



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

**UNIVERSITAT POLITÈCNICA DE
VALÈNCIA**

Doctorado en Telecomunicación

**Investigación y desarrollo de
metodología avanzada de
segmentación de la médula
espinal cervical a partir de
imágenes RM para la ayuda al
diagnóstico y seguimiento de
pacientes de Esclerosis Múltiple**

América Bueno Gómez

DIRECTORES DE TESIS

Ignacio Bosch Roig

Angel Alberich-Bayarri

Valencia, Abril 2024

*A mi familia, refugio constante y seguro.
A mi madre, luchadora y fuerte como un roble, mi mejor amiga, mi piedra angular.
A mi padre, sereno e infatigable.
A mis hermanos, eternos escuderos, siempre daré gracias por poder crecer con vosotros.
A Paula, amiga y consejera, mamá de mi sobrino Diego, mi nueva debilidad.
A Guille, mi compañero en cada remonte. Las subidas son costosas, pero que siempre nos merezca la pena.
A mi yaya América, fuente de amor y dulzura incondicional, compañera del mejor, mi yayo Diego, risueño y despierto como él sólo, a quien siempre he procesado la más grande admiración, a los dos juntos, mis yayos, gracias por esos años tan felices. Siempre os admiraré, siempre os querré con todo mi corazón. Mi aspiración: ser capaz de crear tanto bien como vosotros.
A mis tutores, Angel y Nacho, por vuestra experiencia y guía en el camino.
A Jorge, por su generosidad y siempre atención desinteresada.
A todos los que componéis mi mundo, sabéis quiénes sois, gracias por ese abrazo imperecedero.*



Resumen

La Esclerosis Múltiple (EM) es una enfermedad inflamatoria y autoinmune del sistema nervioso central (SNC) con rasgos de desmielinización y degeneración axonal en el tiempo, y caracterizada por ser muy heterogénea en los síntomas y en el curso de la enfermedad. Suele debutar en personas de entre 20 y 40 años y es una de las principales causas de discapacidad en adultos jóvenes en Estados Unidos y Europa. La Imagen de Resonancia Magnética (RM) es una de las herramientas clínicas más sensibles para la evaluación de los procesos inflamatorios y neurodegenerativos, ya que, mediante ellas, los especialistas son capaces de cuantificar la carga lesional, la diseminación espacial y la evolución longitudinal de la enfermedad en el tiempo. En los últimos años, la evaluación de la médula espinal ha tenido un creciente interés clínico para mejorar el diagnóstico y el fenotipado de la enfermedad, aunque, a diferencia del cerebro, donde se ha concentrado gran parte de la investigación en técnicas de segmentación y volumetría, en médula espinal cervical no existen algoritmos de inteligencia artificial (IA) desarrollados y certificados para práctica clínica que faciliten su identificación y posterior estudio. Es por ello, que nuestro objetivo se centra en investigar y desarrollar un método automático de segmentación de médula cervical en RM, facilitando así una evaluación automática y mejorada de la atrofia de la médula espinal, pues esta puede proporcionar información valiosa sobre la progresión de la enfermedad y sus consecuencias clínicas. De esta forma, el especialista podrá incorporar nueva información en el seguimiento de los pacientes, como la volumetría de la médula, salvando la laboriosa tarea de segmentación manual.

El algoritmo se desarrolló mediante datos del mundo real (*real-world data*) recogidos de manera retrospectiva en 121 pacientes de EM. Se utilizaron 96 de ellos para el entrenamiento del modelo, 25 para test y 13 para la validación del modelo propuesto, estos últimos con fecha de adquisición 5 años más tarde a los datos utilizados para entrenamiento y test. Durante la tesis se trabajaron secuencias de RM adquiridas en un equipo de 3T (SignaHD, GEHC), de tipo 3D axiales potenciadas en T1, dada su mejor resolución y contraste para identificar pequeñas estructuras anatómicas como la médula espinal. El etiquetado manual de los datos fue realizado bajo el consejo y supervisión de dos radiólogos experimentados, entre los cuales se resolvieron posibles discrepancias con un tercer radiólogo, obteniendo finalmente un conjunto de máscaras de médula espinal cervical como *ground-truth*.

Varias fueron las arquitecturas, hiperparámetros y formas de preprocesado aplicados al dataset en busca de la solución óptima. Dada su conocida importancia en la segmentación de imagen

Resumen

médica, la arquitectura U-Net fue el punto de partida, probando algunas de sus modificaciones como Attention-UNet, Deep U-Net o U-Net ++. Tras la ausencia de buenos resultados y una mayor investigación en el campo, se dio con la problemática del desbalanceo de datos, por lo que utilizar hiperparámetros como funciones de pérdidas ponderadas (Tversky Loss, Weighted Cross Entropy, Focal Loss...) y métricas más allá del tradicional 'accuracy' como el empleado coeficiente de correlación de Matthew (MCC), fue determinante para la llegada de segmentaciones cada vez más próximas al *ground truth*, pudiendo así descartar técnicas de preprocesado. Finalmente, para obtener la segmentación deseada, se implementó y entrenó una red neuronal convolucional 2D compuesta por un mecanismo de atención residual y conexiones basadas en la arquitectura U-Net. El mecanismo de atención permitió que el modelo se centrara en aquellas localizaciones de la imagen que son importantes para la tarea de clasificación de los vóxeles correspondientes a la médula cervical, a la vez que retenía la información del resto de estructuras anatómicas, mientras que los bloques residuales nos permitieron solventar problemas de desvanecimiento de gradiente comunes en redes neuronales profundas. El entrenamiento se diseñó con una función de pérdidas local, basada en el índice de Tversky con el fin de controlar el problema de desbalanceo de datos de imagen médica, y un buscador automático de tasa de aprendizaje óptima que nos permitió mejorar la convergencia y rendimiento del modelo.

Finalmente, nuestro método proporcionó una segmentación con una elevada tasa de acierto, obteniendo un valor de 0.95 como MCC en la métrica de entrenamiento y consiguiendo en validación un coeficiente DICE de 0.904 ± 0.101 tomando como referencia la segmentación manual. Esta tasa de acierto se reducía en algunas zonas distales donde la zona de transición entre médula cervical y líquido cefalorraquídeo es difusa y poco nítida debido al efecto de volumen parcial.

Este método automático de segmentación de médula espinal en RM 3D potenciadas en T1 tendrá una trascendencia directa sirviendo como primer paso para acelerar el diagnóstico de la EM e incrementar recursos en el seguimiento de la evolución de la enfermedad con el fin de definir los factores que hay tras el aumento de la discapacidad irreversible, uno de los principales síntomas de la EM. En consideración a ello, además de obtener una herramienta para la segmentación automática de la médula, también creamos un módulo para el cálculo de sus dimensiones, actuando como biomarcador de imagen, lo que será útil y eficaz para la valoración de la atrofia. La atrofia es un indicador directo del daño neuronal y de la pérdida de tejido tanto en el cerebro como en la médula espinal, siendo un factor clave indicador del riesgo de discapacidad en la EM. Al calcular de forma precisa la atrofia, los clínicos pueden evaluar el grado

Resumen

de daño neurológico y seguir su evolución a lo largo del tiempo evitando la laboriosa segmentación manual de toda la estructura medular, pues al tratarse de un enfermedad crónica y progresiva, se necesita de la obtención reiterada de información actualizada. En nuestro estudio, como posibles biomarcadores de imagen, calculamos las dimensiones de las médulas de nuestros pacientes en forma de volumen (mm^3) y sección media (mm^2), y estudiamos la relación entre sección media de la médula espinal cervical con la distribución de las distintas formas clínicas que componen nuestro grupo de estudio. Por último, también estudiamos la relación entre los niveles en la Escala de Discapacidad Extendida de Kurtzke (EDSS) de cada paciente de nuestro grupo de estudio con su valor de sección media de médula espinal cervical calculado por nuestro método.

Resum

L'Esclerosi Múltiple (EM), és una malaltia inflamatòria i autoimmunitària del sistema nerviós central (SNC) amb trets de desmielinització i degeneració axonal en el temps. Es caracteritza per ser molt heterogènia amb els símptomes i curs de la malaltia. És una de les principals causes de discapacitat en adults joves d'entre 20 i 40 anys als Estats Units i Europa. La Imatge de Ressonància Magnètica (RM) és una de les eines més sensibles per a l'avaluació dels processos inflamatoris i neurodegeneratius, ja que, mitjançant elles, els especialistes són capaços de quantificar la càrrega lesional, la disseminació espacial i l'evolució longitudinal de la malaltia en el temps. Als darrers anys, l'evolució de la medul·la espinal ha tingut un creixent interès clínic per tal de millorar el diagnòstic i el fenotipatge de la malaltia, encara que, a diferència del cervell, on s'ha concentrat gran part de la investigació en tècniques de segmentació i volumetria, en medul·la espinal cervical no existeixen algorismes d'intel·ligència artificial (IA) desenvolupats i certificats per a la pràctica clínic que faciliten la seva identificació i posterior estudi. Aquest fet motiva el present estudi, que se centra en la recerca i desenvolupament d'un mètode automàtic de segmentació de medul·la cervical en RM. L'automatització i millora del procés d'avaluació de l'atròfia de la medul·la espinal podrà proporcionar valuosa informació sobre la progressió de la malaltia i les seves conseqüències clíniques. D'aquesta forma, l'especialista podrà incorporar nova informació, tal com la volumetria de la medul·la, al seguiment dels pacients, salvant la laboriosa tasca derivada de la segmentació manual.

L'algoritme proposat al present treball va ser desenvolupat mitjançant dades del món real (*real-world data*) recollides de manera retrospectiva en 121 pacients d'EM. D'aquestes mostres, 96 foren utilitzades per a l'entrenament del model d'IA, 13 per a la validació durant l'entrenament i les 25 restants com a conjunt d'avaluació. Les seqüències d'imatges de RM fetes servir foren adquirides amb un equip 3T (SignaHD, GEHC), de tipus 3D axials potenciats en T1, donada la seua millor resolució i contrast alhora identificar petites estructures anatòmiques com la medul·la espinal. L'etiquetatge de les dades fou realitzat sota la supervisió i consell de dos experimentats radiòlegs, les discrepàncies dels quals, van ser resoltes per un tercer radiòleg. El resultat final fou un conjunt d'imatges RM de referència (*ground truth dataset*) amb les corresponents màscares de segmentació de la medul·la espinal cervical definides pels radiòlegs.

Diverses van ser les arquitectures, hiperparàmetres i tècniques de preprocessat aplicades al conjunt de dades en cerca de la solució òptima. Donada la seua coneguda importància en la segmentació d'imatge mèdica, l'arquitectura U-Net fou el punt de partida. Addicionalment,

Resum

s'incloueren en l'estudi més recent modificacions de l'arquitectura esmentada tals com Attention-UNet, Deep U-Net o U-Net. Un altre punt d'inflexió fou resoldre la problemàtica de la desproporció de representativitat al conjunt de dades utilitzat (*dataset imbalance*). Noves funcions de cost ponderat (*Tversky loss*, *Weighted Cross-Entropy*, *Focal loss*...) foren incloses en l'espai d'hiperparàmetres. Mètriques d'avaluació més adients al problema, tal com el coeficient de correlació de *Matthew (MCC)* (en lloc de la mètrica d'*accuracy*), foren determinants a l'hora d'aconseguir entrenar un model de segmentació d'imatges cada vegada més precís i pròxim a les segmentacions de referència dels radiòlegs i que, a més, permeteren estalviar-se l'ús de tècniques de preprocessat. Finalment, per obtindre la segmentació desitjada, es va implementar i entrenar una xarxa neuronal convolucional 2D composta per un mecanisme d'atenció residual i connexions basades en l'arquitectura U-Net. El mecanisme d'atenció va permetre que el model se centrara en aquelles localitzacions de la imatge més importants per a la tasca de classificació dels corresponents vòxels a la medul·la cervical, a la volta que retenia la informació de la resta d'estructures anatòmiques. Alhora, els blocs residuals, van permetre resoldre els problemes d'esvaïment de gradient, comuns a l'entrenament de xarxes neuronals profundes. L'entrenament es va dissenyar amb una funció de cost local, basada en l'índex *Tversky*, amb el fi de controlar la problemàtica del *dataset imbalance* i, un buscador automàtic de la taxa d'aprenentatge òptima que permetia una millor convergència i rendiment del model.

