tiempo de ejecución, desde el visor se envía únicamente la URL al primer y último corte ya que son los necesarios para generar la matriz afin.
- *dimensions*: Tamaño de la imagen. Dado que sólo se mandan dos archivos DICOM no se puede conocer, a priori, el número de cortes total de la serie. Por esto, se incluye dentro de la petición.

Capítulo 5. Implementación en práctica clínica

- *extra*: Información adicional requerida por el servicio para hacer las transformaciones. Al exportar a DICOM RT Struct se necesita conocer el SOPInstanceUID de los cortes donde se tiene segmentación. Dado que sólo se tienen dos DICOM de la serie de referencia, esta información se envía como un vector dentro de este campo de la petición. Por otro lado, al importar y exportar archivos NIfTI se envía como información extra la extensión del archivo indicando si se encuentra comprimido (.nii.gz) o no (.nii).

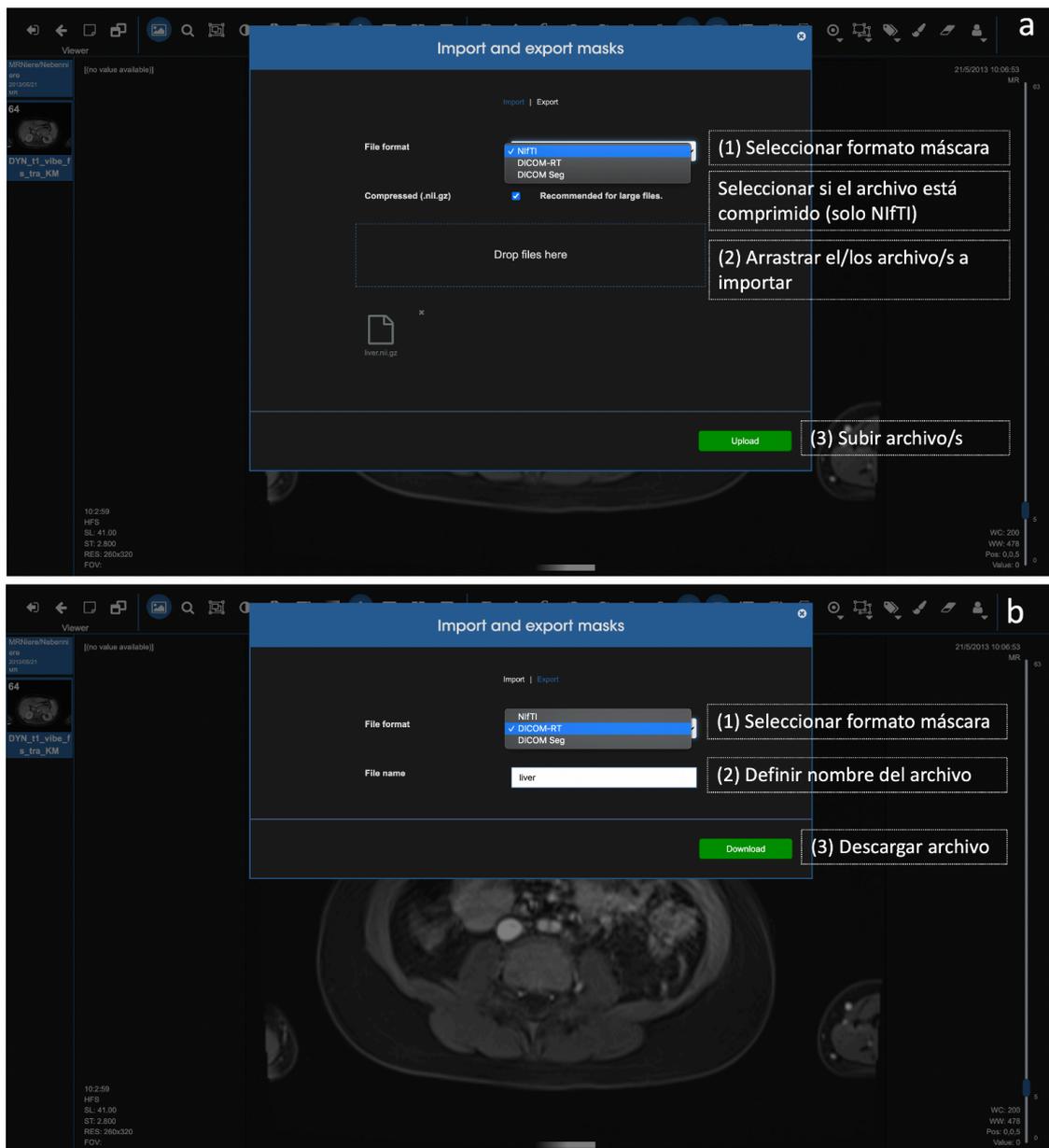


Figura 94. Modal para importar (a) y exportar (b) archivos de segmentación en formatos estándar al visor de Quibim Precision. (a) Para importar en primer lugar hay que definir el formato del archivo entre los definidos y, a continuación, arrastrar el/los archivo/s a importar. (b) Para exportar hay que seleccionar el formato en el que queremos el archivo y establecer un nombre de archivo con el que se descargará la máscara.

Capítulo 5. Implementación en práctica clínica

En la Figura 95 podemos observar los resultados obtenidos tras importar una máscara, originalmente en formato NIFTI en el visor de la plataforma y, posteriormente, exportar la misma máscara a formato DICOM RT Struct.

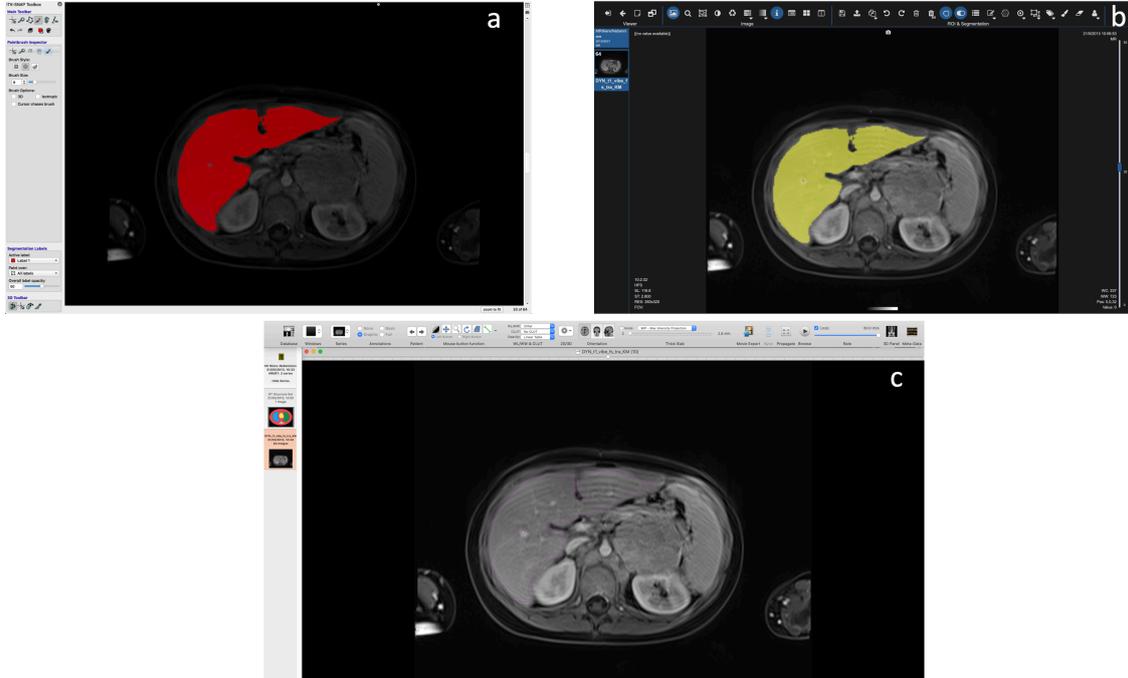


Figura 95. Importar y exportar máscara. (a) Máscara original en formato NifTI visualizada con ITKSnap que se importa al visor de Quibim Precision (b). (c) Máscara exportada en formato DICOM RT Struct visualizada en Horos¹³.

¹³ <https://horosproject.org/>

Capítulo 6. Conclusiones y trabajo futuro

Conclusiones

Las principales conclusiones extraídas del trabajo realizado se pueden resumir en los siguientes puntos.

- Las CNNs han supuesto un cambio de paradigma en la resolución de problemas de segmentación en imagen médica radiológica. Gracias a las CNN podemos implementar metodologías automáticas de segmentación de órganos y lesiones que ofrecen gran precisión. En la presente tesis doctoral se han implementado diferentes aproximaciones para la segmentación del parénquima hepático en imágenes de RM haciendo uso de dos secuencias diferentes (MECSE y DCE) y para la segmentación de la grasa visceral y subcutánea en estudios de TC. Los resultados obtenidos han sido superiores al estado del arte.
- El correcto ajuste de los hiperparámetros, como la tasa de aprendizaje o la función de pérdidas, es de vital importancia para conseguir la convergencia óptima del modelo entrenado. Con este objetivo se ha implementado una rutina de ajuste de hiperparámetros, que, siguiendo un proceso de validación cruzada, evalúa diferentes tasas de aprendizaje y funciones de pérdidas y selecciona aquella configuración que mejores resultados ofrezca en la resolución de cada problema. Esto tiene una gran posibilidad de traslación a otras aplicaciones de IA, con el diseño de procedimientos de entrenamiento que autoajusten los hiperparámetros.
- A la hora de desarrollar un modelo de IA, no solo es importante el entrenamiento y ajuste del propio modelo, sino que también adquiere gran relevancia el preparado previo de la imagen, así como el postprocesado óptimo del resultado obtenido. Todo esto requiere un conocimiento amplio del problema a resolver. De este modo, se han implementado rutinas de preprocesado de la imagen (normalización y redimensionado) y aumento de datos, así como de postprocesado del resultado obtenido (ajuste del umbral o filtrado de regiones segmentadas) específicas para cada problema que han permitido maximizar los resultados alcanzados.
- Los métodos de segmentación automática en imagen médica son de especial relevancia para la cuantificación y caracterización automática de los tejidos. Así, se han implementado y validado rutinas de extracción de biomarcadores de imagen para la cuantificación del parénquima hepático en enfermedades difusas, a través del PDFF y R2*, así como en enfermedades focales, a través de los

descriptores de la curva de captación de contraste, como el $IAUC_{60}$, pico de contraste o el tiempo al pico. Durante la validación de dicha cuantificación se ha visto una correlación perfecta entre los valores cuantificados mediante segmentaciones manuales y aquellos obtenidos con los modelos entrenados.

- Estos métodos son de gran interés en el ámbito clínico ya que permiten ayudar a los expertos a la hora de realizar el diagnóstico. Así, tras pasar todas las validaciones oportunas, se han integrado en una plataforma de análisis de imagen que permite poner al alcance de los radiólogos herramientas de cuantificación automática. Para lograr dicha integración se han realizado una serie de desarrollos complementarios como son: el desarrollo de una herramienta de anotación de máscaras de segmentación, el despliegue de los modelos desarrollados en forma de imágenes Docker o el desarrollo de un servicio que permite importar y exportar máscaras de segmentación en formatos estándar (DICOM RT Struct, DICOM Seg o NIfTI).

Trabajo futuro

Como trabajo futuro, para cada uno de los desarrollos presentados, se plantea lo siguiente:

- En cuanto a la segmentación hepática en secuencias MECSE para la cuantificación del hierro y la grasa, se plantea la división del parénquima hepático, en primer lugar, en lóbulo hepático izquierdo y derecho, y, en segundo lugar, en los diferentes sectores (Figura 96). De este modo, se puede realizar una cuantificación por sectores y conocer la distribución de grasa y hierro en cada uno de ellos. Además, se conoce que otros órganos puede ser de interés en el estudio de la enfermedad difusa hepática, como el páncreas, el bazo o la médula ósea, de este modo, se plantea enriquecer el módulo de caracterización mediante la segmentación de más regiones de interés.

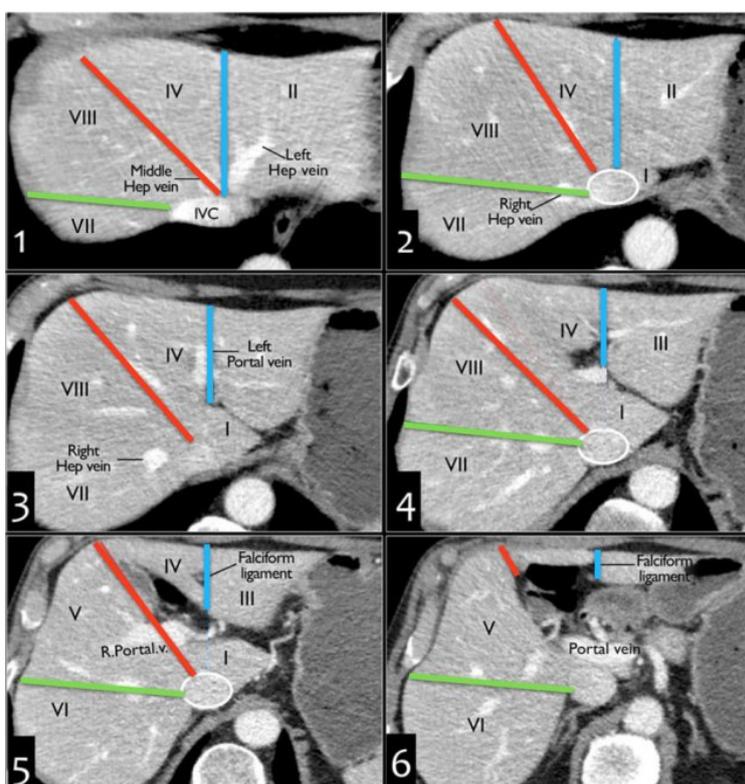


Figura 96. División del hígado en los VI sectores hepáticos. Fuente imagen: radiologyassistant.

- En relación con el modelo de segmentación automática de la grasa visceral y subcutánea, se propone ampliar dicho modelo con la segmentación del músculo y la grasa intramuscular, de este modo, se logrará un análisis completo de la composición corporal (Figura 97) para automatizar, en mayor medida, el estudio de enfermedades metabólicas.

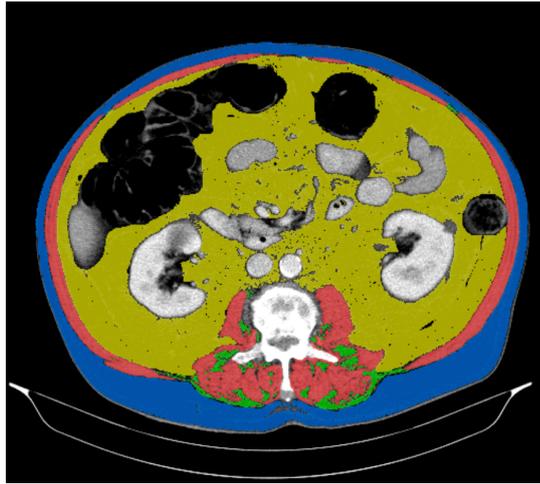


Figura 97. Imagen de TC con análisis de la composición corporal completo: grasa visceral (amarillo), grasa subcutánea (azul), músculo (rojo) y grasa intramuscular (verde).

- Finalmente, en cuanto a la segmentación hepática en secuencias de perfusión, se propone el enriquecimiento de la caracterización de las lesiones hepáticas focales mediante la segmentación automática de éstas. Para ello, se evaluará el valor añadido de emplear los propios mapas paramétricos obtenidos de la cuantificación de la secuencia de perfusión como entrada al modelo en forma de múltiples canales. Esto permitiría la caracterización del área tumoral y peritumoral (Figura 98). Además, la parcelación del hígado en los distintos sectores, en este caso también adquiere gran relevancia ya que permitirá la localización de las lesiones y un estudio volumétrico más preciso del hígado.

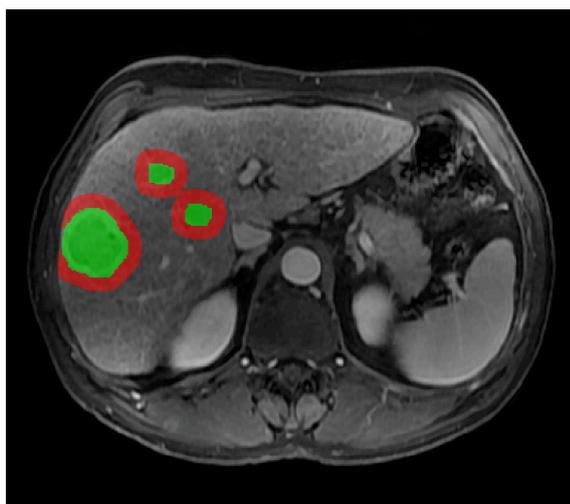


Figura 98. Segmentación de las lesiones focales presentes en el parénquima hepático (verde) del área peritumoral (rojo).

Referencias

Referencias

1. Onesi SO, Ignatius UE. Metabolic syndrome: Performance of five different diagnostic criterias. *Indian J Endocrinol Metab.* 2014 Jul;18(4):496-501.
2. Qiang G, Kong HW, Fang D, et al. The obesity-induced transcriptional regulator TRIP-Br2 mediates visceral fat endoplasmic reticulum stress-induced inflammation. *Nat Commun.* 2016 Apr 25;7:11378.
3. Ibrahim MM. Subcutaneous and visceral adipose tissue: structural and functional differences. *Obes Rev.* 2010 Jan;11(1):11–8.
4. Fox CS, Massaro JM, Hoffmann U, et al. Abdominal visceral and subcutaneous adipose tissue compartments: association with metabolic risk factors in the Framingham Heart Study. *Circulation.* 2007 Jul 3;116(1):39-48.
5. Brozek J, Grande F, Anderson JT, Keys A. Densitometric analysis of body composition: Revision of some quantitative assumptions. *Ann N Y Acad Sci.* 1963 Sep 26;110:113-40.
6. Kullberg J, Brandberg J, Angelhed JE, et al. Whole-body adipose tissue analysis: comparison of MRI, CT and dual energy X-ray absorptiometry. *Br J Radiol.* 2009 Feb;82(974):123–30.
7. Younossi ZM. Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis: Implications for liver transplantation. *Liver Transpl.* 2018 Feb;24 (2):166-70.
8. Pietrangelo A. Genetics, genetic testing, and management of hemochromatosis: 15 years since hepcidin. *Gastroenterology.* 2015 Oct; 149(5):1240–1251.e4.
9. Pietrangelo A. Iron in NASH, chronic liver diseases and HCC: How much iron is too much? *J Hepatol.* 2009 Feb;50(2):249–51.
10. Kew MC. Hepatic iron overload and hepatocellular carcinoma. *Liver Cancer.* 2014 Mar;3(1):31–40.
11. Bravo A, Sheth S, Chopra S, 2001. Liver Biopsy. *New England Journal of Medicine,* 344(7), pp.495-500.
12. Ratziu V, Charlotte F, Heurtier A, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology.* 2005 Jun;128(7):1898-906.
13. Butensky E, Fischer R, Hudes M, et al. Variability in hepatic iron concentration in percutaneous needle biopsy specimens from patients with transfusional hemosiderosis. *Am J Clin Pathol.* 2005 Jan;123(1):146-52.

14. Deugnier Y, Turlin B. Pathology of hepatic iron overload. *World J Gastroenterol*. 2007 Sep 21;13(35):4755-60
15. Villeneuve JP, Bilodeau M, Lepage R, et al. Variability in hepatic iron concentration measurement from needle-biopsy specimens. *J Hepatol*. 1996 Aug; 25(2):172– 177.
16. Bedossa P, Carrat F. Liver biopsy: The best, not the gold standard. 2009. *J Hepatol* 50(1):1–3.
17. Singal AG, Lampertico P, Nahon P. Epidemiology and surveillance for hepatocellular carcinoma: New trends. *J Hepatol*. 2020 Feb; 72(2):250-61.
18. Farhang Zangneh H, Wong WWL, Sander B, et al. Cost effectiveness of hepatocellular carcinoma surveillance after a sustained virologic response to therapy in patients with hepatitis C virus infection and advanced fibrosis. *Clin Gastroenterol Hepatol*. 2019 Aug; 17(9), pp.1840-49.e16.
19. Becker AK, Tso DK, Harris AC, Malfair D, Chang SD. Extrahepatic metastases of hepatocellular carcinoma: A spectrum of imaging findings. *Can Assoc Radiol J*. 2014 Feb; 65(1):60-6.
20. Vilana R, Forner A, García Á, Ayuso C, Bru C. Hepatocellular Carcinoma: Diagnosis, staging, and treatment strategy. *Radiología (English Edition)*. 2010 Sept-Oct; 52(5): 385-98.
21. Shen W, Punyanitya M, Wang Z, et al. Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. *J Appl Physiol*. 2004 Dec; 97(6):2333-8.
22. Tong Y, Udupa JK, Torigian DA. Optimization of abdominal fat quantification on CT imaging through use of standardized anatomic space: A novel approach. *Med Phys*. 2014 Jun; 41(6): 063501.
23. Bozzetti F. Forcing the vicious circle: sarcopenia increases toxicity, decreases response to chemotherapy and worsens with chemotherapy. *Ann Oncol*. 2017 Sep 1; 28(9):2107-18.
24. Prado CM, Baracos VE, McCargar LJ, et al. Sarcopenia as a determinant of chemotherapy toxicity and time to tumor progression in metastatic breast cancer patients receiving capecitabine treatment. *Clin Cancer Res*. 2009 Apr 15; 15(8): 2920-6.