Els resultats proporcionats pel nostre mètode de segmentació automàtica, presentaren una elevada taxa d'encert, obtinguen un valor de 0.95 com coeficient de correlació de *Matthew* en la mètrica d'entrenament i aconseguint en validació un coeficient DICE de 0.904 ± 0.101 prenent com a referència la segmentació manual. Aquesta taxa d'encert es reduïa en algunes zones distals on la zona de transició entre la medul·la cervical i el líquid cefalorraquidi és difús i poc nítid, degut a l'efecte de volum parcial.

Aquest mètode de segmentació automàtica de medul·la espinal en RM 3D potenciat en T1, tindrà una transcendència directa, augmentant la prestesa al diagnòstic de l'EM. A més, incrementarà els recursos en el seguiment de l'evolució de la malaltia, els quals, permetran definir amb major precisió els factors que hi ha darrere de l'augment de la discapacitat irreversible, un dels principals símptomes de l'EM. En consideració a això, a més de l'eina de segmentació automàtica, també hem desenvolupat un mòdul per al càlcul de les seues dimensions, el que serà útil per a una eficaç valoració de l'atròfia. L'atròfia és un indicador directe del dany neuronal i de la perduda de teixit tant al cervell com a la medul·la espinal, sent un indicador clau del risc de discapacitat en l'EM. En calcular de forma precisa l'atròfia, els clínics poden avaluar el grau de dany neurològic i seguir la seua evolució al llarg del temps evitant la

Resum

laboriosa tasca que suposa la segmentació manual de tota l'estructura medul·lar. Aquest fet incrementa significativament l'eficiència del procés, ja que en tractar-se l'EM d'una malaltia crònica i progressiva, es necessita l'obtenció reiterada d'informació actualitzada. Al nostre estudi, com a possibles biomarcadors d'imatge, calcularem les dimensions de les medul·les dels nostres pacients en forma de volum (mm^3) i secció mitjana (mm^2), i estudiarem la relació entre secció mitjana de la medul·la espinal cervical amb la distribució de les distintes formes clíniques que componen el nostre grup d'estudi. Per acabar, també estudiarem la relació entre els nivells en l'escala de discapacitat estesa de Kurtzke (EDSS) de cada pacient amb el seu valor de secció mitjana de medul·la espinal cervical calculada pel nostre mètode.

Abstract

Multiple Sclerosis (MS) is an inflammatory and autoimmune disease of the central nervous system (CNS) with features of demyelination and axonal degeneration over time, and characterised by being very heterogeneous in symptoms, disease course and outcome. It usually affects people between 20 and 40 years old and is one of the leading causes of disability in young adults in the United States and Europe. Magnetic Resonance Imaging (MRI) is one of the most sensitive clinical tools for the evaluation of inflammatory and neurodegenerative processes, as specialists are able to quantify the lesion burden, spatial spread and longitudinal evolution of the disease over time. In recent years, the evaluation of the spinal cord has been of increasing clinical interest to improve the diagnosis and phenotyping of the disease, although, unlike the brain, where much of the research has focused on segmentation and volumetric techniques, in the cervical spinal cord there are no artificial intelligence (AI) algorithms developed and certified for clinical practice to facilitate its identification and subsequent study. Therefore, our aim is to investigate and develop an automatic method of cervical cord segmentation in MRI, thus facilitating an automatic and improved assessment of spinal cord atrophy, which can provide valuable information on the progression of the disease and its clinical consequences. In this way, the specialist will be able to incorporate new information into patient follow-up, such as spinal cord volumetry, saving the time-consuming task of manual segmentation.

The algorithm was developed using real-world data collected retrospectively from 121 MS patients. Of these, 96 were used for model training, 25 for testing and 13 for validation of the proposed model, the latter with an acquisition date 5 years later than the data used for training and testing. During the thesis, 3D axial T1-weighted MRI sequences acquired in 3T equipment (SignaHD, GEHC) were used, given their better resolution and contrast to identify small anatomical structures such as the spinal cord. Manual labelling of the data was performed under the advice and supervision of two experienced radiologists, between whom possible discrepancies were resolved with a third radiologist, resulting in a set of cervical spinal cord masks as ground-truth.

Several architectures, hyperparameters and forms of pre-processing were applied to the dataset in search of the optimal solution. Given its known importance in medical image segmentation, the U-Net architecture was the starting point, testing some of its variations such as Attention-UNet, Deep U-Net or U-Net. After the absence of good results and further research in the field, the problem of data imbalance was identified, so the use of hyperparameters such as weighted

Abstract

loss functions (Tversky Loss, Weighted Cross Entropy, Focal Loss...) and metrics beyond the traditional 'accuracy' such as Matthew's Correlation Coefficient (MCC), was decisive for the arrival of segmentations increasingly closer to ground truth, thus being able to discard pre-processing techniques. Finally, to obtain the desired segmentation, a 2D convolutional neural network (CNN) composed of a residual attention mechanism and connections based on the U-Net architecture was implemented and trained. The attention mechanism allowed the model to focus on those image locations that are important for the classification task of the voxels corresponding to the cervical cord, while retaining the information of the rest of the anatomical structures. Residual blocks allowed us to solve common gradient fading problems in deep neural networks. Training was designed with a local loss function, based on the Tversky index in order to control the medical image data imbalance problem, and an automatic optimal learning rate finder that allowed us to improve the convergence and performance of the model.

Finally, our method provided a segmentation with a high success rate, obtaining a value of 0.95 as MCC in the training metric and obtaining in validation a DICE coefficient of 0.904 ± 0.101 taking manual segmentation as a reference. This success rate was reduced in some distal areas where the transition zone between cervical cord and cerebrospinal fluid is diffuse and unclear due to the partial volume effect. This automatic segmentation method of the spinal cord in 3D T1-weighted MRI will be of direct importance as a first step to accelerate the diagnosis of MS and increase resources in disease evolution monitoring in order to define the factors behind the increased irreversible disability, one of the main MS symptoms. This is the reason why, in addition to obtaining a tool for the automatic segmentation of the spinal cord, we also created a module for the calculation of its dimensions, which will be useful and effective for the assessment of atrophy. Atrophy is a direct indicator of neuronal damage and tissue loss in both the brain and spinal cord, and is a key risk factor for disability in MS. By accurately calculating atrophy, clinicians can assess the degree of neurological damage and follow its evolution over time, avoiding time-consuming manual segmentation of the entire spinal cord structure, as EM is a chronic and progressive disease, it is necessary to obtain updated information on a regular basis. In our study, we calculated the dimensions of our patients' cords, as possible imaging biomarkers, in terms of volume (mm^3) and mean section (mm^2), and studied the relationship between the mean section of the cervical spinal cord with the distribution of the different clinical forms in our study group. Finally, we also studied the relationship between the Kurtzke Expanded Disability Status Scale (EDSS) levels of each patient in our study group with their cervical spinal cord midsection value calculated by our method.

Motivación

La esclerosis múltiple (EM) es una enfermedad crónica del sistema nervioso central (SNC) que afecta al cerebro y la médula espinal. Es una enfermedad autoinmune, lo que significa que el sistema inmunológico del cuerpo ataca de manera errónea a la mielina, la capa protectora que recubre las fibras nerviosas. Este daño a la mielina interrumpe la transmisión eficiente de los impulsos nerviosos causando una amplia variedad de síntomas que varían según la persona y la progresión de la enfermedad. Aunque se ha avanzado en la comprensión de la EM, las causas exactas aún no se comprenden completamente. Se cree que una combinación de factores genéticos y ambientales puede contribuir al desarrollo de la enfermedad. Los síntomas más comunes incluyen fatiga, debilidad muscular, problemas de equilibrio y coordinación, dificultades visuales, hormigueo, entumecimiento y problemas cognitivos. Hay varios subtipos de EM, y la forma en que se manifiesta puede variar. Los tipos más comunes son la EM recurrente-remitente (EMRR), la EM secundaria progresiva (EMSP) y la EM primaria progresiva (EMPP).

El diagnóstico de la EM es un proceso complejo, pues los síntomas son variados y pueden ocasionar la confusión con otras enfermedades similares en sintomatología. Para llevarlo a cabo, se utilizan una combinación de pruebas clínicas, estudios de imagen como la resonancia magnética (RM) y análisis de líquido cefalorraquídeo (LCR).

La interpretación de la RM es un proceso no trivial, y la evaluación de la médula espinal se obvia en muchos casos por lo laborioso que resulta la segmentación manual de la médula, y lo sutiles que son sus cambios morfológicos. Es por ello que, en atención a mejorar el flujo de trabajo de los radiólogos, facilitar la interpretación de la médula y cubrir la necesidad de comprender mejor esta enfermedad neurológica y desarrollar tratamientos más efectivos, se propone desarrollar un método de segmentación automática de médula espinal a partir de imágenes de RM mediante técnicas de Inteligencia Artificial (IA) que permita automatizar la caracterización morfológica, proponiendo nuevos biomarcadores de imagen. Con ello se pretende facilitar el diagnóstico y seguimiento de la enfermedad, pues, al tratarse de una patología crónica y progresiva, resulta conveniente la existencia de herramientas que nos faciliten realizar un seguimiento constante y sean cuantitativas. Este seguimiento, nos permitirá caracterizar daños neurológicos característicos de la enfermedad y su progresión (como la atrofia de la médula espinal), identificar la EM en cada etapa o estudiar el efecto de los tratamientos en su progresión y control de los síntomas. Estos tratamientos pueden incluir desde medicamentos modificadores

Motivación

de la enfermedad, terapias de rehabilitación o medicamentos para el manejo de síntomas específicos. De esta manera, se pretende desarrollar terapias más específicas y personalizadas que sean efectivas y sin efectos secundarios, pues no todos los pacientes responden de la misma forma.

La EM tiene un impacto significativo en la calidad de vida de quienes la padecen. Investigar esta enfermedad es crucial para encontrar formas de aliviar los síntomas, mejorar la función neurológica y proporcionar a los pacientes una mejor calidad de vida, proporcionando esperanza a las personas afectadas por la EM y a sus familias.

Organización de la tesis doctoral

Esta tesis doctoral ha sido organizada en los siguientes capítulos y apartados:

Capítulo 1 se introduce la enfermedad de EM tanto a nivel biológico como clínico, así como conceptos relacionados con su diagnóstico mediante imagen. Se introducen conceptos sobre imagen médica y técnicas de IA en segmentación de imagen médica.

Capítulo 2 se introduce la hipótesis científica a la problemática y los objetivos como metas específicas a alcanzar para conseguir el propósito general de la tesis.

Capítulo 3 se describen las técnicas utilizadas para el desarrollo de la metodología implementada, desde la adquisición de la base de datos hasta el diseño de la arquitectura de IA entrenada para nuestro método propuesto.

Capítulo 4 se presentan los resultados obtenidos y se elabora una comparación con el estado del arte a modo de discusión.

Capítulo 5 se presentan las principales conclusiones y líneas futuras de investigación.

Referencias conjunto de fuentes de información utilizadas durante el desarrollo de la tesis.

Anexo I se introduce la detección de lesiones en médula espinal como línea futura y se describe el estado de su desarrollo actual.

Anexo II se presentan las contribuciones científicas aportadas mediante el desarrollo de la presente tesis, incluyendo congresos y publicación en revista.

Anexo III contiene el índice de abreviaturas utilizados durante la presente tesis.