25. Karatas F, Erdem GU, Sahin S, et al. Obesity is an independent prognostic factor of decreased pathological complete response to neoadjuvant chemotherapy in breast cancer patients. *Breast*. 2010 Apr; 32:237-44.
26. França M, Alberich-Bayarri Á, Martí-Bonmatí L, et al. Accurate simultaneous quantification of liver steatosis and iron overload in diffuse liver diseases with MRI. *Abdom Radiol (NY)*. 2017 May; 42(5):1434-43.
27. Westphalen AC, Qayyum, A, Yeh BM, et al. Liver fat: effect of hepatic iron deposition on evaluation with opposed-Phase MR imaging. *Radiology*. 2007 Feb; 242(2):450-55.
28. Liu CY, McKenzie CA, Yu H, Brittain JH, Reeder SB. Fat quantification with IDEAL gradient echo imaging: Correction of bias from T1 and noise. *Magn Reson Med*. 2007 Aug; 58(2):354-64.
29. Stejskal E, Tanner J. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *J Chem Phys*. 1965;42(1):288-92.
30. Dregely I, Prezzi D, Kelly-Morland C, Rocchia E, Neji R, Goh V. Imaging biomarkers in oncology: Basics and application to MRI. *J Magn Reson Imaging*. 2018 Jul;48(1):13-26.
31. deSouza NM, Achten E, Alberich-Bayarri A, et al; European Society of Radiology. Validated imaging biomarkers as decision-making tools in clinical trials and routine practice: current status and recommendations from the EIBALL* subcommittee of the European Society of Radiology (ESR). *Insights Imaging*. 2019 Aug 29;10(1):87.
32. Martí Bonmatí L, Alberich-Bayarri A, García-Martí G, et al. Imaging biomarkers, quantitative imaging, and bioengineering. *Radiología (English Edition)*. 2012 May-Jun; 54(3):269-78.
33. Schmeel FC. Variability in quantitative diffusion-weighted MR imaging (DWI) across different scanners and imaging sites: is there a potential consensus that can help reducing the limits of expected bias? *Eur Radiol*. 2009 May; 29(5):2243-45.
34. Wood JC. Use of magnetic resonance imaging to monitor iron overload. *Hematol Oncol Clin North Am*. 2014 Aug; 28(4):747-64, vii.
35. Hernando D, Kühn JP, Mensel B, et al. R2* estimation using “in-phase” echoes in the presence of fat: The effects of complex spectrum of fat. *J Mag Reson Imaging*. 2013 Mar; 37(3):717-26.

36. Meisamy S, Hines CD, Hamilton G, et al. Quantification of hepatic steatosis with T1-independent, T2*-corrected MR imaging with spectral modeling of fat: blinded comparison with MR spectroscopy. *Radiology*. 2011 Mar; 258(3):767-75.
37. Yu H, Shimakawa A, Hines C, et al. Combination of complex-based and magnitude-based multiecho water-fat separation for accurate quantification of fat-fraction. *Magn Reson in Med*. 2011 Jul; 66(1):199-206.
38. Kühn JP, Hernando D, Muñoz del Rio A, et al. Effect of multiplex spectral modeling of fat for liver iron and fat quantification: correlation of biopsy with MR imaging results. *Radiology*. 2012 Oct; 265(1):133-42.
39. Hamilton G, Yokoo T, Bydder M, et al. In vivo characterization of the liver fat 1H MR spectrum. *NMRBiomed*. 2011 Aug; 24(7):784-90.
40. França M, Alberich-Bayarri Á, Martí-Bonmatí L, et al. Accurate simultaneous quantification of liver steatosis and iron overload in diffuse liver diseases with MRI. *Abdom Radiol (NY)* 2017;42(5):1434–1443.
41. Wood JC, Enriquez C, Ghugre N, et al. MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. *Blood*. 2005 Aug 15; 106(4):1460–65.
42. Garbowski MW, Carpenter J-P, Smith G, et al. Biopsy-based calibration of T2* magnetic resonance for estimation of liver iron concentration and comparison with R2 Ferriscan. *J Cardiovasc Magn Reson*. 2014; 16(1):40.
43. Hankins JS, McCarville MB, Loeffler RB, et al. R2* magnetic resonance imaging of the liver in patients with iron overload. *Blood*. 2009. 113(20):4853–55.
44. Deugnier Y, Turlin B. Pathology of hepatic iron overload. *World J Gastroenterol*. 2007; 13(35):4755-60.
45. Reeder SB, Yokoo T, França M, et al. Quantification of Liver Iron Overload with MRI: Review and Guidelines from the ESGAR and SAR. *Radiology*. 2023 Apr;307(1):e221856.
46. Woodhams R, Kakita S, Hata H, et al. Diffusion-weighted imaging of mucinous carcinoma of the breast: evaluation of apparent diffusion coefficient and signal intensity in correlation with histologic findings. *AJR Am J Roentgenol*. 2009 Jul; 193(1):260-6.

47. Surov A, Meyer HJ, Wienke A. Correlation between apparent diffusion coefficient (ADC) and cellularity is different in several tumors: a meta-analysis. *Oncotarget*. 2017 May. 8(35):59492-9.
48. Oussous SA, Boujraf S, Kamaoui I. Assessment of the focal hepatic lesions using diffusion tensor magnetic resonance imaging. *J Med Signals Sens*. 2016 Apr-Jun; 6(2):99-105.
49. El-Nahas M, Sadek A, Gwely N, Elrakhawy M, Al-Naghy E. MR diffusion imaging and ADC measurement in solid and cystic mediastinal masses for differentiating benign from malignant lesions. *Egypt J Radiol Nucl Med*. 2018; 49(4):976-85.
50. Tschudi Y, Pollack A, Punnen S. Automatic detection of prostate tumor habitats using diffusion MRI. *Sci Rep*. 2018 Nov; 8(1):16801.
51. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology*. 1988 Aug; 168(2):497–505.
52. Szubert-Franczak AE, Naduk-Ostrowska M, Pasicz K, Podgórska J, Skrzyński W, Cieszanowski A. Intravoxel incoherent motion magnetic resonance imaging: basic principles and clinical applications. *Pol J Radiol*. 2020 Nov 18;85:e624-e635.
53. Steven A, Zhuo J, Melhem E, 2014. Diffusion Kurtosis Imaging: An Emerging Technique for Evaluating the Microstructural Environment of the Brain. *American Journal of Roentgenology*, 202(1):W26-W33.
54. Wu E, Cheung M, 2010. MR diffusion kurtosis imaging for neural tissue characterization. *NMR in Biomedicine*, 23(7):836-848.
55. Tabesh A, Jensen J, Ardekani B, Helpert J, 2010. Estimation of tensors and tensor-derived measures in diffusional kurtosis imaging. *Magnetic Resonance in Medicine*, 65(3):823-836.
56. Liu X, Zhou L, Peng W, Wang H, Zhang Y, 2015. Comparison of stretched-Exponential and monoexponential model diffusion-Weighted imaging in prostate cancer and normal tissues. *Journal of Magnetic Resonance Imaging*, 42(4):1078-1085.
57. Bergers, G.; Benjamin, L.E. Tumorigenesis and the angiogenic switch. *Nat. Rev. Cancer* 2003, 3, 401–410.

58. Yadav, L.; Puri, N.; Rastogi, V.; Satpute, P.; Sharma, V. Tumour Angiogenesis and Angiogenic Inhibitors: A Review. *J. Clin. Diagn. Res.* 2015, 9, XE01–XE05.
59. Sengupta A, Gupta RK, Singh A. Evaluation of B1 inhomogeneity effect on DCE-MRI data analysis of brain tumor patients at 3T. *J Transl Med.* 2017 Dec 2;15(1):242.
60. Sung K, Daniel BL, Hargreaves BA. Transmit B1+ field inhomogeneity and T1 estimation errors in breast DCE-MRI at 3 tesla. *J Magn Reson Imaging.* 2013;38(2):454-9.
61. de Bazelaire CM, Duhamel GD, Rofsky NM, Alsop DC. MR imaging relaxation times of abdominal and pelvic tissues measured in vivo at 3.0 T: preliminary results. *Radiology.* 2004 Mar;230(3):652-9.
62. Nam JG, Kang KM, Choi SH, et al. Comparison between the Prebolus T1 Measurement and the Fixed T1 Value in Dynamic Contrast-Enhanced MR Imaging for the Differentiation of True Progression from Pseudoprogession in Glioblastoma Treated with Concurrent Radiation Therapy and Temozolomide Chemotherapy. *AJNR Am J Neuroradiol.* 2017 Dec; 38(12):2243-50.
63. Conte GM, Altabella L, Castellano A. Comparison of T1 mapping and fixed T1 method for dynamic contrast-enhanced MRI perfusion in brain gliomas. *Eur Radiol.* 2019 Jul; 29(7):3467-79.
64. Tofts PS, Brix G, Buckley DL, et al. Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusable tracer: standardized quantities and symbols. *J Magn Reson Imaging.* 1999; 10:223232.
65. Tofts PS, Kermode AG. Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. *Magn Reson Med.* 1991; 17:357-367.
66. Curry H. The method of steepest descent for non-linear minimization problems. *Quarterly of Applied Mathematics.* 1944; 2(3):258-61.
67. Rumelhart D, Hinton G, Williams R. Learning representations by back-propagating errors. *Nature.* 1986; 323(6088):533-6.
68. Tang F, Liang S, Zhong T, Huang X, Deng X, Zhang Y, Zhou L. Postoperative glioma segmentation in CT image using deep feature fusion model guided by multi-sequence MRIs. *Eur Radiol.* 2020 Feb;30(2):823-832.

69. Aldojo N, Lukas S, Dewey M, Penzkofer T. Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network. *Eur Radiol.* 2020 Feb;30(2):1243-1253.
70. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing.* 2018, 321: 321-331.
71. Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis.* 2015. 115, 211-252
72. Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009: 248-255.
73. Krizhevsky A, Sutskever I; Hinton GE. ImageNet classification with deep convolutional neural networks. *Communications of the ACM.* 2017; 60(6): 84–90
74. Kaiming H, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016: 770-778.
75. Huang G, Liu Z, van der Maaten L, Weinberger K. *Densely Connected Convolutional Networks.* [online] arXiv.org. 2020.
76. Tang YX, Tang YB, Peng Y, et al. Automated abnormality classification of chest radiographs using deep convolutional neural networks. *NPJ Digit Med.* 2020 May; 3:70.
77. Zhang C, Zhao J, Niu J, Li D. New convolutional neural network model for screening and diagnosis of mammograms. *PLoS One.* 2020 Aug; 15(8):e0237674.
78. Wang S, Zhou M, Liu Z, et al. Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. *Med Image Anal.* 2017 Aug; 40:172-83.
79. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.* MICCAI 2015. *Lecture Notes in Computer Science*, vol 9351. Springer, Cham
80. Duong MT, Rudie JD, Wang J, et al. Convolutional neural network for automated FLAIR lesion segmentation on clinical brain MR imaging. *AJNR Am J Neuroradiol.* 2019 Aug; 40(8):1282-90.

81. Wang K, Mamidipalli A, Retson T, et al. Automated CT and MRI liver segmentation and biometry using a generalized convolutional neural network. *Radiol Artif Intell.* 2019 Mar; 1(2):180022.
82. Milletari F, Navab N, Ahmadi S. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. 2020. arXiv: 1606.04797.
83. Lee C, Xie S, Gallagher P, Zhang Z, Tu Z. Deeply-supervised nets. 2015. arXiv: 1409.5185.
84. Dou Q, Chen H, Jin Y, Yu L, Qin J, Heng P. 3D deeply supervised network for automatic liver segmentation from CT volumes. 2020. arXiv: 1607.00582.
85. Oktay O, Schlemper J, Folgoc L. Attention U-Net: learning where to look for the pancreas. 2020. arXiv: 1804.03999.
86. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. 2016. arXiv: 1511.07122.
87. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. 2017. arXiv: 1703.06870.
88. Girshick R. Fast R-CNN. 2015. arXiv: 1504.08083.
89. Martí-Aguado D, Alberich-Bayarri Á, Martín-Rodríguez J, et al. Differences in multi-echo chemical shift encoded MRI proton density fat fraction estimation based on multifrequency fat peaks selection in non-alcoholic fatty liver disease patients. *Clin Radiol.* 2020 Nov; 75(11):880.e5-880.e12.
90. Martí-Aguado D, Rodríguez-Ortega A, Mestre-Alagarda C, et al. Digital pathology: accurate technique for quantitative assessment of histological features in metabolic-associated fatty liver disease. *Aliment Pharmacol Ther.* 2021 Jan;53(1):160-171.
91. Martí-Aguado D, Jiménez-Pastor A, Alberich-Bayarri Á, et al. Automated Whole-Liver MRI Segmentation to Assess Steatosis and Iron Quantification in Chronic Liver Disease. *Radiology.* 2021. 16:211027.
92. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology.* 2005; 41:1313-21.
93. Deugnier Y, Turlin B. Pathology of hepatic iron overload. *World J Gastroenterol.* 2007; 13(35):4755-60.
94. Murphy, A., Baba, Y. Windowing (CT). Reference article, *Radiopaedia.org.* (accessed on 14 Apr 2022) <https://doi.org/10.53347/rID-52108>

95. Ioffe S, Szegedy C. Batch Normalization: Accelerating deep network training by reducing internal covariate shift. 2015. arXiv: 1502.03167.
96. Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*. 2011;12: 2121–21.
97. Zeiler M. ADADELTA: An Adaptive Learning Rate Method. 2012. arXiv: 1212.5701.
98. Tieleman T, Geoffrey H. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA Neural Netw. Mach. Learn*. 2012; 4:26–31.
99. Kingma DP, Ba JL. Adam: a Method for Stochastic Optimization. *International Conference on Learning Representations*. 2015. arXiv: 1412.6980.
100. Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab*. 1948; 5(4):1–34.
101. Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945. 26(3):297–302.
102. Salehi SSM, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. 2017. arXiv: 1706.05721.
103. Tversky A. Features of similarity. *Psychological Review*. 1977; 84(4):327-52.
104. Abraham N, Khan NM. A novel focal Tversky loss function with improved attention U-Net for lesion segmentation. 2018. arXiv: 1810.07842.
105. Taghanaki S, Zheng Y, Kevin Zhou S, et al. Combo loss: Handling input and output imbalance in multi-organ segmentation. *Computerized Medical Imaging and Graphics*. 2019; 75:24-33.
106. Rohrer M, Bauer H, Mintorovitch J, Requardt M, Weinmann HJ. Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. *Invest Radiol*. 2005;40(11):715-24.
107. Yeung M, Sala E, Schönlieb C, Rundo L. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*. 2022. 95:102026.

108. Jimenez-Pastor A, Alberich-Bayarri A, Lopez-Gonzalez R, et al. Precise whole liver automatic segmentation and quantification of PDFF and R2* on MR images. *Eur Radiol.* 2021; 10.1007/s00330-021-07838-5.
109. Christ PF, Ettliger F, Grün F et al (2017). Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. arXiv: 1702.05970.
110. Wang K, Mamidipalli A, Retson T et al (2019) Automated CT and MRI liver segmentation and biometry using a generalized convolutional neural network. *Radiol Artif Intell.* 1(2):180022.
111. Lavdas I, Glocker B, Kamnitsas K et al (2017) Fully automatic, multiorgan segmentation in normal whole body magnetic resonance imaging (MRI), using classification forests (CFs), convolutional neural networks (CNNs), and a multi-atlas (MA) approach. *Med Phys* 44(10):5210–5220.
112. Jansen MJA, Kuijf HJ, Nielke M et al (2019) Liver segmentation and metastases detection in MR images using convolutional neural networks. *J Med Imaging (Bellingham)* 6(4):044003
113. Stocker D, Bashir MR, Kannengiesser SAR, Reiner CS (2018) Accuracy of automated liver contouring, fat fraction, and R2* measurement on gradient multiecho magnetic resonance images. *J Comput Assist Tomogr* 42(5):697–706.
114. Gonzalez Ballester M, Zisserman AP, Brady M. 2002. Estimation of the partial volume effect in MRI. *Medical Image Analysis*, 6(4):389-405.
115. Yan W, Huang L, Xia L et al (2020) MRI manufacturer shift and adaptation: increasing the generalizability of deep learning segmentation for MR images acquired with different scanners. *Radiol ArtifIntell* 2(4):e190195.
116. Zhu JY, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, in *IEEE International Conference on Computer Vision (ICCV)*, 2017.
117. Dulai PS, Sirlin CB, Loomba R. MRI and MRE for non-invasive quantitative assessment of hepatic steatosis and fibrosis in NAFLD and NASH: clinical trials to clinical practice. *J Hepatol* 2016;65(5):1006–1016.
118. Tang A, Desai A, Hamilton G, et al. Accuracy of MR imaging–estimated proton density fat fraction for classification of dichotomized histologic steatosis grades in nonalcoholic fatty liver disease. *Radiology* 2015;274(2):416–425.

119. Kukuk GM, Hittatiya K, Sprinkart AM, et al. Comparison between modified Dixon MRI techniques, MR spectroscopic relaxometry, and different histologic quantification methods in the assessment of hepatic steatosis. *Eur Radiol* 2015;25(10):2869–2879.
120. Bashir MR, Wolfson T, Gamst AC, et al. Hepatic R2* is more strongly associated with proton density fat fraction than histologic liver iron scores in patients with nonalcoholic fatty liver disease. *J Magn Reson Imaging* 2019;49(5):1456–1466.
121. Muresan B, Sánchez Juan C, Artero A, Hernández Machancoses A, Almendros-Blanco P, Montoro A, Roselló Ferrando J, Íñigo Valdenebro R, Ríos Ríos J, Soriano del Castillo J, Quirante Cascales J, López-Torrecilla J. 2019. Measurement of body composition in cancer patients using CT planning scan at the third lumbar vertebra. *Nutrición Hospitalaria*.
122. Faron A, Sprinkart A, Kuetting D, Feisst A, Isaak A, Endler C, Chang J, Nowak S, Block W, Thomas D, Attenberger U, Luetkens J. 2020. Body composition analysis using CT and MRI: intra-individual intermodal comparison of muscle mass and myosteatorsis. *Scientific Reports*, 10(1).
123. Nowak S, Theis M, Wichtmann B, Faron A, Froelich M, Tollens F, Geißler H, Block W, Luetkens J, Attenberger U, Sprinkart A. 2021. End-to-end automated body composition analyses with integrated quality control for opportunistic assessment of sarcopenia in CT. *European Radiology*, 32(5), pp.3142-3151.
124. Weston, A., Korfiatis, P., Kline, T., Philbrick, K., Kostandy, P., Sakinis, T., Sugimoto, M., Takahashi, N. and Erickson, B., 2019. Automated Abdominal Segmentation of CT Scans for Body Composition Analysis Using Deep Learning. *Radiology*, 290(3), pp.669-679.
125. Ma D, Chow V, Popuri K, Beg M. 2022. Comprehensive Validation of Automated Whole Body Skeletal Muscle, Adipose Tissue, and Bone Segmentation from 3D CT images for Body Composition Analysis: Towards Extended Body Composition. *arXiv:2106.00652*.
126. Choe, J., Lee, S., Do, K., Lee, G., Lee, J., Lee, S. and Seo, J., 2019. Deep Learning-based Image Conversion of CT Reconstruction Kernels Improves Radiomics Reproducibility for Pulmonary Nodules or Masses. *Radiology*, 292(2), pp.365-373.

127. Selim M, Zhang J, Fei B, Zhang GQ, Ge GY, Chen J. Cross-Vendor CT Image Data Harmonization Using CVH-CT. *AMIA Annu Symp Proc.* 2022 Feb 21;2021:1099-1108. PMID: 35308983; PMCID: PMC8861670.
128. Jansen M, Kuijf H, Nielke M, Veldhuis W, Wessels F, Viergever M, Pluim J, 2019. Liver segmentation and metastases detection in MR images using convolutional neural networks. *Journal of Medical Imaging*, 6(04):1.
129. Gross M, Spektor M, Jaffe A, Kucukkaya A, Iseke S, Haider S, Strazzabosco M, Chapiro J, Onofrey J. 2021. Improved performance and consistency of deep learning 3D liver segmentation with heterogeneous cancer stages in magnetic resonance imaging. *PLOS ONE*, 16(12):e0260630.
130. Xuangrui Li et al. The first step for neuroimaging data analysis: DICOM to NIfTI conversion.