Organización de la tesis doctoral

Resumen.....	iii
Resum.....	vi
Abstract.....	ix
Motivación.....	xi
Organización de la tesis doctoral.....	xiii
1 Capítulo 1. Introducción.....	1
1.1 CLÍNICA DE LA ESCLEROSIS MÚLTIPLE (EM).....	2
1.1.1 <i>Introducción a la Esclerosis Múltiple</i>	2
1.1.2 <i>Fenotipos y cursos clínicos de la enfermedad</i>	4
1.1.3 <i>Modificadores de actividad y progresión en EM</i>	10
1.1.4 <i>Escala de Discapacidad Extendida de Kurtzke (EDSS)</i>	12
1.2 IMAGEN MÉDICA.....	14
1.2.1 <i>Modalidades</i>	15
1.2.2 <i>El formato DICOM</i>	21
1.2.3 <i>Conversión de DICOM a Nifti</i>	22
1.2.4 <i>Imagen de resonancia magnética en Esclerosis Múltiple</i>	23
1.2.5 <i>Biomarcadores de imagen</i>	31
1.3 SEGMENTACIÓN EN IMAGEN MÉDICA.....	35
1.3.1 <i>Machine learning en segmentación de imagen médica</i>	36
1.3.2 <i>Aprendizaje profundo en segmentación de imagen</i>	39
1.3.3 <i>Arquitectura empleada: Residual Attention-UNet</i>	62
1.3.4 <i>Hiperparámetros en entrenamiento y test</i>	73
2 Capítulo 2. Hipótesis y objetivos.....	80
2.1 HIPÓTESIS.....	81
2.2 OBJETIVOS.....	81
3 Capítulo 3. Materiales y métodos.....	83
3.1 BASES DE DATOS.....	85
3.2 ANOTACIÓN/ETIQUETADO DE LA BASE DE DATOS.....	86
3.3 HARDWARE Y SOFTWARE.....	87
3.3.1 <i>Anaconda y entorno de GPU</i>	87
3.3.2 <i>Etiquetado manual: ITK-SNAP</i>	88
3.3.3 <i>Conversión de formato DICOM a Nifti</i>	89
3.4 PREPROCESADO DE DATOS.....	89
3.4.1 <i>Redimensionado</i>	90
3.4.2 <i>Normalización</i>	90
3.4.3 <i>Conversión de imágenes a tensor</i>	91
3.5 TÉCNICAS DE INTELIGENCIA ARTIFICIAL EMPLEADAS.....	92
3.5.1 <i>Arquitecturas evaluadas durante la investigación</i>	92
3.6 CONFIGURACIÓN DEL ENTRENAMIENTO Y MÉTRICAS EN TEST.....	95
3.7 BIOMARCADORES DE IMAGEN: CÁLCULO DE MÉTRICAS ESPACIALES DE LA MÉDULA CERVICAL.....	100
4 Capítulo 4. Resultados y discusión.....	103
4.1 RESULTADOS EN ENTRENAMIENTO.....	104
4.2 RESULTADOS EN TEST.....	106
4.3 RESULTADOS EN VALIDACIÓN.....	109
4.4 EJEMPLOS DE RESULTADOS CUALITATIVOS.....	111
4.4.1 <i>Discusión del método de segmentación propuesto</i>	119
4.5 COMPARACIÓN DE RESULTADOS ENTRE NUESTRO MÉTODO Y OTROS MODELOS DE SEGMENTACIÓN.....	121
4.6 RELACIÓN ENTRE BIOMARCADORES EXTRAÍDOS, FORMAS CLÍNICAS Y EDSS EN EM.....	127
4.6.1 <i>Discusión de estudio de relación entre biomarcadores de imagen extraídos, formas clínicas y EDSS</i>	131

Organización de la tesis doctoral

5	Capítulo 5. Conclusiones y líneas futuras	136
	Referencias	140
	Anexo I: Detección automática de lesiones en médula cervical	158
1	INTRODUCCIÓN	159
2	LESIONES EN MÉDULA ESPINAL	160
3	SECUENCIAS EMPLEADAS.....	162
3.1	<i>Secuencia Sagital T2</i>	162
3.2	<i>Secuencia STIR</i>	164
4	REGISTRO DE SECUENCIAS DE IMÁGENES MÉDICAS.....	168
4.3	<i>Antecedentes matemáticos</i>	169
4.4	<i>Componentes de registro</i>	170
5	METODOLOGÍA PROPUESTA.....	174
5.1	<i>Base de datos</i>	175
5.2	<i>Registro de imágenes</i>	175
5.3	<i>Detección de lesiones</i>	183
5.4	<i>Correlación entre lesiones y enfermedad</i>	187
	Anexo II: Contribuciones científicas	188
	Anexo III: Índice de abreviaturas	191
	Tabla de ilustraciones.....	192

Capítulo 1. Introducción

Capítulo 1. Introducción

La presente tesis doctoral recoge el trabajo llevado a cabo para alcanzar el objetivo principal, el cual es investigar y desarrollar una herramienta de IA automática capaz de segmentar la médula espinal cervical en pacientes de Esclerosis Múltiple a partir de secuencias de RM recolectadas en un entorno hospitalario común, así como combinarla con la extracción de biomarcadores de imagen de la médula. En el presente capítulo de Introducción, abordamos conceptos que nos permitirán contextualizar dicho objetivo primario, comenzando con la descripción de la esclerosis múltiple como enfermedad, tanto a nivel biológico como en el diagnóstico y seguimiento clínico, continuando con la imagen médica y su implicación a la hora de diagnosticar dicha enfermedad y finalizando con una sección sobre IA en el ámbito de la segmentación en imagen médica, compuesta por definiciones generales sobre redes convolucionales y exposiciones particulares sobre la arquitectura diseñada y empleada para conseguir nuestro objetivo. Dicha arquitectura, definida como Residual Attention-UNet, nos permitirá solventar el problema de desbalanceo de datos, introducido por el pequeño tamaño de la médula espinal cervical, nuestra región de interés, con respecto a la totalidad de la imagen, de manera que veremos cómo es capaz de provocar que el modelo se centre en estas pequeñas regiones relevantes de la imagen a la vez que es capaz de preservar información del resto de la imagen, dando lugar a una segmentación lo más fidedigna a la realidad posible.

1.1 Clínica de la Esclerosis Múltiple (EM)

La esclerosis múltiple (EM) es una enfermedad con un gran impacto socioeconómico en todo el mundo. Pese a que, con el tiempo y gracias al esfuerzo de profesionales y pacientes, se ha conseguido aumentar su conocimiento y atención, mejorando así la calidad de vida de las personas que la padecen, todavía sigue teniendo un elevado impacto en los pacientes y familiares que sufren la enfermedad, provocando discapacidad y reducción en la esperanza de vida. Es por ello necesario continuar con esta labor de investigación y difusión para perfeccionar su comprensión y mejorar la atención médica y calidad de vida de los pacientes.

1.1.1 Introducción a la Esclerosis Múltiple

La esclerosis múltiple (EM) es una enfermedad crónica inflamatoria del sistema nervioso central (SNC) con rasgos de desmielinización y degeneración axonal, y caracterizada por ser muy heterogénea en los síntomas, en el curso de la enfermedad y en los resultados (Compston & Coles, 2008). Sintomatológicamente, suele presentarse por primera vez en personas de entre 20 y 40 años y es una de las principales causas de discapacidad en adultos jóvenes en Estados

Capítulo 1. Introducción

Unidos y Europa (*Overview of the Epidemiology, Diagnosis, and Disease Progression Associated with Multiple Sclerosis - PubMed, n.d.*).

La desmielinización, es la pérdida de mielina, una sustancia grasa que rodea y aísla las fibras nerviosas en el SNC, incluyendo la médula espinal. Está compuesta principalmente por lípidos y proteínas y tiene un color blanco. La función principal de la mielina en la médula espinal es facilitar la transmisión rápida y eficiente de los impulsos eléctricos a lo largo de las fibras nerviosas. Actúa como un aislante eléctrico, formando una capa protectora alrededor de las fibras nerviosas llamada "vaina de mielina". Esta vaina interrumpe la propagación de la señal eléctrica a lo largo de la fibra y fuerza al impulso a "saltar" de un nodo de Ranvier (espacio entre las secciones de mielina) al siguiente. Este proceso se conoce como "conducción saltatoria".

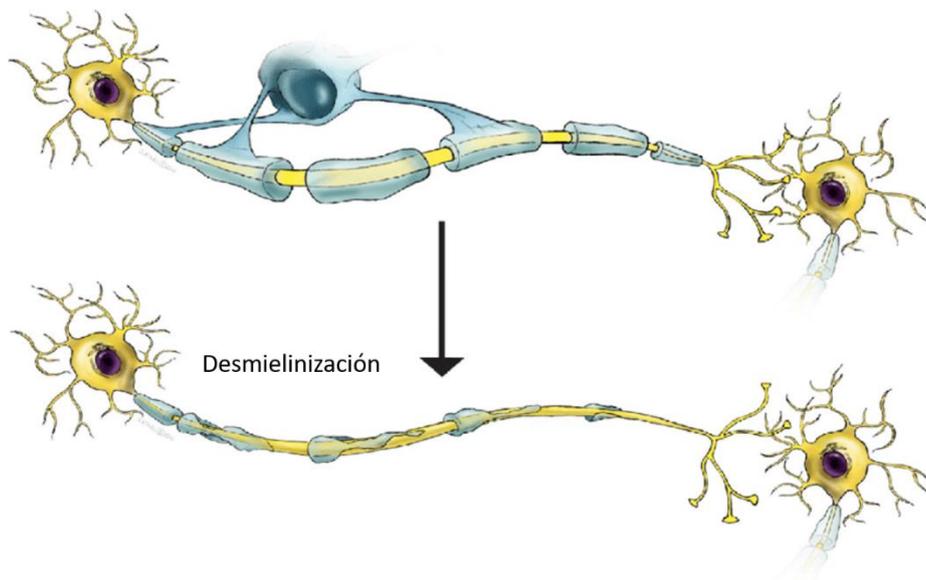


Figura 1. Representación gráfica de la desmielinización. (R. H. Miller et al., 2017)

La conducción saltatoria permite que los impulsos nerviosos viajen a velocidades mucho más altas que si tuvieran que recorrer toda la longitud de la fibra nerviosa sin mielina. Esto es especialmente importante en la médula espinal, donde se transmiten y procesan rápidamente las señales sensoriales y motoras desde y hacia el cerebro. La mielina ayuda a acelerar la velocidad de transmisión de los impulsos, lo que permite una respuesta rápida y coordinada del sistema nervioso.

Además de su función de conducción saltatoria, la mielina también proporciona soporte y protección a las fibras nerviosas en la médula espinal. Ayuda a mantener la integridad estructural de las fibras y las protege contra daños o lesiones.

Capítulo 1. Introducción

La EM se considera una enfermedad autoinmune en la que las células del sistema inmunológico, llamadas linfocitos T y linfocitos B, ingresan al SNC y atacan la mielina, provocando su daño o destrucción. Como resultado de la desmielinización, se forman lesiones o placas en áreas dispersas del cerebro y la médula espinal. Estas lesiones interrumpen o bloquean la transmisión de los impulsos nerviosos, lo que puede causar una amplia variedad de síntomas, dependiendo de las áreas del SNC afectadas.

Los síntomas de la EM pueden variar enormemente entre los individuos y pueden incluir fatiga, debilidad muscular, dificultades para caminar, problemas de equilibrio y coordinación, trastornos visuales, problemas de memoria y concentración, trastornos del habla y hormigueo o entumecimiento en diversas partes del cuerpo. La desmielinización en la EM también puede causar inflamación en el SNC, lo que agrava aún más los daños y la formación de lesiones. Con el tiempo, puede ocurrir una pérdida permanente de las funciones neurológicas afectadas.

Aunque no se conoce la causa exacta de la EM, se cree que factores genéticos y ambientales, incluyendo las infecciones fúngicas (Benito-León & Laurence, 2017), desempeñan un papel en el desarrollo de la enfermedad. El tratamiento de la EM generalmente se enfoca en el manejo de los síntomas, la reducción de las recaídas y la ralentización de la progresión de la enfermedad a través de medicamentos inmunomoduladores y terapias de rehabilitación.

1.1.2 Fenotipos y cursos clínicos de la enfermedad

La EM es una enfermedad heterogénea, puede variar considerablemente de una persona a otra en términos de características clínicas y manifestaciones. Esta variabilidad ha llevado a la identificación de varios fenotipos diferentes que ayudan a los médicos a entender y tratar la enfermedad de manera más efectiva. Estos fenotipos son conjuntos de estas características clínicas y manifestaciones (síntomas, patrones de discapacidad, evolución de la enfermedad...) que se emplean para clasificar y categorizar a los pacientes con EM en grupos que comparten similitudes en cuanto a cómo se manifiesta la enfermedad en ellos.

Los primeros fenotipos de esclerosis múltiple definidos formalmente son: EM Recurrente Remitente (EMRR); EM Primaria Progresiva (EMPP); EM Secundaria Progresiva (EMSP) y EM Progresiva Recurrente (EMPR), fueron propuestos en 1996 por la *U.S. National Multiple Sclerosis Society (NMSS)*, comité de prevención en pruebas clínicas de EM, debido a la creciente necesidad de estandarizar la terminología en el campo de la enfermedad. Se consideró que la falta de esta terminología sería perjudicial, no sólo para la práctica clínica y la comunicación entre especialistas, sino también para los futuros avances en investigación clínica. De esta manera,

Capítulo 1. Introducción

esta terminología unificada garantizaría el correcto diseño de los ensayos clínicos y la homogeneidad de la población reclutada, proporcionando así las bases necesarias para obtener buenos resultados. Sin embargo, el comité también fue consciente de que la naturaleza de los fenotipos era puramente clínica y que, por lo tanto, podría cambiar con el tiempo (Lublin & Reingold, 1996).