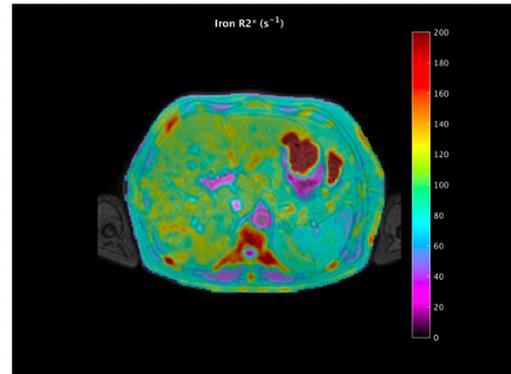
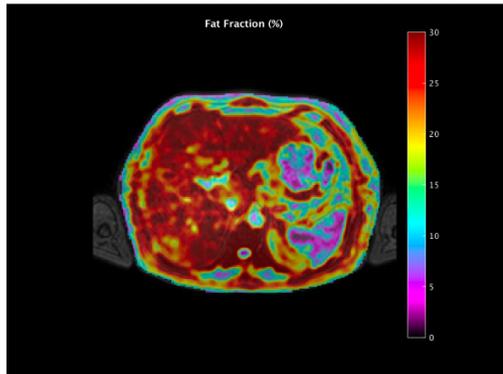
Anexo I.
Informe módulo
de análisis

Abdomen

Liver fat and iron concentration

VERSION 3.0.1

Imaging Center	123	Patient Name	Liver_Fat_Iron
Modality	MR	Patient ID	Liver_Fat_Iron
Study Description	Rm, Abdómen	Patient Sex	M
Study Date	20150818	Birthdate	undefined

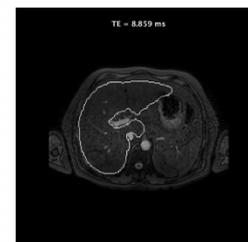
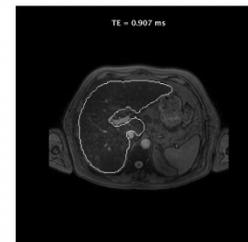
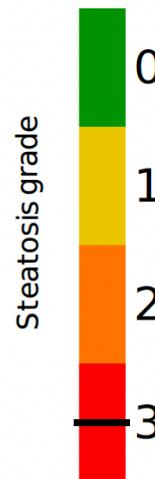


	PD Fat Fraction (%)
Mean	29.28
Std	8.30
p25	24.74
p75	34.38

PDFF Median: 29.49 %

	Iron ($\mu\text{mol/g}$)	$R2^*$ (s^{-1})
Mean	21.32	100.75
Std	5.03	33.53
p25	18.89	84.53
p75	23.16	113.01

Iron Median: 21.03 $\mu\text{mol/g}$



Anexo II. Cabecera
DICOM: Objetos RT
Struct y Seg

A continuación, se listan los campos de la cabecera principales que se han empleado para la construcción de los objetos DICOM RT Struct y DICOM Seg. Para seguir el estándar DICOM fue necesario incluir todos los campos obligatorios definidos por el mismo y, además, se incluyeron una serie de campos opcionales que se consideraron de interés. Para cada campo de la cabecera listado se define su clave, si es requerido u opcional y el tipo de dato (*Value Representation*)¹⁴.

DICOM RT Struct

Campos con información del paciente

Todos los campos definidos en esta sección toman los mismos valores que los que encontramos en los campos correspondientes en la cabecera de la serie de referencia.

- *Patient Name* (0010,0010 – Requerido – *Person Name*): Nombre del paciente. Cuando el estudio se encuentre anonimizado/pseudonimizado este parámetro se encontrará vacío o con un código identificador del paciente (de un ensayo clínico, por ejemplo).
- *Patient ID* (0010,0020 – Requerido – *Long String*): Identificador único del paciente. Cuando el estudio se encuentre anonimizado/pseudonimizado este parámetro se encontrará vacío o con un código identificador del paciente (de un ensayo clínico, por ejemplo).
- *Patient's Birth Date* (0010,0030 – Requerido – *Date*): Fecha de Nacimiento del paciente. Cuando el estudio se encuentre anonimizado/pseudonimizado este parámetro se encontrará vacío o con una fecha dada por defecto.
- *Patient's Sex* (0010,0040 – Requerido – *Code String*): Toma uno de los siguientes valores:
 - M → Sexo masculino.
 - F → Sexo femenino.
 - O → No disponible.

Campos con información del estudio

¹⁴ http://dicom.nema.org/dicom/2013/output/chtml/part05/sect_6.2.html

Anexo II. Cabecera DICOM: Objetos RT Struct y Seg

Todos los campos definidos en esta sección toman los mismos valores que los que encontramos en los campos correspondientes en la cabecera de la serie de referencia.

- *Study Date* (0008,0020 – Requerido – *Date*): Fecha en la que se realizó la prueba de imagen.
- *Study Time* (0008,0030 – Requerido – *Time*): Instante en el que se realizó la prueba de imagen.
- *Accession Number* (0008,0050 – Requerido – *Short String*): Número creado por el RIS que identifica la orden para la realización del estudio.
- *Referring Physician's Name* (0008,0090 – Requerido – *Person Name*): Nombre del médico que lleva al paciente.
- *Study Description* (0008,1030 – Opcional – *Long String*): Descripción del estudio.
- *Study Instance UID* (0020,000D – Requerido – *Unique Identifier*): Identificador único del estudio al que pertenece la serie sobre la que se ha realizado la segmentación.
- *Study ID* (0020,0010 – Requerido – *Short String*): Identificador del estudio definido por el usuario o la máquina.

Campos con información de la serie

- *Modality* (0008,0060 – Requerido – *Code String*): Modalidad de la serie. Para el almacenamiento de este tipo de archivos toma siempre el valor 'RTSTRUCT'
- *Series Description* (0008,103E – Opcional – *Long String*): Descripción de la serie. Se ha decidido darle el valor constante: 'RT Structure Set'.
- *Series Instance UID* (0020,000E – Requerido – *Unique Identifier*): Identificador único de la serie generada.
- *Series Number* (0020,0011 – Requerido – *Integer String*): Número único de la serie generada. Por defecto, se ha decidido darle un valor de '40000'.

Campos con información del equipo empleado

- *Manufacturer* (0008,0070 – Requerido – *Long String*): Fabricante de la máquina empleada para la adquisición de las imágenes. Va a tomar el mismo valor que el presente en la cabecera de la serie de referencia.

Campos con información del archivo dado

- *SOP Class UID* (0008,0016 – Requerido – *Unique Identifier*): Identificador único del tipo de modalidad empleada. Para objetos RT Struct toma el valor por defecto: ‘1.2.840.10008.5.1.4.1.1.481.3’.
- *SOP Instance UID* (0008,0018 – Requerido – *Unique Identifier*): Identificador única del archivo generado.

Campos con información del conjunto de ROIs generado

- *Structure Set Label* (3006,0002 – Requerido – *Short String*): Etiqueta definida por el usuario para el conjunto de ROIs (e.g, Grasa abdominal).
- *Structure Set Date* (3006,0008 – Requerido – *Date*): Fecha en la que se modificó por última vez el conjunto de ROIs.
- *Structure Set Time* (3006,0009 – Requerido – *Time*): Momento en el que se modificó por última vez el conjunto de ROIs.
- *Referenced Frame Of Reference Sequence* (3006,0010 – Opcional - *Sequence*): Frame o frames de referencia a los que la ROI se encuentra asociado. Dado que, en nuestro caso, la ROI está asociada siempre a una única serie, vamos a definir una única entrada con la siguiente información:
 - *Frame Of Reference UID* (0020,0052 – Requerido – *Unique Identifier*): Mismo valor que el que se encuentra en la cabecera de la serie de referencia.
 - *RT Referenced Study Sequence* (3006,0012 – Opcional - *Sequence*): Estudio o estudios de referencia a los que la ROI se encuentra asociada. Dado que, en nuestro caso, la ROI está asociada siempre a un único estudio, vamos a definir una única entrada con la siguiente información:

Anexo II. Cabecera DICOM: Objetos RT Struct y Seg

- *Referenced SOP Class UID* (0008,1150 – Requerido – *Unique Identifier*): Toma el valor 1.2.840.10008.3.1.2.3.2.
- *Referenced SOP Instance UID* (0008,1155 – Requerido – *Unique Identifier*): Valor del identificador único del estudio de referencia definido por el campo *Study Instance UID* (0020,000D).
- *RT Referenced Series Sequence* (3006,0014 – Requerido – *Sequence*): Serie o series de referencia a los que la ROI se encuentra asociada. Dado que, en nuestro caso, la ROI está asociada a una única serie, se define una única entrada con la siguiente información:
 - *Series Instance UID* (0020,000E – Requerido – *Unique Identifier*): Mismo valor que el que se encuentra en la cabecera de la serie de referencia.
 - *Contour Image Sequence* (3006,0016 – Requerido – *Sequence*): Conjunto de imágenes que forman la serie de referencia sobre las que se delinea la ROI. Tiene tantas entradas como cortes tiene la serie y, cada una, se define por la siguiente información:
 - *Referenced SOP Class UID* (0008,1150 – Requerido – *Unique Identifier*): Mismo valor que la cabecera *SOP Class UID* (0008,0016) de la serie de referencia.
 - *Referenced SOP Instance UID* (0008,1155 – Requerido – *Unique Identifier*): Valor del *SOP Instance UID* (0008,0018) del corte de referencia.
- *Structure Set ROI Sequence* (3006,0020 – Requerido – *Sequence*): Información general de cada una de las ROIs delineadas. Contiene la siguiente información:
 - *ROI Number* (3006,0022 – Requerido – *Integer String*): Número único de la ROI dada.
 - *Referenced Frame Of Reference UID* (3006,0024 – Requerido – *Unique Identifier*): Valor de la cabecera *Frame Of Reference UID* (0020,0052) de la serie de referencia.
 - *ROI Name* (3006,0026 – Requerido – *Long String*): Nombre/etiqueta de la ROI (e.g, Grasa visceral, Grasa subcutánea).