Esta nomenclatura se convirtió rápidamente en una parte inherente de la práctica clínica y de investigación y fue, en parte, decisiva en el proceso de aprobación de nuevos estudios terapéuticos de EM. Sin embargo, conforme aumentaba el conocimiento de la patología, se volvían evidentes las limitaciones de los fenotipos de origen puramente clínicos y la falta de imágenes y de correlaciones biológicas.

En 2012, el comité (apoyado por NMSS y el comité europeo para el tratamiento y la investigación en EM) revisó los fenotipos clínicos originales con el objetivo de mejorar la terminología a la vez que se incorporan como herramientas de diagnóstico la imagen y los biomarcadores de fluido corporales como el líquido cefalorraquídeo (LCR), la sangre o el suero sanguíneo. El comité recomendó mantener los fundamentos de los fenotipos originales de 1996 pero añadiendo una caracterización mejorada introduciendo nuevos descriptores de actividad y de progresión. El comité también informó sobre dos nuevos cursos de la EM: el síndrome radiológicamente aislado (*Radiologically Isolated Syndrome, RIS*) y síndrome clínicamente aislado (*Clinically Isolated Syndrome, CIS*) (Lublin et al., 2014). Estos términos se utilizan para describir situaciones en las que los síntomas neurológicos o las lesiones en el SNC se presentan antes de que se cumplan los criterios para un diagnóstico definitivo de EM. Son etapas iniciales en las que se sigue la evolución del paciente para determinar si la enfermedad progresa y cumple con los criterios para un diagnóstico de EM. Es importante destacar que ni el CIS ni el RIS por sí solos cumplen con los criterios para un diagnóstico definitivo de EM, pero estos estados iniciales pueden ser indicativos de un mayor riesgo de desarrollar EM en el futuro. Puesto que, en la actual práctica clínica, los especialistas evalúan cuidadosamente a las personas con CIS o RIS para determinar si cumplen con los criterios para el diagnóstico de EM y si es necesario iniciar un tratamiento o un seguimiento más cercano, definimos a continuación estos dos cursos.

Capítulo 1. Introducción

1.1.2.1 Síndrome radiológicamente aislado (RIS)

Aunque el RIS no es considerado un fenotipo diferenciado de EM (Lublin et al., 2014), el incremento en su frecuencia de aparición en las imágenes de RM ha aumentado la concienciación en la comunidad de EM. En estos pacientes se encuentran anomalías en las imágenes de RM que parecen indicar desmielinización, aunque no presenten síntomas o signos clínicos.

Dado el alto riesgo que estos pacientes presentan de desarrollar en un futuro Esclerosis Múltiple Clínicamente Definitiva (*Clinically definite Multiple Sclerosis, CDMS*), resulta de gran importancia la identificación de factores con relevancia pronóstica para el estudio a largo plazo de futuros resultados. La importancia del RIS y la necesidad de una caracterización precisa de su evolución y riesgos asociados, está respaldada no solo por el hecho de que un porcentaje significativo de pacientes pasa a tener CDMS, sino también por la preocupación de que existan diagnósticos erróneos y la posible exposición a agentes terapéuticos.

1.1.2.2 Síndrome clínicamente aislado (CIS)

El término *CIS* describe un primer evento clínico señal de enfermedad desmielinizante en el SNC pero que aún no cumple con el requisito de diseminación en el tiempo para llegar al diagnóstico de *CDMS*. Diseminación en el tiempo indica la ocurrencia de múltiples episodios de brotes o síntomas neurológicos a lo largo del tiempo, criterio clave de diagnóstico junto a la diseminación en el espacio, la cual se refiere a la afectación de áreas diferentes del SNC, es decir, que los síntomas deben involucrar múltiples regiones anatómicas. Estos dos criterios, en conjunto, ayudan a los especialistas a diagnosticar y caracterizar adecuadamente la EM en un paciente. Los síntomas en *CIS* suelen ser monofocales, evolucionan de forma aguda o subaguda durante días o semanas y afectan al nervio óptico, la médula espinal, el tronco encefálico o el cerebelo. Al igual que otros ataques de EM, el episodio puede durar al menos 24 horas y ocurre en ausencia de fiebre o infección (D. H. Miller et al., 2012).

De manera similar al RIS, ciertas características demográficas y observadas en imagen de RM influyen en el riesgo de derivar a un diagnóstico de CDMS. La presencia y número de lesiones en sustancia blanca en T2 y un aspecto inusual del LCR, definido como un índice IgG alto (Inmunoglobulina G) o la presencia de OCBs (bandas oligoclonales), son los dos predictores más utilizados en la práctica clínica.

Capítulo 1. Introducción

La mayoría de los pacientes con CIS (50-70%) presentarán anomalías en las imágenes de RM cerebrales en T2 que se asemejarán a lesiones desmielinizantes, pero no tendrán síntomas asociados. Estas lesiones tienen un papel predictivo en relación con el paso a CDMS, el cual ha sido demostrado en varios estudios observacionales a largo plazo y muestra una ratio de conversión de hasta un 80% en un periodo de seguimiento de 20 años. El riesgo de pasar a CDMS además de por la presencia, también está relacionado significativamente con el número de lesiones (Brex et al., 2002; D. H. Miller et al., 2012; Tintore et al., 2010).

1.1.2.3 Esclerosis Múltiple Remitente Recurrente (EMRR)

La EMRR (Esclerosis Múltiple Remitente Recurrente), es el fenotipo más común en EM, presentándose en un 85% de los pacientes. Está caracterizada por la alternancia de periodos de recaídas en disfunciones neurológicas y periodos de relativa estabilidad clínica libre de nuevos síntomas neurológicos, como podemos ver en el gráfico de su evolución temporal en la Figura 2. La frecuencia de las recaídas puede variar entre pacientes, pero normalmente no sobrepasa 1.5 veces por año. Durante el periodo de recaída, pueden aparecer algunos síntomas neurológicos como debilidad, deterioro del equilibrio, deterioro de la agudeza visual y visión doble, durando al menos 24 horas sin presencia de infección o alteración metabólica. Las recaídas dan lugar a déficits residuales en casi la mitad de los episodios, lo que lleva a una acumulación gradual del deterioro (Lublin et al., 2003). La magnitud de la patología inflamatoria y la frecuencia de las recaídas, más notoria en adultos jóvenes, decrece conforme avanza la edad y la enfermedad (Boiko et al., 2002).

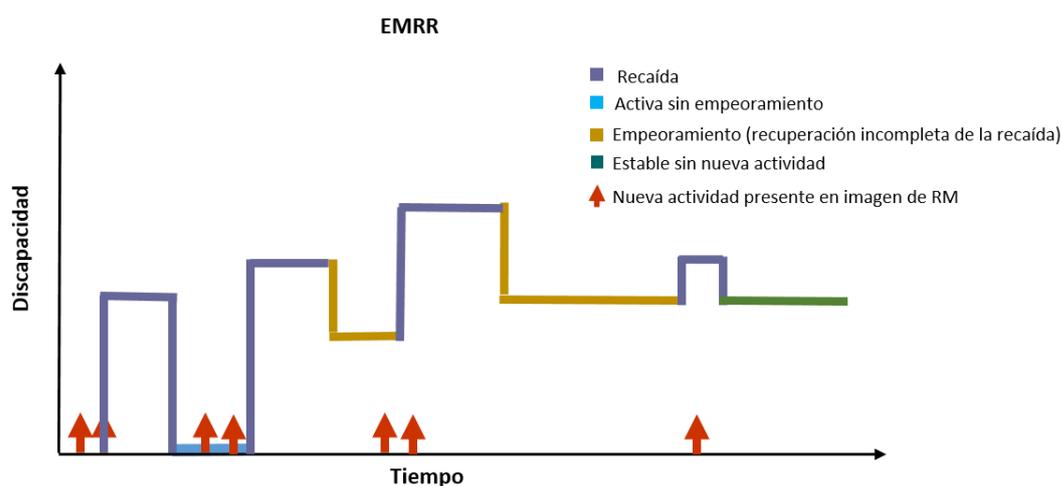


Figura 2. Trayectoria de la enfermedad de EMRR. Presencia de recaídas con leves remisiones entre ellas. Las recaídas pueden durar entre 24-48 horas.

Existen estudios que muestran una relación entre las infecciones y el aumento de la tasa de recaída, de su duración y de una mayor acumulación de discapacidad, pero todavía no ha sido

Capítulo 1. Introducción

identificado ningún patógeno en concreto (Vollmer, 2007). Por ejemplo, algunas infecciones como las respiratorias, las urinarias o las fúngicas, la gastroenteritis, situaciones de estrés o el embarazo y sus cambios bioquímicos asociados pueden afectar al desencadenamiento de recaídas, ya que estas coyunturas pueden estimular una respuesta inmunológica que a su vez afecte la inflamación en el SNC.

1.1.2.4 Esclerosis Múltiple Secundaria Progresiva (EMSP)

La mayoría de los pacientes con EMRR que no son tratados, con el tiempo pueden derivar a EMSP existiendo investigaciones que muestran que en un tiempo medio de 19 años se entra en fase progresiva desde que se inicia la EMRR (Rovaris et al., 2006). El diagnóstico de la EMSP es establecido frecuentemente de forma retrospectiva, años después de que la progresión haya empezado. A nivel individual, es difícil determinar con exactitud en qué momento del transcurso de la enfermedad el paciente hace la transición, pudiendo llegar a ser varios los años de incertidumbre en el diagnóstico. Las razones más comunes para la existencia de esta incertidumbre que suele ser de 2.9 ± 0.8 años de media según un estudio de población (Sand et al., 2014), es que los síntomas iniciales son muy sutiles y normalmente fluctuantes, por lo que indican signos de una progresión temprana pero también reticencia a ser progresiva, aumentando así la ansiedad de los pacientes en lo que se refiere a un pronóstico y la falta de terapias aprobadas.

Se han identificado unos cuantos predictores de la conversión a EMSP, por ejemplo, las personas que presentan la EMRR a una edad más avanzada suelen tener una progresión más rápida a EMSP, o también el hecho de pertenecer al género masculino, aunque esto último no está demostrado consistentemente en todos los estudios. La aparición de carga lesional en la médula espinal y una recuperación incompleta tras un periodo de recaída, son factores que también están relacionados con una progresión más rápida (Rovaris et al., 2006).

Fenotípicamente, el transcurso de la EMSP no es uniforme y está compuesto tanto por periodos de progresión con posibles episodios de recaída como por periodos de discapacidad relativamente estable (Figura 3).

Capítulo 1. Introducción

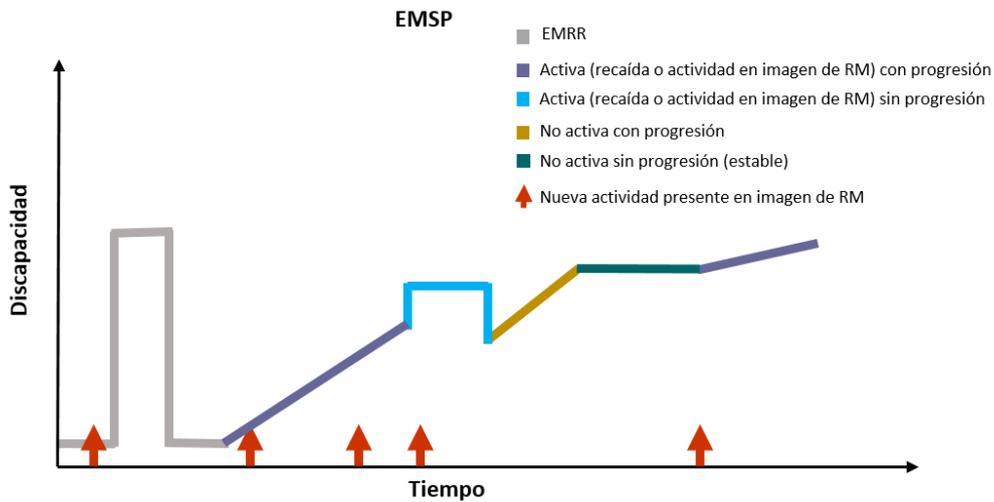


Figura 3. Trayectoria de la EMSP, Similar a la fase temprana-media de la EMRR, mostrando una pérdida más continua de las funciones físicas y cognitivas.

La patología implicada en la EMSP es compleja y poco conocida, ya que implica algo de inflamación persistente, aunque en menor medida que en la EMRR, además de neurodegeneración causada por disfunción mitocondrial y el resultante daño axonal.

Se cree que los cambios inflamatorios presentes en la enfermedad progresiva están compartimentados dentro del SNC tras una barrera hematoencefálica o blood-brain barrier (BBB) cerrada o reparada. Esta inflamación compartimentada es la fuerza impulsora detrás de la expansión de las lesiones existentes y los cambios difusos en la sustancia blanca de apariencia normal. Además, se pueden encontrar áreas inflamadas focales en las meninges de pacientes progresivos.