- *ROI Generation Algorithm* (3006,0036 – Requerido – *Code String*): Puede tomar los siguientes valores:
 - AUTOMATIC → ROI obtenida de manera automática (e.g., algoritmo de IA).
 - SEMIAUTOMATIC → ROI obtenida de manera semi-automática (e.g., crecimiento de regiones donde las semillas se definen de forma manual).
 - MANUAL → ROI delineada de manera completamente manual.

Campos con información del contorno de cada una de las ROIs.

- *ROI Contour Sequence* (3006,0039 – Requerido – *Sequence*): Información de los contornos que forman cada una de las ROIs. Para cada ROI se define la siguiente información:
 - *ROI Display Color* (3006,002A – Opcional – *Integer String*): Color con el que mostrar la ROI cuando se abre en un visor. Está formado por un triplete de valores entre 0 y 255 definiendo el color en codificación RGB. Para cada ROI se define un color predefinido en función del valor que tenga en la matriz de origen.
 - *Referenced ROI Number* (3006,0084 – Requerido – *Integer String*): Número de la ROI que se ha definido en la sección *Structure Set ROI Sequence* (0036,0020).
 - *Contour Sequence* (3006,0040 – Requerido – *Sequence*): Contorno o contornos que definen la ROI. Para cada uno de los contornos se define la siguiente información:
 - *Contour Geometry Type* (3006,0042 – Requerido – *Code String*): Tipo de geometría. Puede tomar los siguientes valores:
 - POINT → Punto único
 - OPEN_PLANAR → Contorno abierto que contiene puntos coplanares (todos se encuentran en el mismo corte).
 - OPEN_NONPLANAR → Contorno abierto que contiene puntos que no son coplanares (puntos en diferentes cortes).

Anexo II. Cabecera DICOM: Objetos RT Struct y Seg

- **CLOSED_PLANAR** → Contorno cerrado (polígono) que contiene puntos coplanares.

En nuestro caso, se va a emplear siempre el valor 'CLOSED_PLANAR'.

- *Number of Contour Points* (3006,0046 – Requerido – *Integer String*): Número de puntos que definen el contorno definido en *Contour Data* (3006,0050).
- *Contour Data* (3006,0050 – Requerido – *Decimal String*): Secuencia de tripletes (x, y, z) que definen cada uno de los puntos del contorno (en mm) en coordenadas físicas (sistema de coordenadas basado en el paciente).
- *Contour Image Sequence* (3006,0016 – Opcional – *Sequence*): Conjunto de imágenes que contienen el contorno dado. Contiene la siguiente información:
 - *Referenced SOP Class UID* (0008,1150 – Requerido – *Unique Identifier*): *SOP Class UID* (0008,0016) de la serie de referencia.
 - *Referenced SOP Instance UID* (0008,1150 – Requerido – *Unique Identifier*): *SOP Instance UID* (0008,0018) de la imagen que define el corte donde se ha delineado el contorno dado.

Campos con información de la identificación e interpretación de las ROIs.

- *RT ROI Observations Sequence* (3006,0080 – Requerido – *Sequence*): Conjunto de observaciones de cada una de las ROIs definidas. Contiene la siguiente información:
 - *Observation Number* (3006,0082 – Requerido – *Integer String*): Número único de la observación.
 - *Referenced ROI Number* (3006,0084 – Requerido – *Integer String*): Número único de la ROI a la que se refiere dicha observación definido en el campo *ROI Number* (3006,0022) dentro de *Structure Set ROI Sequence* (3006,0020).

Anexo II. Cabecera DICOM: Objetos RT Struct y Seg

- *RT ROI Interpreted Type* (3006,00A4 – Requerido – *Code String*): Tipo de la ROI. Puede tomar un valor dado dentro de una lista definida en el estándar o permanecer vacío. En nuestro caso se decidió dejarlo vacío por el desconocimiento, a priori, del tipo de ROI.
- *ROI Interpreter* (3006,00A6 – Requerido – *Person Name*): Nombre de la persona que está realizando la interpretación. En nuestro caso, se decidió establecer una cadena de texto vacía para este campo.

DICOM Seg

Los campos correspondientes a la información del paciente, del estudio, de la serie y del equipo empleado son compartidos entre cualquier archivo DICOM, por lo que los mismos que se han listado previamente para el objeto DICOM RT Struct se incluyen en los archivos DICOM Seg. Para los campos de paciente, estudio y equipo, los valores que se emplean son, también, los que obtenemos directamente de los campos correspondientes en la serie de referencia. Sin embargo, determinados campos con la información de la serie toman valores ligeramente diferentes:

- *Modality* (0008,0060 – Requerido – *Code String*): Modalidad de la serie. Para el almacenamiento de este tipo de archivos toma siempre el valor ‘SEG’.
- *Series Description* (0008,103E – Opcional – *Long String*): Descripción de la serie. Se ha decidido darle el valor constante: ‘Segmentation Object’.
- *Series Instance UID* (0020,000E – Requerido – *Unique Identifier*): Identificador único de la serie generada.
- *Series Number* (0020,0011 – Requerido – *Integer String*): Número único de la serie generada. Por defecto, se ha decidido darle un valor de ‘50000’.

Además, se encuentran una serie de campos específicos de este formato.

Campos con información de la imagen

- *Image Type* (0008,0008 – Requerido – *Code String*): Tipo de imagen. Está formado por dos o más componentes. En este caso se van a emplear 2 componentes: [‘DERIVED’, ‘PRIMARY’]. Indicando que es una imagen derivada

cuyos píxeles se han obtenido de alguna manera de los píxeles originales, y primaria, es decir, una imagen creada como resultado directo de la adquisición del paciente.

- *Instance Number* (0020,0013 – Requerido – *Integer String*): Número que identifica la imagen. Como este formato está formado por una única imagen tomará el valor ‘1’.
- *Patient Orientation* (0020,0020 – Requerido condicionalmente – *Code String*): Orientación de las filas y columnas de la imagen. Está formado por dos componentes que pueden tomar los valores:
 - A → anterior
 - P → posterior
 - R → derecha
 - L → izquierda
 - H → cabeza
 - F → pies

Este campo toma el mismo valor que el de la imagen de origen.

Campos con información de los píxeles de la imagen

- *Samples per Pixel* (0028,0002 – Requerido – *Unsigned Short*): Número de canales de la imagen almacenada. Por ejemplo, en imágenes en escala de grises tomaría un valor de 1 y en imágenes en RGB de 3. En este caso, para el almacenamiento de la máscara, se emplea un único canal, por lo que toma un valor de 1.
- *Photometric Interpretation* (0028,0004 – Requerido – *Code String*): Tipo de representación de los píxeles de la imagen. Se empleará MONOCHROME2.
- *Rows* (0028,0010 – Requerido – *Unsigned Short*): Número de filas de la imagen.
- *Columns* (0028,0011 – Requerido – *Unsigned Short*): Número de columnas de la imagen.
- *Bits Allocated* (0028,0100 – Requerido – *Unsigned Short*): Número de bits asignados a cada píxel. Se empleará 1 bit generalmente.
- *Bits Stored* (0028,0101 – Requerido – *Unsigned Short*): Número de bits almacenados por cada píxel. Se empleará un valor de 1.

Anexo II. Cabecera DICOM: Objetos RT Struct y Seg

- *High Bit* (0028,0102 – Requerido – *Unsigned Short*): Bit más significativo de un pixel. Toma un valor una unidad inferior al valor encontrado en el campo *Bits Stored*, por lo tanto, tomará un valor de 0.
- *Pixel Representation* (0028,0103 – Requerido – *Unsigned Short*): Representación de datos de los píxeles. Se empleará un valor de 0.
- *Pixel Data* (7FE0,0010 – Requerido condicionalmente – *Other Byte String* ó *Other Word String*): *Stream* de datos que forman los píxeles de la imagen.

Campos con información de la máscara de segmentación

- *Segmentation Type* (0062,0001 – Requerido – *Code String*): Tipo de codificación empleada para el almacenamiento de la máscara de segmentación. Existen dos opciones:
 - BINARY → Toma un valor 1 o 0 en función de si el píxel/vóxel se ha segmentado o no, respectivamente.
 - FRACTIONAL → Toma un valor entre 0 y el valor indicado en el campo *Maximum Fractional Value* (0062,000E). En caso de aplicar este tipo de codificación, será necesario especificar, a su vez, el campo *Segmentation Fractional Type* (0062,0010).Dado que las máscaras segmentadas definen si un píxel completo está segmentado o no, se va a emplear, por defecto, la codificación binaria ('BINARY').
- *Maximum Fractional Value* (0062,000E – Requerido condicional – *Unsigned Short*): Requerido únicamente si el campo *Segmentation Type* (0062,0001) contiene el código FRACTIONAL. Indica el porcentaje que supone una ocupación completa del. En *Pixel Data* (7FE0,0010) no podrá haber valores superiores a este.
- *Fractional Type* (0062,0010 – Requerido condicional – *Code String*): Indica el tipo de fraccionalidad aplicada al almacenamiento de la máscara. Puede tomar los siguientes valores:
 - PROBABILITY → Define la probabilidad, como la relación entre el valor del píxel y el valor almacenado en el campo *Maximum Fractional Value* (0062,000E), de que la región segmentada ocupe el área completa del píxel.