Capítulo 1. Introducción

1.1.2.5 Esclerosis múltiple primaria progresiva (EMPP)

Alrededor del 10%-20% de los pacientes desarrollarán el fenotipo EMPP caracterizado por la falta de la fase remitente recurrente inicial y la progresión continua desde el inicio de la enfermedad (Compston & Coles, 2008). A nivel individual, la progresión en el paciente no es uniforme y se pueden dar recaídas al igual que periodos de relativa estabilidad de la enfermedad (Figura 4).

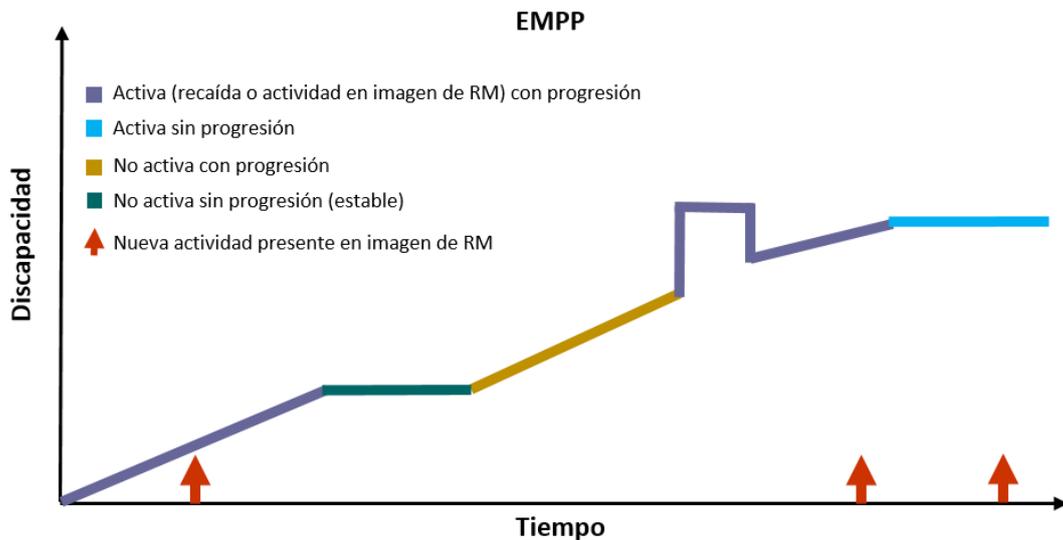


Figura 4. Transcurso de EMPP. No muestra recaídas, pero sí una pérdida gradual de las funciones físicas y cognitivas con los años. Puede haber presencia de leves periodos de estabilidad o mejoras, pero poco frecuente.

El aumento de información en forma de datos clínicos, genéticos y de imágenes sugiere que la EMPP es una parte del espectro de la enfermedad de la EM y que cualquier diferencia patológica con la EMSP es relativa en lugar de absoluta. Se ha demostrado, estudiando la evolución natural de la enfermedad, que la discapacidad progresa en paralelo en pacientes tanto con EMPP como con EMSP, sufran o no recaídas (Kremenutzky et al., 2006). El hecho de que ~10% de los pacientes con RIS desarrollen EMPP respalda aún más la teoría de que la ausencia de la fase RR en pacientes EMPP está causada por lesiones en el SNC clínicamente silenciosas (Okuda et al., 2014; Ransohoff et al., 2015).

Al igual que en EMSP, la patología en EMPP es compleja e incluye neurodegeneración que tiene lugar junto a una inflamación de leve a moderada.

1.1.3 Modificadores de actividad y progresión en EM

La incorporación de dos nuevos modificadores de actividad y progresión, a los fenotipos ya antes establecidos, fue propuesta en la revisión de 2013 basándose en el avance del conocimiento de la progresión clínica de la EM y en el papel cada vez más importante de las imágenes de RM en la investigación y atención clínica. Mientras que los fenotipos definidos en 1996 nos aportaban

Capítulo 1. Introducción

una descripción de la enfermedad más estática, la incorporación de los descriptores de actividad y progresión mejora la caracterización del transcurso de la enfermedad y otros aspectos como el pronóstico, los tratamientos y los resultados tanto en atención clínica como en investigación (Kang et al., 2014).

Que haya actividad en la enfermedad se define por la presencia de recaída clínica o una nueva observación anómala en RM, tanto en pacientes con recaídas como progresivos. El comité sugirió una evaluación anual de la actividad de la enfermedad a través de un examen clínico tanto en pacientes con EMRR como progresivos. La evaluación anual de la actividad en RM en pacientes con recaídas es claramente beneficiosa, pero no se llegó a un consenso sobre la frecuencia de evaluación en pacientes progresivos. Debido a la correlación entre la actividad cerebral y espinal en las RM, no se recomienda la realización de exploraciones en la médula espinal mediante RM a menos que haya nuevos síntomas vertebrales (J. C. Bot & Barkhof, 2009).

La inclusión de esta nueva terminología descriptiva da lugar a varios fenotipos de EM, como EMRR activa, ya sea en pacientes de EMRR tanto con actividad clínica como radiológica o EMPP-no activa en pacientes de EMPP que no presentan crisis agudas o actividad radiológica dentro de un determinado periodo de tiempo.

El descriptor de progresión de la enfermedad aplica tanto a pacientes con EMSP como con EMPP. Como la progresión en EM no es uniforme y los pacientes pueden permanecer relativamente estables en el tiempo, se recomienda una evaluación anual. Debido a la falta de biomarcadores de imagen o inmunológicos, la progresión de la enfermedad se determina sólo de manera clínica, combinando hallazgos objetivos y el historial del paciente. De nuevo, combinando características de actividad y progresión, se pueden observar varios fenotipos de EM, como EMSP activa y progresiva en pacientes con EMSP con actividad clínica o de imagen y progresión de la discapacidad (Lublin et al., 2014).

La incorporación de los modificadores fenotípicos de actividad y progresión es sólo un primer paso hacia una terminología más específica y por lo tanto hacia un cuidado del paciente más individualizado.

Capítulo 1. Introducción

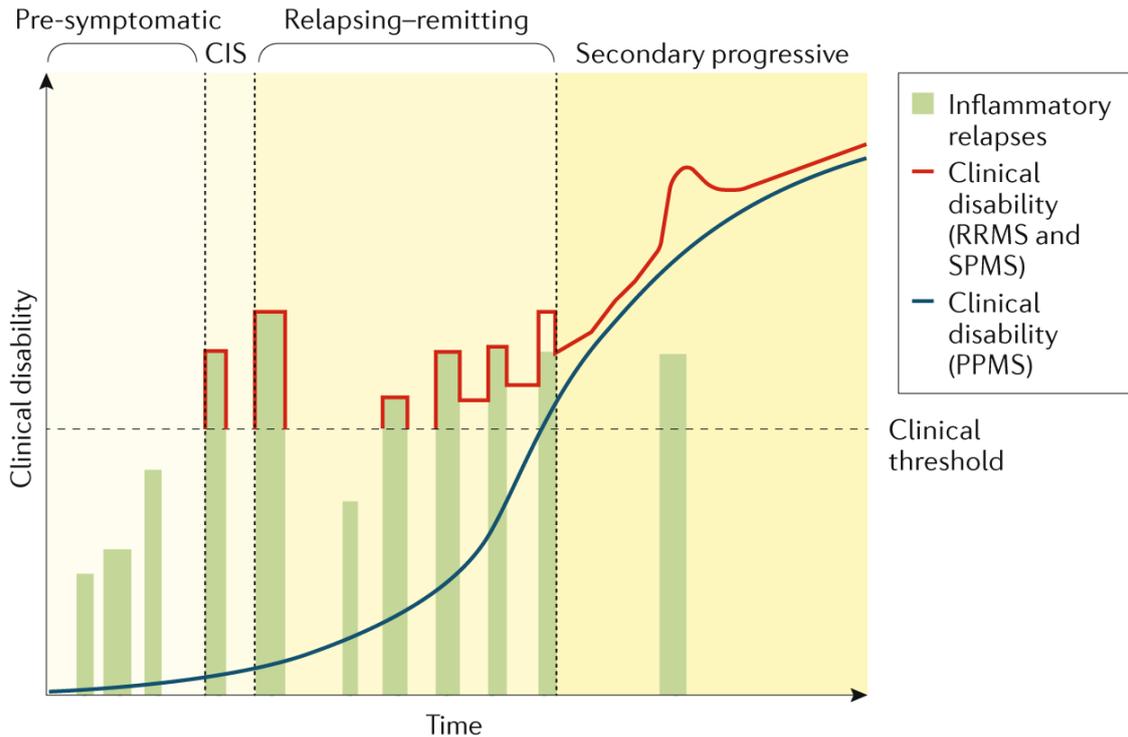


Figura 5. Curso clínico de la EM (Filippi et al., 2018). Si no se trata, la EMRR puede evolucionar a EMSP, donde los niveles de discapacidad aumentan continuamente sin períodos de recaída ni remisiones.

1.1.4 Escala de Discapacidad Extendida de Kurtzke (EDSS)

La Escala de Discapacidad Extendida de Kurtzke (EDSS) (Hohol et al., 1999) es la base de un método para cuantificar la discapacidad en la EM y monitorizar los cambios a nivel de discapacidad en el tiempo, tanto en ensayos clínicos como en la rutina de evaluación de los pacientes que padecen la enfermedad, de manera que complementa a las demás consideraciones clínicas y pruebas para brindar una atención integral a los pacientes con esta enfermedad.

La escala fue desarrollada por el neurólogo John F. Kurtzke en 1983 como una evolución de su escala previa (DSS, Disability Status Scale) la cual contenía 10 niveles distintos. La EDSS contiene puntuaciones entre 0 y 10 en incrementos de 0.5, los cuales van indicando cada vez mayores niveles de discapacidad (ver Tabla 1). La puntuación asignada por neurólogos al paciente está basada en medidas de deterioro en ocho sistemas funcionales (SF): piramidal (debilidad muscular o dificultad para mover las extremidades), cerebeloso (ataxia, pérdida de equilibrio, coordinación o temblor), tronco encefálico (problemas con el habla, la deglución, nistagmo), sensorial (entumecimiento o pérdida de sensaciones), función intestinal y vesical, función visual y funciones cerebrales (problemas con el pensamiento y la memoria). Estos sistemas funcionales representan una red de neuronas en el cerebro responsables de unas tareas en particular. Los rangos de 1 a 4.5 se refieren a personas con EM que son capaces de andar sin ningún tipo de ayuda mientras que los rangos de 5.0 a 9.5 se caracterizan por la discapacidad para andar. Esta

Capítulo 1. Introducción

dependencia tan fuerte de la escala, con la suficiencia para caminar como principal medida de discapacidad, suele ser un motivo de crítica.

Tabla 1. Resumen descriptivo de las distintas puntuaciones en la EDSS. Punt.: puntuación. ADC: ausencia de discapacidad para caminar. Total. dep: totalmente dependiente. Contenido adaptado de (Kurtzke, 1983)

Punt.	Descripción
0	Examen neurológico normal, sin discapacidad en ningún SF
1.0	Sin discapacidad, signos mínimos en un SF
1.5	Sin discapacidad, signos mínimos en más de un SF.
2.0	Discapacidad mínima en un SF.
2.5	Discapacidad leve en un SF o mínima en dos SF.
3.0	Discapacidad moderada en un SF, o discapacidad leve en tres o cuatro SFs. ADC.
3.5	Discapacidad moderada en un SF y discapacidad por encima de la mínima en varios. ADC.
4.0	Discapacidad significativa, pero autosuficiencia en torno a las 12 horas diarias. Capacidad de caminar 500 m.
4.5	Discapacidad significativa, pero independencia durante la mayor parte del día. Capacidad de trabajar una jornada completa, con algunas limitaciones para ciertas actividades o necesidad de una ayuda mínima. Capacidad de caminar 300 m.
5.0	Discapacidad suficientemente severa como para impedir las actividades diarias y el trabajo de una jornada completa sin modificaciones. Capacidad de caminar 200 m.
5.5	Discapacidad limitante de actividades diarias. Capacidad de caminar 100 m.
6.0	Necesidad de algún tipo de apoyo para caminar 100 m, con o sin descanso.
6.5	Necesidad de dos tipos de apoyos para caminar unos 20 m sin descansar.
7.0	Incapacidad de caminar más de unos 5 m, incluso con ayuda. Esencialmente limitado a moverse en silla de ruedas, aunque hay capacidad de moverla uno mismo y se transfiere sin ayuda. Activo en la silla de ruedas durante 12 horas al día.
7.5	Incapacidad de dar más de unos pocos pasos. Limitación a la silla de ruedas y con apoyo para transferencia. Capacidad para mover la silla, pero no todo el día si la silla es convencional y carece de motor.
8.0	Esencialmente confinado en cama o silla de ruedas movida por otra persona. Capacidad de pasar gran parte del día fuera de la cama, con mantenimiento de muchas funciones de autocuidado. El uso efectivo de los brazos es habitual.
8.5	Confinado en la cama la mayor parte del día. Capacidad de emplear los brazos y mantenimiento de algunas funciones de autocuidado.
9.0	Confinado en cama y total. dep. Capacidad para comunicarse y comer.
9.5	Confinado en cama y total. dep. Incapacidad para comunicarse de forma efectiva y de tragar.
10	Muerte por EM.

Esta escala tiene varias implicaciones en el seguimiento de la enfermedad y todos los pacientes deberían conocer o tener en su informe médico la valoración EDSS, ya que permite tener una medida cuantitativa de la discapacidad y su evolución a lo largo del tiempo, fundamental para tomar decisiones sobre el tratamiento y la atención a medida que la enfermedad evoluciona (Hohol et al., 1999). De igual modo, la EDSS se utiliza para evaluar la eficacia de los tratamientos para la EM de manera que los profesionales pueden comparar las puntuaciones de EDSS antes y después para determinar si está teniendo un impacto positivo en la progresión de la enfermedad y la discapacidad, además, resulta útil para estudiar la necesidad de rehabilitación, el manejo de síntomas y la planificación a largo plazo para el cuidado del paciente. La EDSS

Capítulo 1. Introducción

también facilita la comprensión, proporcionando una forma estandarizada de comunicar la discapacidad y la progresión de la enfermedad entre profesionales de la salud y pacientes.

A pesar de que su uso está muy extendido y su valoración repercute en la rutina clínica, no se tienen en cuenta otros aspectos provocados también por la EM y que influyen en la calidad de vida de los afectados, como: las secuelas cognitivas, el dolor, la depresión, la falta de energía, la fatiga o su repercusión social (Ziemssen, 2009). Además, las puntuaciones de la EDSS pueden variar por la complejidad de las reglas de puntuación y la subjetividad del examen neurológico.

En relación con la atrofia medular, al estar asociada con la pérdida de células nerviosas y tejido en el SNC, ocurre que, en la mayoría de los casos, conforme la enfermedad progresa, esta atrofia contribuye significativamente a la discapacidad medida con la EDSS. Encontrar una relación entre la EDSS y la atrofia en EM es complejo y multifactorial, ya que, aunque es frecuente que a medida que la atrofia progresa se refleje en un aumento en la discapacidad, la EDSS puede no capturar completamente la complejidad de la atrofia y otros cambios estructurales en el SNC. Los avances en técnicas de imagenología, como la RM, permiten una evaluación más detallada de la atrofia de la médula espinal en EM, de manera que las mediciones de médula espinal en estudios de RM pueden proporcionar información valiosa sobre la progresión de la enfermedad y la relación con la discapacidad, incluso cuando las medidas clínicas, como la EDSS, pueden no ser suficientemente sensibles para capturar cambios sutiles.

1.2 Imagen médica

La imagen médica es la base metodológica de la presente tesis doctoral y se centra en la utilización de diversas tecnologías para visualizar el cuerpo humano o partes de este con el propósito de diagnosticar, monitorizar y examinar condiciones médicas. Estas imágenes proporcionan información valiosa a los profesionales de la salud sobre la anatomía, función y patologías de los tejidos y órganos internos. De manera más específica, nos referimos a la imagen médica como una imagen en forma de matriz 2D o 3D cuyos elementos contienen niveles de intensidad medidos por un instrumento o equipo de captación cuyos resultados son imágenes. A un elemento de una matriz 2D lo conocemos como píxel, mientras que a un elemento de una matriz de una imagen volumétrica lo llamamos voxel (VOLUME pixEL).

El tipo de técnica de imagen, es decir, el tipo de escáner o sonda, utilizado para adquirir una imagen médica se conoce como modalidad.

Capítulo 1. Introducción

1.2.1 Modalidades

Las modalidades de imagen médica son diferentes técnicas y tecnologías utilizadas para capturar imágenes del cuerpo humano desde distintos ángulos y perspectivas. La medicina moderna se beneficia enormemente de una variedad de tecnologías de imagen médica que permiten a los profesionales de la salud explorar el cuerpo humano de manera detallada y precisa. Desde los pioneros rayos X, que revelaron el mundo interior de los tejidos óseos, hasta los ultrasonidos, que utilizan ondas sonoras para crear imágenes en tiempo real, y la medicina nuclear, que rastrea la distribución de sustancias radiactivas en el cuerpo, cada modalidad ofrece una perspectiva única. Las radiografías han sido durante mucho tiempo la piedra angular de la imagen médica, proporcionando una visión bidimensional de las estructuras óseas y permitiendo diagnósticos rápidos y fundamentales. Las tomografías computarizadas (TC), por otro lado, han llevado la visualización tridimensional a nuevas alturas, revelando con precisión las complejidades de los tejidos blandos y los órganos internos. Los ultrasonidos, utilizando ondas sonoras de alta frecuencia, brindan la ventaja de la visualización en tiempo real, siendo una opción valiosa para la monitorización en el ámbito obstétrico y la evaluación de tejidos superficiales. Por su parte, la medicina nuclear permite estudiar la función fisiológica a través de la detección de sustancias radiactivas, desempeñando un papel crucial en la detección temprana de enfermedades y el seguimiento de procesos metabólicos.

En este amplio espectro de tecnologías de imagen médica, la RM emerge como una herramienta excepcionalmente poderosa y versátil. A diferencia de muchas modalidades, la RM no utiliza radiación ionizante, lo que la convierte en una opción segura y valiosa para una variedad de pacientes. En los siguientes párrafos, nos sumergiremos en los principios fundamentales y las aplicaciones de la RM en EM, explorando cómo esta tecnología ha transformado nuestra capacidad para visualizar y comprender la anatomía de los pacientes.

1.2.1.1 *Imagen de Resonancia Magnética*

1.2.1.1.1 Ciencia en la adquisición

Una imagen de RM es una imagen 3D producida por un campo magnético. Se generan por el intercambio de energía de radio frecuencia (RF) entre el cuerpo de un paciente y el instrumento de imagen médica, lo que es posible gracias a las propiedades magnéticas intrínsecas del cuerpo humano; específicamente, de los átomos de hidrógeno (protones) que son los más abundantes en nuestros tejidos. Los campos magnéticos se utilizan para alinear la magnetización nuclear de estos átomos de hidrógeno ubicados en las moléculas del agua que constituye el cuerpo del paciente y los campos de RF para alterar sistemáticamente el alineamiento de esa

Capítulo 1. Introducción

magnetización, causando que los núcleos de hidrógeno produzcan un campo magnético rotacional detectable por el escáner.

Debido a que los tejidos de nuestro cuerpo contienen distintas cantidades de agua, las imágenes de RM detectan el campo electromagnético del núcleo (protones) para distinguir las diferencias en intensidad y la forma de estos. Basándonos en la rapidez con la que los protones liberan su exceso de energía una vez que el pulso de RF se ha apagado, podremos diferenciar entre los diferentes tejidos del cuerpo. La velocidad (o tiempo) de relajación es el factor más importante a la hora de producir el contraste en la imagen entre los distintos tipos de tejido. Existen dos tipos de tiempos de relajación en imágenes de RM, T1 (tiempo de relajación longitudinal) y T2 (tiempo de relajación transversal). Las imágenes T1 mapean la densidad de los protones dentro de los tejidos grasos como la médula ósea de los cuerpos vertebrales. Como el LCR no contiene grasa, aparece oscuro en las imágenes ponderadas T1. En cambio, las imágenes T2 mapean la densidad de protones de los tejidos grasos y acuosos. Por ello, en las imágenes ponderadas T2 el LCR aparece brillante. Podemos encontrar una explicación más detallada de la ciencia en la adquisición mediante la técnica de RM en el anexo I.

1.2.1.1.2 Secuencias

Son varios los tipos de imágenes de RM que se recomiendan para la detección de la EM en la médula espinal, entre los cuales, hay varios niveles de eficacia diagnóstica debido a las diferencias de contraste. Es esencial utilizar secuencias de imagen específicas que proporcionen una buena visualización y diferenciación de las estructuras anatómicas y patológicas.

- Las secuencias ponderadas en T1 son útiles para obtener una buena visualización anatómica de la médula espinal y los tejidos circundantes. Estas imágenes proporcionan un contraste adecuado entre los diferentes tejidos y permiten evaluar la morfología general de la médula espinal.
- Las secuencias ponderadas en T2 son buenas para detectar cambios en la señal de los tejidos blandos y líquidos, lo que facilita la identificación de edema, inflamación o lesiones traumáticas en la médula espinal.
- La secuencia STIR (*Short Tau Inversion Recovery*), la cual suprime la señal de la grasa y realza la señal de las lesiones de médula espinal.
- La secuencia ponderada en difusión mide el movimiento de las moléculas de agua en los tejidos lo que nos podría ser útil para ayudar a evaluar la integridad de las fibras nerviosas y detectar áreas de daño axonal o isquemia.

Capítulo 1. Introducción

- Las secuencias de realce de contraste, en la que utilizando un agente de contraste como el Gadolinio para mejorar la visualización de ciertas estructuras o lesiones se puede llegar a detectar tumores, malformaciones vasculares y lesiones inflamatorias.

En el desarrollo de esta tesis decidimos emplear una secuencia T1 en su plano transversal para desarrollar el algoritmo de segmentación automática de médula espinal ya que, debido al alto contraste que proporciona, permite una buena diferenciación de las estructuras anatómicas y además presenta una alta resolución espacial, mejorando la visualización de estructuras pequeñas como es la médula espinal.

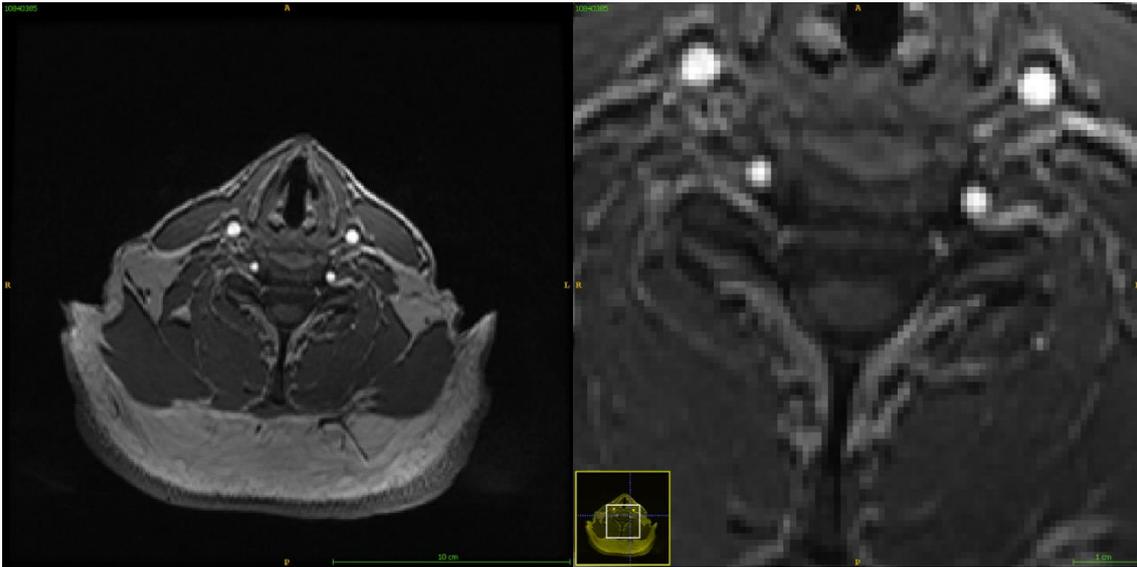


Figura 6. Izquierda: Imagen componente de secuencia transversal 3D potenciada en T1. Derecha: ampliación de la región perteneciente a la médula.

Capítulo 1. Introducción

1.2.1.1.2.1 Secuencia axial 3D ponderada T1

En la secuencia axial T1, las imágenes se adquieren en un plano transversal del cuerpo del paciente, esto se refiere a un plano anatómico en el cual las imágenes se adquieren de manera perpendicular al eje longitudinal del cuerpo. Por lo tanto, en una secuencia axial, las imágenes se toman recorriendo el cuerpo en una dirección horizontal, recorriendo al paciente desde los pies hasta la cabeza.