Anexo II. Cabecera DICOM: Objetos RT Struct y Seg

- OCCUPANCY → Define la proporción del volumen del píxel ocupado por la propiedad segmentada como la relación entre el valor del píxel y el valor del campo *Maximum Fractional Value* (0062,000E)
- *Content Label* (0070,0080 – Requerido – *Code String*): Etiqueta empleada para identificar la región segmentada.
- *Content Description* (0070,0081 – Requerido – *Long String*): Descripción de la región segmentada.
- *Content Creator's Name* (0070,0084 – Requerido – *Person Name*): Nombre de la persona que ha creado la segmentación. Para evitar problemas de protección de datos, este campo se deja vacío.
- *Segment Sequence* (0062,0002 – Requerido - *Sequence*): Secuencia del conjunto de regiones segmentadas. Para cada elemento de la secuencia se incluyen los siguientes campos:
 - *Segmented Property Category Code Sequence* (0062,0003 – Requerido – *Sequence*): Secuencia que define lo que contiene el archivo de segmentación. Esta secuencia está formada por un único elemento que contiene los siguientes campos:
 - *Code Value* (0008,0100 – Requerido condicional – *Short String*): Identificador de la ontología que representa la región segmentada en su totalidad (considerando todas las clases). En nuestro caso vamos a generar tres archivos diferentes:
 - Segmentación hepática en MECSE
 - Segmentación hepática en perfusión
 - Segmentación de grasa abdominalDada la similitud entre las dos primeras, ambas se van a asociar a la misma ontología (hígado) mientras que la tercera se va a asociar a una diferente (grasa) siendo los identificadores
 - Hígado: 1020004 (de la colección SNOMED CT)
 - Grasa: RID1538 (de la colección RadLex)
 - *Coding Scheme Designator* (0008,0102 – Requerido condicional – *Short String*): Colección empleada para definir la ontología incluida en el campo anterior. Como hemos visto se han empleado dos colecciones diferentes:

Anexo II. Cabecera DICOM: Objetos RT Struct y Seg

- SNOMED CT: A la que se le asigna el valor ‘SCT’
- RadLex: A la que se le asigna el valor ‘RADLEX’
- *Code Meaning* (0008,0104 – Requerido – *Long String*): Significado de la ontología empleada en los dos campos anteriores. En este caso, se le asignará el nombre de la región delineada (Hígado y Grasa según el caso).
- *Segment Number* (0062,0004 – Requerido – *Unsigned Short*): Identificador del número la región segmentada, debe de tener un valor único empezando con un valor de 1 para la primera región segmentada e incrementando de 1 en 1 en las siguientes regiones.
- *Segment Label* (0062,0005 – Requerido – *Long String*): Etiqueta de la región segmentada. Debe tomar el mismo valor que el campo *Code Meaning* (0008,0104) de *Segmented Property Type Code Sequence* (0062,000F).
- *Segment Algorithm Type* (0062,0008 – Requerido – *Code String*): Tipo de algoritmo empleado para generar la segmentación. Puede tomar uno de los siguientes valores:
 - AUTOMATIC → Segmentación automática
 - SEMIAUTOMATIC → Segmentación obtenida con intervención del usuario.
 - MANUAL → Segmentación realizada completamente por el usuario de forma manual.

De este modo, si se almacena la máscara según se obtiene de los modelos entrenados tomará el valor ‘AUTOMATIC’, mientras que si previamente sufre algún tipo de corrección tomará el valor ‘SEMIAUTOMATIC’.

- *Segment Algorithm Name* (0062,0009 – Requerido condicional – *Long String*): Requerido si el campo *Segment Algorithm Type* (0062,0008) toma un valor distinto a MANUAL, como es nuestro caso. Este campo contendrá el nombre del algoritmo empleado.
- *Segmented Property Type Code Sequence* (0062,000F – Requerido - *Sequence*): Secuencia que define las propiedades de la región segmentada. Esta secuencia está formada por un único elemento con los siguientes campos:

Anexo II. Cabecera DICOM: Objetos RT Struct y Seg

- *Code Value* (0008,0100 – Requerido condicional – *Short String*): Identificador de la ontología que identifica la región delineada. Para cada una de las regiones se ha seleccionado una ontología:
 - Hígado: 1020004 (de la colección SNOMED CT)
 - Grasa visceral: RID50365 (de la colección RadLex)
 - Grasa subcutánea: RID50366 (de la colección RadLex)
- *Coding Scheme Designator* (0008,0102 – Requerido condicional – *Short String*): Colección empleada para definir la ontología incluida en el campo anterior. Como hemos visto se han empleado dos colecciones diferentes:
 - SNOMED CT: A la que se le asigna el valor ‘SCT’.
 - RadLex: A la que se le asigna el valor ‘RADLEX’.
- *Code Meaning* (0008,0104 – Requerido – *Long String*): Significado de la ontología empleada en los dos campos anteriores. En este caso, se le asignará el nombre de la región delineada (Hígado, Grasa visceral o Grasa subcutánea según el caso).

Campos con información de series multi-frame (presente aunque solo haya un frame).

- *Content Date* (0008,0023 – Requerido – *Date*): Fecha en la que se creó la máscara.
- *Content Time* (0008,0033 – Requerido – *Time*): Momento en el que se creó la máscara.
- *Number of Frames* (0028,0008 – Requerido – *Integer String*): Número de frames. Dado que la máscara se almacena en un único DICOM, cada uno de los cortes de esta se almacenan como frames. Por lo tanto, el número de frames coincidirá con el número de cortes de la serie original.
- *Shared Functional Groups Sequence* (5200,9220 – Requerido – *Sequence*): Se asume que todas las máscaras almacenadas tienen las mismas dimensiones, resolución y orientación que la imagen original sobre las que se han delineado.
 - *Derivation Image Sequence* (0008,9124 – Requerido – *Sequence*): Secuencia que define sobre qué imágenes se ha realizado la segmentación. Tiene tantos elementos como regiones anotadas y, para cada una, imágenes segmentadas se tengan. Para cada secuencia se encuentran los siguientes campos:

Anexo II. Cabecera DICOM: Objetos RT Struct y Seg

- *Source Image Sequence* (0008,2112 – Requerido – *Sequence*): Secuencia que define la imagen sobre la que se ha realizado la anotación correspondiente. Está formado por un único elemento con la siguiente información:
 - *Referenced SOP Class UID* (0008,1150 – Requerido – *Unique Identifier*): SOP Class de referencia, se emplea el valor de la serie de referencia.
 - *Reference SOP Instance UID* (0008,1155 – Requerido – *Unique Identifier*): SOP Instance UID de la imagen correspondiente en la serie de referencia.
- *Derivation Code Sequence* (0008,9215 – Requerido condicional – *Sequence*): descripción del modo de creación del frame. Está formada por una secuencia de un único elemento con la siguiente información:
 - *Code Value* (0008,0100 – Requerido condicional – *Short String*): Adquiere el valor ‘113076’.
 - *Code Scheme Designator* (0008,0102 – Requerido condicional – *Short String*): Adquiere el valor ‘DCM’.
 - *Code Meaning* (0008,0104 – Requerido – *Long String*): Adquiere el valor ‘Segmentation’.
- *Frame Content Sequence* (0020,9111 – Requerido condicional – *Sequence*): Secuencia que contiene información general de la máscara. Tiene tantos elementos como regiones anotadas y, para cada una, imágenes segmentadas se tengan. Para cada secuencia se encuentra el campo *Dimension Index Values* (0020,9157 – Requerido condicional – *Unsigned Long*), el cual está formado por un vector de dos componentes, la primera indica el índice de la máscara (empezando en 1) y la segunda el índice del corte segmentado dentro de la máscara (empezando en 1).
- *Plane Position Sequence* (0020,9113 – Requerido – *Sequence*): Secuencia que identifica la posición del plano en el frame. Tiene tantos elementos como regiones anotadas y, para cada una, imágenes segmentadas se tengan. Para cada secuencia se encuentra el campo *Image Position Patient*

- (0020,0032 – Requerido condicional – *Decimal String*) que contiene el valor del mismo campo presente en la imagen asociada correspondiente.
- *Plane Orientation Sequence* (0020,9116 – Requerido – *Sequence*): Secuencia que define la orientación de las máscaras. Tendrá un único elemento con el campo *Image Orientation Patient* (0020,0037 – Requerido condicional – *Decimal String*), que define la orientación del paciente y tomará un valor igual al valor de dicho campo en la serie original.
 - *Pixel Measures Sequence* (0028,9110 – Requerido – *Sequence*): Secuencia que define las características físicas de los vóxeles de las máscaras. Tendrá un único elemento con los siguientes campos, todos ellos con valores iguales a los correspondientes en la serie de referencia:
 - *Slice Thickness* (0018,0050 – Requerido condicional – *Decimal String*), que define el espesor del corte.
 - *Spacing Between Slices* (0018,0088 – Requerido condicional – *Decimal String*), que define el espaciado entre cortes.
 - *Pixel Spacing* (0028,0030 – Requerido condicional – *Decimal String*), que define el tamaño del píxel en el plano de adquisición.
 - *Segment Identification Sequence* (0062,000A – Requerido – *Sequence*): Secuencia que identifica la región segmentada correspondiente. Tendrá un único elemento con el campo *Referenced Segment Number* (0062,000B – Requerido – *Unsigned Short*), que contiene el índice de la máscara correspondiente (empezando en 1).
 - *Dimension Organization Sequence* (0020,9221 – Requerido – *Sequence*): Secuencia que contiene el identificador de la organización. Contiene un único elemento con el campo *Dimension Organization UID* (0020,9164 – Requerido – *Unique Identifier*). Para crear este se ha implementado un método de creador de UID acorde al estándar DICOM.
 - *Dimension Index Sequence* (0020,9222 – Requerido condicional – *Sequence*): Secuencia que contiene los índices empleados para especifica la dimensión del objeto multi-frame. Contiene un único elemento con los siguientes campos:
 - *Dimension Organization UID* (0020,9162 – Requerido – *Unique Identifier*): Contiene el valor especificado en *Dimension Organization*

Anexo II. Cabecera DICOM: Objetos RT Struct y Seg

UID del campo *Dimension Organization Sequence*, definido en el punto anterior.

- *Dimension Index Pointer* (0020,9165 – Requerido – *Attribute Tag*): Contiene el identificador del campo DICOM que se emplea para identificar al atributo conectado con el índice de la máscara, es decir, el campo (0062, 000B).
- *Functional Group Pointer* (0020,9167 – Requerido condicional – *Attribute Tag*): Contiene el identificador del campo DICOM de la *Functional Group Sequence* que contiene el atributo referenciado por el *Dimension Index Pointer*. Por lo tanto, con tiene el campo (00062, 000A).
- *Dimension Organization Type* (0020,9311 – Opcional – *Code String*): Dimensión de la instancia. Se emplea el valor '3D'.

Anexo III. Publicaciones

En este anexo se presentan las publicaciones que han surgido durante la realización de la presente tesis doctoral. Estas se han dividido en capítulos de libro, comunicaciones orales en congresos nacionales e internacionales, presentación de posters y, finalmente, publicaciones científicas. En cuanto a estas últimas se han diferenciado entre aquellas relacionadas con la temática de la tesis doctoral y otras adicionales que han surgido de la colaboración en diferentes proyectos de investigación.