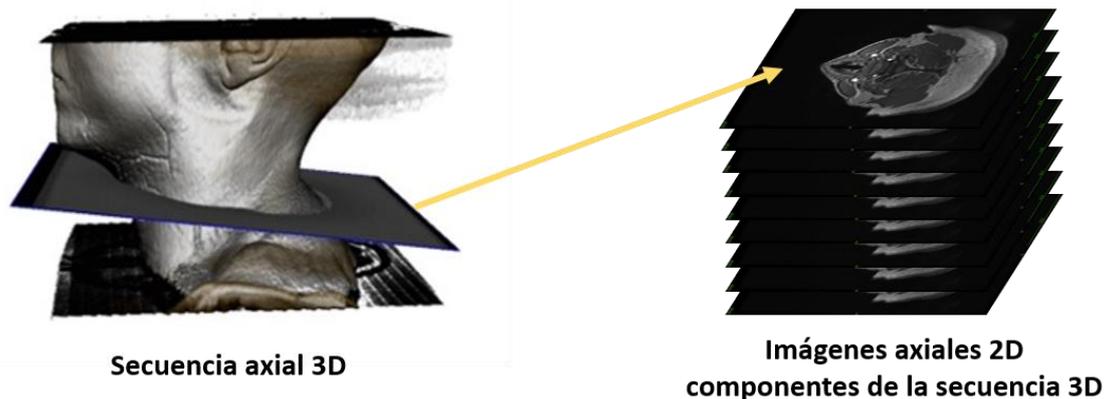


Figura 7. Ejemplo ilustrativo de imágenes axiales 2D componentes de la secuencia axial 3D.

La potenciación T1 implica resaltar las características de los tejidos según su tiempo de relajación longitudinal. Esto se logra utilizando pulsos de radiofrecuencia específicos en una secuencia de RM. Durante la potenciación T1, los pulsos de RF se diseñan para invertir la dirección de los momentos magnéticos de los protones en los tejidos, de manera que los protones que originalmente estaban en estado de equilibrio en la dirección del campo magnético ahora se invierten. Los protones excitados comienzan a regresar a su estado de equilibrio en la dirección del campo magnético durante el proceso conocido como relajación longitudinal o recuperación T1. Durante esta fase, se genera la señal de resonancia que se capta a medida que los protones vuelven a su estado de equilibrio en función del T1 de los diferentes tejidos. Esto resalta los tejidos con tiempos de T1 más cortos en las imágenes resultantes. Por ejemplo, los tejidos con contenido graso y muscular tienden a tener tiempos de T1 más cortos, por lo que aparecerán hiperintensos en una imagen potenciada por T1. En general, los líquidos como el LCR aparecerán hipointensos debido a su largo T1. La imagen final se procesa mediante técnicas de reconstrucción que asignan valores de intensidad a cada píxel de la imagen, reflejando las diferencias en la recuperación de los tejidos en función de sus tiempos de T1.

Los contrastes suelen ser administrados para mejorar las imágenes de RM de los órganos y tejidos por cambio en los tiempos de relajación de los protones del agua. Estas sustancias tienen la capacidad de acumularse en determinados tejidos, modificando temporalmente sus

Capítulo 1. Introducción

propiedades electromagnéticas y, por tanto, destacándose en las imágenes obtenidas en la RM. Normalmente acortan la velocidad de todos los procesos de relajación; sin embargo, cada sustancia influye de una manera especial. En nuestro caso, para la adquisición de las secuencias axiales 3D, se utiliza el tipo de agente de contraste T1 o agente de contraste positivo, en concreto el Gadolinio, acortando los tiempos de relajación en la componente longitudinal. Estos contrastes son administrados por vía intravenosa y excretados por vía renal, con una vida media de 1 a 2 horas desapareciendo de la sangre a las 24 horas. En el interior del cuerpo humano, con el contraste distribuido tanto en los vasos como en los tejidos, se produce un acortamiento de los tiempos de relajación del momento magnético, lo que mejora muy significativamente la señal en la relajación T1.

Estas secuencias axiales 3D potenciadas T1 están formadas a su vez por imágenes 2D, tal como se muestra en la Figura 7 y son fundamentales para visualizar diversos aspectos de los tejidos y estructuras internas del cuerpo, en este caso específico, la médula espinal cervical.

Elegir una secuencia axial potenciada T1 para nuestro método, tiene otras virtudes además del alto contraste entre los diferentes tipos de tejidos del cuerpo, lo que facilita la diferenciación entre estructuras anatómicas y es particularmente útil para identificar bordes y límites claros entre tejidos blandos. Estas secuencias también ofrecen una alta resolución espacial, lo que es crucial para una segmentación precisa, ya que cuanto mejor sea la resolución, más detalles se podrán capturar en las imágenes, lo que resulta importante a la hora de adquirir imágenes de estructuras como la médula espinal, la cual es una estructura pequeña y móvil, afectada por artefactos de movimiento ocasionados por la respiración, el movimiento cardíaco, la pulsación del LCR y el flujo sanguíneo. Además, la presencia de artefactos en este tipo de secuencia suele ser muy pequeña debido a su alta relación señal-ruido, tal como explicamos en la sección 1.2.1.1.3.

Por otra parte, cumple la característica de datos procedentes del mundo real (*Real-world data*, RWD), al ser una de las secuencias más comunes en RM, siendo ampliamente utilizada en entornos clínicos y de investigación por lo que la mayoría de los equipos de RM están configurados para adquirirlas y los profesionales de la salud suelen estar familiarizados con la interpretación de estas imágenes. Muchas enfermedades y condiciones médicas pueden ser identificadas y evaluadas mediante imágenes T1, lo que hace que esta secuencia sea fundamental para el proceso de segmentación y análisis de estructuras anatómicas y debido a su tiempo de adquisición relativamente corto con respecto a otras secuencias de RM, lo que es beneficioso para pacientes que pueden tener dificultades para permanecer inmóviles durante

Capítulo 1. Introducción

períodos prolongados, las secuencias T1 son esenciales en el diagnóstico clínico y en la planificación de tratamientos.

1.2.1.1.3 Artefactos en Imágenes RM

Aunque de manera infrecuente, las imágenes de RM pueden presentar artefactos debidos a diversas razones y pueden afectar de manera negativa tanto a la interpretación correcta por parte de los clínicos como el rendimiento de los modelos de IA. Los tipos de artefacto más frecuentes son: movimiento del paciente, de solapamiento, de susceptibilidad magnética, de truncamiento, de desplazamiento químico, de repliegue, de ángulo mágico...).

Los más comunes encontrados en el análisis de EM son tanto el efecto del volumen parcial como el 'bias field noise'. Vamos a explicar cada uno de ellos.

En una imagen de RM, cada píxel (o vóxel en 3D) representa la señal devuelta por un tipo de tejido en respuesta al campo electromagnético. Si un píxel representa la información de un solo tejido, tendrá un nivel de intensidad único característico, pero cuando hay varios tejidos próximos, es posible que la señal devuelta sea la combinación de las señales devueltas por varios tejidos diferentes adyacentes. Esto se traduce en una zona difusa con bordes no nítidos en la que un vóxel (o más) representa la señal combinada devuelta por varios tejidos, conocido como efecto de volumen parcial, y dificulta el poder diferenciar donde está el límite entre ellos, lo que provoca que sea difícil determinar dónde empieza una estructura y termina otra. De darse este efecto, podría afectar en nuestro objetivo de identificación de medula espinal cervical, ya que el hecho de que unos píxeles sean considerados pertenecientes a la médula espinal o no, dependerá de que los distintos tejidos estén bien diferenciados. En la Figura 8 podemos ver un ejemplo.

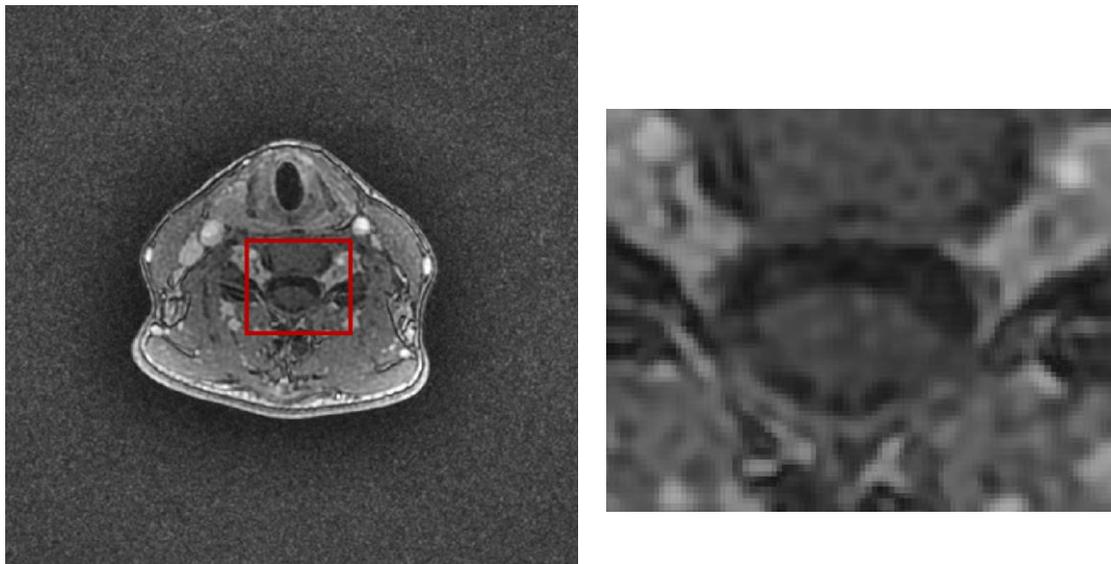


Figura 8. Efecto del volumen parcial en una región de la imagen (izquierda) y misma región aumentada (derecha).

Capítulo 1. Introducción

El “bias field noise” se produce por heterogeneidades en el campo electromagnético generado que se reflejan como zonas de iluminación no uniforme en la imagen, tal como se muestra en la Figura 9. Este efecto también puede llegar a dificultar la segmentación, siendo en algunos casos necesaria la aplicación de técnicas de pre-procesamiento para uniformizar la imagen.

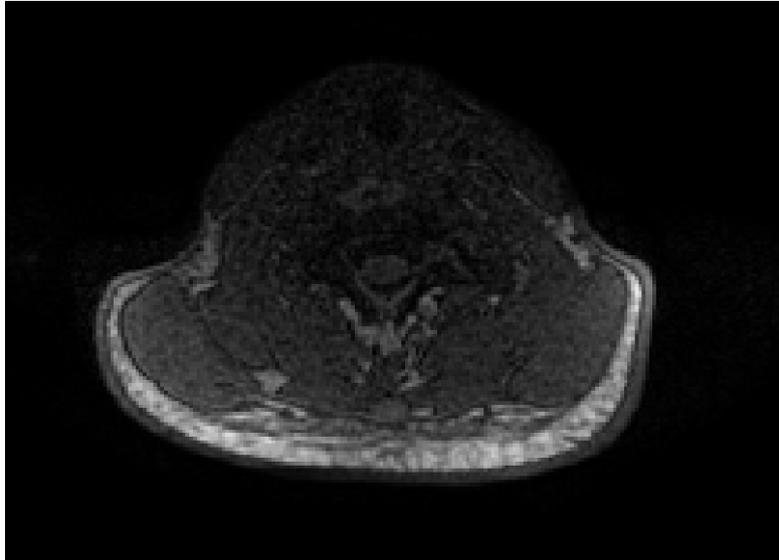


Figura 9. Artefacto de ruido bias-field.

En nuestro estudio, empleamos imágenes de RM potenciadas T1 las cuales tienen la característica de presentar una buena relación señal-ruido (SNR), lo que significa que la señal (la información deseada) es relativamente fuerte en comparación con el ruido (señales no deseadas). Esto mejora la calidad de las imágenes y facilita la identificación de estructuras anatómicas, disminuyendo la presencia de estos artefactos que acabamos de presentar.

1.2.2 El formato DICOM

Las técnicas de imagen médica introducidas en el apartado 1.2.1, generan imágenes que pueden interpretarse mediante distintos sistemas de visualización. Las características que las difieren y caracterizan (distintos formatos, escalas de colores, fabricantes de equipos, volumetría...) hacen necesaria la interoperabilidad, requerimiento fundamental para la colaboración entre profesionales de la salud y para la transferencia segura entre diferentes instalaciones médicas. Este requerimiento junto a otros que vemos a continuación, nos lo proporciona el formato DICOM.