Capítulos de libro

- Alberich-Bayarri A, Jiménez-Pastor A, López-González R, García-Castro F. (2019) How to Develop Artificial Intelligence Applications. In: Ranschaert E., Morozov S., Algra P. (eds) Artificial Intelligence in Medical Imaging. Springer, Cham. https://doi.org/10.1007/978-3-319-94878-2_5.
- Mayorga-Ruiz I, Jiménez-Pastor A, Fos-Guarinos B, López-González R, García-Castro F, Alberich-Bayarri A. (2019) The Role of AI in Clinical Trials. In: Ranschaert E., Morozov S., Algra P. (eds) Artificial Intelligence in Medical Imaging. Springer, Cham. https://doi.org/10.1007/978-3-319-94878-2_16.
- Alberich-Bayarri, A., Jiménez-Pastor, A., Mayorga-Ruiz, I. (2022). Imaging Biomarkers and Their Meaning for Molecular Imaging. In: Veit-Haibach, P., Herrmann, K. (eds) Artificial Intelligence/Machine Learning in Nuclear Medicine and Hybrid Imaging. Springer, Cham. https://doi.org/10.1007/978-3-031-00119-2_6
- Alberich-Bayarri, A., Jiménez-Pastor, A., Ferrer, B., Terol, M.J., Mayorga-Ruiz, I. (2022). Imaging Biobanks for Molecular Imaging: How to Integrate ML/AI into Our Databases. In: Veit-Haibach, P., Herrmann, K. (eds) Artificial Intelligence/Machine Learning in Nuclear Medicine and Hybrid Imaging. Springer, Cham. https://doi.org/10.1007/978-3-031-00119-2_8

Comunicaciones orales

- Jimenez-Pastor A, Alberich Bayarri A, Lopez-Gonzalez R, Alegre A, Torregrosa A, Marti-Bonmati L. (2018) Automatic liver segmentation on late phase MR images by using a Multi-Planar Convolutional Neural Network. EuSoMII Annual Meeting, Valencia, España.

Anexo III. Publicaciones

- Jimenez-Pastor A, Alberich-Bayarri A, Garcia-Castro F, Marti-Bonmati L. (2019) Automatic visceral fat characterization on CT scans through Deep Learning and CNN for the assessment of metabolic syndrome. 25th European Congress of Radiology (ECR 2019). Vienna, Austria.
- Jimenez-Pastor A, Alberich Bayarri A, Lopez-Gonzalez R, Garcia-Junco M, Marti-Aguado D, Marti-Bonmati L (2019). Zero-click liver PDFF and R2* quantification by deeply-supervised convolutional neural networks applied to MECSE MR. ESGAR 2019: 30th Annual meeting and postgraduate course (European Society of Gastrointestinal and Abdominal Radiology). Roma, Italia.
- Jimenez-Pastor A. (2021) AI in automated organs segmentation. 27th European Congress of Radiology (ECR 2021). Online.
- Jimenez-Pastor A. (2022) Generative adversarial networks in medical imaging. 28th European Congress of Radiology (ECR 2022). Vienna, Austria.

Posters

- Jimenez-Pastor A, Lopez-Gonzalez R, Garcia-Castro F, Mañas-Garcia A, Hernandez-Navarro R, Marti-Bonmati L, Alberich-Bayarri A. (2018) Big data generation by fully automated biomarker quantification in medical imaging using deeply supervised convolutional neuronal networks. 3rd TOPIM TECH – Big Data in Imaging (European Society for Molecular Imaging). Creta, Grecia.
- Jimenez-Pastor A, Alberich Bayarri A, Lopez-Gonzalez R, Marti-Bonmati L. (2019) Deeply supervised networks for the automated liver segmentation and quantification on MECSE-MRI. 25th European Congress of Radiology. Vienna, Austria.

Artículos JCR

- Jimenez-Pastor A, Alberich-Bayarri A, Lopez-Gonzalez R, Marti-Aguado D, França M, San Martin Bachmann R, Mazzucco J, Marti-Bonmati L. (2021) Precise whole liver automatic segmentation and quantification of PDFF and R2* on MR images. Eur Radiol; 10.1007/s00330-021-07838-5.

- Martí-Aguado D, Jiménez-Pastor A, Alberich-Bayarri Á, Rodríguez-Ortega A, Alfaro-Cervello C, Mestre-Alagarda C, Bauza M, Gallén-Peris A, Valero-Pérez E, Ballester MP, Gimeno-Torres M, Pérez-Girbés A, Benlloch S, Pérez-Rojas J, Puglia V, Ferrández A, Aguilera V, Escudero-García D, Serra MA, Martí-Bonmatí L. (2021) Automated Whole-Liver MRI Segmentation to Assess Steatosis and Iron Quantification in Chronic Liver Disease. *Radiology*.:211027. doi: 10.1148/radiol.2021211027.

Otros artículos JCR publicados en escenarios clínicos en el marco colaborativo de la tesis

- Puig J, Biarnés C, Daunis-I-Estadella P, Blasco G, Gimeno A, Essig M, Balaña C, Alberich-Bayarri A, Jimenez-Pastor A, Camacho E, Thio-Henestrosa S, Capellades J, Sanchez-Gonzalez J, Navas-Martí M, Domenech-Ximenes B, Del Barco S, Puigdemont M, Leiva-Salinas C, Wintermark M, Nael K, Jain R, Pedraza S. (2019) Macrovascular Networks on Contrast-Enhanced Magnetic Resonance Imaging Improves Survival Prediction in Newly Diagnosed Glioblastoma. *Cancers (Basel)*. 11(1):84. doi: 10.3390/cancers11010084. PMID: 30646519; PMCID: PMC6356693.
- Jimenez-Pastor A, Alberich-Bayarri A, Fos-Guarinos B, Garcia-Castro F, Garcia-Juan D, Glocker B, Marti-Bonmati L. (2020) Automated vertebrae localization and identification by decision forests and image-based refinement on real-world CT data. *Radiol Med*. 125(1):48-56. doi: 10.1007/s11547-019-01079-9. Epub 2019 Sep 14. PMID: 31522345.
- Martí-Bonmatí L, Alberich-Bayarri Á, Ladenstein R, Blanquer I, Segrelles JD, Cerdá-Alberich L, Gkontra P, Hero B, García-Aznar JM, Keim D, Jentner W, Seymour K, Jiménez-Pastor A, et al. (2020). PRIMAGE project: predictive in silico multiscale analytics to support childhood cancer personalised evaluation empowered by imaging biomarkers. *Eur Radiol Exp*. 4(1):22. doi: 10.1186/s41747-020-00150-9. PMID: 32246291; PMCID: PMC7125275.
- Puig J, Biarnés C, Pedraza S, Vilanova JC, Pamplona R, Fernández-Real JM, Brugada R, Ramos R, Coll-de-Tuero G, Calvo-Perxas L, Serena J, Ramió-Torrentà L, Gich J, Gallart L, Portero-Otin M, Alberich-Bayarri A, Jimenez-

Anexo III. Publicaciones

- Pastor A, et al. (2020). The Aging Imageomics Study: rationale, design and baseline characteristics of the study population. *Mech Ageing Dev.* 189:111257. doi: 10.1016/j.mad.2020.111257. Epub 2020 May 11. PMID: 32437737.
- Bruixola G, Remacha E, Jiménez-Pastor A, Dualde D, Viala A, Montón JV, Ibarrola-Villava M, Alberich-Bayarri Á, Cervantes A. (2021). Radiomics and radiogenomics in head and neck squamous cell carcinoma: Potential contribution to patient management and challenges. *Cancer Treat Rev.* 99:102263. doi: 10.1016/j.ctrv.2021.102263. Epub 2021 Jul 26. PMID: 34343892.
 - Nan Y, Ser JD, Walsh S, Schönlieb C, Roberts M, Selby I, Howard K, Owen J, Neville J, Guiot J, Ernst B, Pastor A, Alberich-Bayarri A, Menzel MI, Walsh S, Vos W, Flerin N, Charbonnier JP, van Rikxoort E, Chatterjee A, Woodruff H, Lambin P, Cerdá-Alberich L, Martí-Bonmatí L, Herrera F, Yang G. Data harmonisation for information fusion in digital healthcare: A state-of-the-art systematic review, meta-analysis and future research directions. *Inf Fusion.* 2022 Jun;82:99-122. doi: 10.1016/j.inffus.2022.01.001. PMID: 35664012; PMCID: PMC8878813.
 - Herrero Vicent C, Tudela X, Moreno Ruiz P, Pedralva V, Jiménez Pastor A, Ahicart D, Rubio Novella S, Meneu I, Montes Albuixech Á, Santamaria MÁ, Fonfria M, Fuster-Matanzo A, Olmos Antón S, Martínez de Dueñas E. Machine Learning Models and Multiparametric Magnetic Resonance Imaging for the Prediction of Pathologic Response to Neoadjuvant Chemotherapy in Breast Cancer. *Cancers (Basel).* 2022 Jul 19;14(14):3508. doi: 10.3390/cancers14143508. PMID: 35884572; PMCID: PMC9317428.
 - Jimenez-Pastor A, Lopez-Gonzalez R, Fos-Guarinos B, Garcia-Castro F, Wittenberg M, Torregrosa-Andrés A, Marti-Bonmati L, Garcia-Fontes M, Duarte P, Gambini JP, Bittencourt LK, Kitamura FC, Venugopal VK, Mahajan V, Ros P, Soria-Olivas E, Alberich-Bayarri A. Automated prostate multi-regional segmentation in magnetic resonance using fully convolutional neural networks. *Eur Radiol.* 2023 Jan 24. doi: 10.1007/s00330-023-09410-9. Epub ahead of print. PMID: 36690774.

Anexo III. Publicaciones

- Veiga-Canuto D, Cerdà-Alberich L, Jiménez-Pastor A, Carot Sierra JM, Gomis-Maya A, Sangüesa-Nebot C, Fernández-Patón M, Martínez de Las Heras B, Taschner-Mandl S, Düster V, Pötschger U, Simon T, Neri E, Alberich-Bayarri Á, Cañete A, Hero B, Ladenstein R, Martí-Bonmatí L. Independent Validation of a Deep Learning nnU-Net Tool for Neuroblastoma Detection and Segmentation in MR Images. *Cancers (Basel)*. 2023 Mar 6;15(5):1622. doi: 10.3390/cancers15051622. PMID: 36900410; PMCID: PMC10000775.
- Kondylakis H, Kalokyri V, Sfakianakis S, Marias K, Tsiknakis M, Jimenez-Pastor A, Camacho-Ramos E, Blanquer I, Segrelles JD, López-Huguet S, Barelle C, Kogut-Czarkowska M, Tsakou G, Siopis N, Sakellariou Z, Bizopoulos P, Drossou V, Lalas A, Votis K, Mallol P, Marti-Bonmati L, Alberich LC, Seymour K, Boucher S, Ciarrocchi E, Fromont L, Rambla J, Harms A, Gutierrez A, Starmans MPA, Prior F, Gelpi JL, Lekadir K. Data infrastructures for AI in medical imaging: a report on the experiences of five EU projects. *Eur Radiol Exp*. 2023 May 8;7(1):20. doi: 10.1186/s41747-023-00336-x. PMID: 37150779; PMCID: PMC10164664.