Todas las modalidades médicas de imagen actuales como rayos-X, ultrasonidos, TC y RM se apoyan en el formato DICOM. DICOM es un protocolo de almacenamiento y de comunicación. Su éxito se basa en la habilidad de integrar modalidades médicas fabricadas por diferentes

Capítulo 1. Introducción

proveedores, lo que era casi imposible antes de su llegada. La integración de equipos médicos de diferentes proveedores solía ser un gran problema, esa es la razón por la que DICOM se ha convertido en el estándar de la industria.

DICOM no sólo almacena la imagen médica, sino que también contiene meta-datos que contienen atributos y propiedades de ésta, facilitando la identificación y el acceso a variables dentro de un mismo archivo. La cabecera consta de 128 bytes a modo de preámbulo y 4 bytes con el prefijo "DICM". El cuerpo del archivo se forma por una secuencia de Data Sets, que representan objetos del mundo real y que a su vez están constituidos por *Data Elements*, que son atributos codificados del objeto. Cada *Data Element* es identificado y clasificado por un *TAG* o etiqueta, es decir, un identificador único. Existe una gran variedad de *Data Elements* y se debe tener en cuenta que no siempre estarán definidos en su totalidad dentro de un mismo archivo, la presencia de estos depende del tipo de estudio. Los metadatos contienen información importante que debe mantenerse dentro del archivo para garantizar que no se separen unos de otros.

1.2.3 Conversión de DICOM a NifTi

Pese a que los datos de imagen médica se almacenan y transfieren en la práctica clínica usual en formato DICOM, su interoperabilidad puede ser limitada fuera de estos entornos, por lo que es recomendable para el procesamiento, análisis y compartición de datos en el campo de la neuroimagen y la investigación médica, utilizar formatos de imagen más simples que optimizan este tipo de tareas. Estos formatos tienen una estructura más sencilla y estandarizada que mejoran la interoperabilidad y retienen sólo un limitado conjunto de metadatos relevantes de la imagen. En el desarrollo de nuestro método, y con este propósito, incluimos un paso inicial de conversión de imágenes del formato DICOM al formato NifTi (*Neuroimaging Informatics Technology Initiative*).

1.2.3.1 Formato NifTI (Neuroimaging Informatics Technology Initiative)

Con este fin de superar algunas limitaciones y complejidades del formato DICOM fue desarrollado el formato NifTI, un estándar en neuroimagen que se utiliza para el almacenamiento y análisis de datos de imágenes en el campo de la investigación neurológica y médica, el cual es el utilizado durante el desarrollo de esta tesis.

La estructura de un archivo NifTI consta de dos partes principales: el encabezado (*header*) y los datos de la imagen. El encabezado contiene información sobre la dimensión, resolución, orientación, metadatos y otros detalles relevantes de la imagen, mientras que los datos de la imagen son un arreglo tridimensional que representa la intensidad de los píxeles o voxels. Uno

Capítulo 1. Introducción

de los beneficios clave de NIfTI es la inclusión de información de orientación espacial que permite una interpretación y registro consistentes de las imágenes. Esto asegura que las imágenes estén correctamente alineadas y sean comparables entre sí. Además, permite definir un punto de origen en las coordenadas de la imagen, lo que es útil para establecer un punto de referencia en el espacio que puede ser importante para análisis específicos.

Este formato es ampliamente adoptado en la investigación de imagen médica, permitiendo a los investigadores combinar numerosas bibliotecas y herramientas de software diseñadas para leer, escribir y manipular archivos NIfTI como programas para visualización, procesamiento y análisis de datos de imagen médica. El formato NIfTI se ha convertido en un estándar de facto en el campo de la neuroimagen debido a su simplicidad, flexibilidad y facilidad de uso.

En este proyecto, fue el formato de imagen que elegimos utilizar, por lo que las imágenes que obtenemos del Sistema de Comunicación y Archivo de Imágenes (*PACS, Picture Archiving and Communication System*) en formato DICOM fueron convertidas a dicho formato, NIfTI. El módulo de Python *dcm2niix*, nos permite convertir a la extensión “.nii” (NIfTI) una carpeta DICOM que puede contener varios archivos pertenecientes a una sola secuencia o multi-secuencia. Esta fue la herramienta utilizada para convertir nuestros casos de DICOM a NIfTI, incluida en el visor MRICroGL (*NITRC: MRICroGL: Tool/Resource Info*, n.d.). Existen otras alternativas para dicha tarea, como el módulo *dcm2nii* de la librería Nipype de Python (*Neuroimaging in Python - Pipelines and Interfaces — Nipy Pipeline and Interfaces Package*, n.d.) o la librería *dicom2nifti* (*Dicom2nifti — Dicom2nifti Documentation*, n.d.).

1.2.4 Imagen de resonancia magnética en Esclerosis Múltiple

La adquisición de imágenes de RM en la región cerebral y la médula espinal nos aporta mucha sensibilidad a la hora de detectar anomalías en la EM. El criterio diagnóstico actual para EM incluye una serie de características específicas para imágenes de RM que podemos observar en la Tabla 2 con el fin de proporcionarnos evidencias de diseminación en espacio y tiempo, permitiéndonos así un diagnóstico más temprano de la enfermedad tras la presencia en el paciente de un síndrome clínico aislado con desmielinización (CIS).

Tabla 2: Criterio diagnóstico de McDonald revisado en 2010 para EM (Traboulsee et al., 2016).

Características mínimas en imágenes de RM para considerar diseminación en el espacio

1 Lesión infratentorial

1 Lesión yuxtacortical (cerca de la corteza)

1 Lesión periventricular (cerca de los ventrículos)

1 Lesión en médula espinal

Capítulo 1. Introducción

Estas revisiones simplifican el criterio de McDonald de 2005 a la vez que preservan la sensibilidad y especificidad de diagnóstico, solucionan la aplicabilidad entre poblaciones y mantiene una buena sensibilidad y especificidad si lo comparamos con el que existía previamente (Poser et al., 1983). Aun así, hay que tener en cuenta que las lesiones en sustancia blanca suelen ser comunes conforme aumenta la edad, por lo que el criterio de aplicación de RM debería ser utilizado con cuidado en aquellos pacientes que muestren síntomas atípicos de EM o cuando la aparición de síntomas se da en pacientes mayores de 40 años. Esta recomendación es particularmente importante cuando hay determinados factores que se conoce que han causado hiperintensidades en T2, como la hipertensión, el tabaquismo, la diabetes, el colesterol alto y las migrañas.

La RM se usa cada vez más para el seguimiento de pacientes con EM definitiva, con el fin de seguir la progresión y las alteraciones silenciosas clínicamente y monitorizar la respuesta a terapia. El contraste de Gadolinio resalta las lesiones o cambios en T1, las cuales se podrían considerar como biomarcadores de imagen que señalan a una nueva inflamación. Es más frecuente observar actividad nueva en las imágenes de resonancia que en síntomas clínicos como las recaídas (J. O. Harris et al., 1991).

A continuación, vamos a realizar una revisión de los protocolos de imagen de RM y las guías clínicas en cuanto a los avances de la tecnología de imagen y las nuevas evidencias clínicas para diagnosticar y monitorizar la EM con imágenes de RM.

1.2.4.1 Protocolos

Los protocolos fueron actualizados en base a nuevos datos, mediciones y opiniones de consenso, los cuales se desarrollaron a partir de una serie de reuniones organizadas por el *CMSC (Consortium of Multiple Sclerosis Centers)* (C. Harris et al., 2003), formado por radiólogos y neurólogos tanto procedentes de ámbitos académicos como de la práctica común en el tratamiento de EM y representantes de la *American Academy of Neurology*, la *RSNA (Radiological Society of North America)*, la *ASNR (American Society of Neuroradiology)* y la cooperación *NAIMS (National Institutes of Health and the North American Imaging in Multiple Sclerosis Cooperative)*. Los protocolos son de rutina cerebral y médula espinal pues son los comúnmente utilizados tanto en el diagnóstico como el seguimiento de la EM.

Capítulo 1. Introducción

Protocolo 1: RM cerebral

El protocolo de imagen cerebral proporciona las secuencias mínimas requeridas para manejar el diagnóstico y monitorización de la EM. Este protocolo se sintetiza en la Tabla 3. Estas secuencias pueden ser adquiridas por varios escáneres clínicos e incluyen preferiblemente imágenes 3D T1, 3D T2-FLAIR, 3D T2, y con contraste de Gadolinio ponderada T1, todas con un espesor de corte $\leq 3\text{mm}$ y una secuencia DWI $\leq 5\text{mm}$ de espesor. Se pueden añadir secuencias adicionales propias de otras patologías, dependiendo de las necesidades individuales del paciente y las preferencias de los expertos.

Tabla 3: Protocolo de Imagen de RM en cerebro (Traboulsee et al., 2016).

Parámetros	Descripción
Intensidad de campo	Los escaneos deben ser de buena calidad, con una buena SNR y resolución (resolución de pixel $\leq 1 \times 1 \text{ mm}$).
Orientación de escáner	Plano subcaloso para reformatear secciones oblicuas axiales.
Cobertura	Cerebro completo
Espesor y espaciado	$\leq 3\text{mm}$, sin espaciado
Secuencias principales	Secuencias de eco de gradiente potenciadas en T1 preparadas con inversión-recuperación. Dosis de Gadolinio, 0.1 mmol/kg durante 30 segundos. FLAIR T2WI 3D sagital. 3D T2WI. 2D axial DWI. 3D FLASH post Gadolinio. Las series 3D serían reconstruidas a espesor de 3 mm para la visualización y siguiente comparación de recuento de lesiones.
Secuencias opcionales	<i>Secuencia de densidad protónica (Proton Density, PD)</i> Eco de espín transversal T1 pre o post Gadolinio. Imagen potenciada en susceptibilidad (<i>Susceptibility Weighted Imaging, SWI</i>) para la identificación de la vena central dentro de lesiones T2.

Los escaneos deben ser de buena calidad con una SNR y una resolución espacial adecuada (resolución de pixel $\leq 1 \times 1 \text{ mm}$). La reconstrucción se recomienda que sea a 0.5mm, sin embargo,

Capítulo 1. Introducción

esto puede ser un factor limitante en equipos más antiguos, particularmente aquellos que operan con intensidades de campo más bajas limitarse en algunos equipos antiguos, particularmente en aquellos que operan con intensidades de campo más bajas.

La cobertura debería incluir el cerebro completo y la orientación de las secuencias axiales (adquisiciones de secuencias 2D o reformato de las secuencias 3D) debe ser a lo largo de la línea subcallosa ya que contar con un reposicionamiento consistente es esencial para detectar cambios a lo largo del tiempo.

La mayoría de los equipos son capaces de realizar adquisiciones 3D con vóxeles isotrópicos ≤ 1.2 mm. Los datos pueden ser reformateados para dar secciones axiales o sagitales de 3 mm para lecturas clínicas. Si se utiliza una adquisición 2D, el espesor de corte debe ser ≤ 3 mm y sin espaciado entre cortes. Para la detección de las lesiones de fosa posterior se puede usar imágenes ponderadas T2 o 3D FLAIR y se recomienda que se adquiera, como apoyo, una secuencia 3D T2 o una axial de 3mm fast/turbo spin-echo que pueda servir de respaldo.

Protocolo 2: Imagen de RM en médula espinal

Este protocolo cobra especial relevancia en esta tesis doctoral, ya que ha sido el origen de las adquisiciones de RM utilizadas para el trabajo. En nuestro método empleamos secuencias axiales 3D ponderadas T1 tras la aplicación de contraste de Gadolinio pues son adecuadas para caracterizar la morfología de las estructuras anatómicas (en nuestro caso, la médula espinal cervical) y su contraste.

El campo de visión (FOV, *field of view*) de la adquisición de la imagen RM debe cubrir, como mínimo, la médula espinal de la zona cervical ya que las lesiones que no se manifiestan clínicamente son más comunes y mejor visualizadas en esa zona, de manera que no sería necesaria la examinación de la médula espinal de la parte torácica de manera rutinaria, a no ser que existan sintomatología o signos clínicos a ese nivel. Un resumen del protocolo se recoge en la Tabla 4. No tiene por qué ser necesario examinar la columna torácica rutinariamente si no hay síntomas clínicos o indicios a este nivel. Para la detección de lesiones de difícil visualización se recomienda adquirir principalmente dos secuencias, una debe ser la Sagital potenciada T2 y una segunda que puede ser una secuencia PD, una secuencia STIR o una potenciada T1 de recuperación de inversión con reconstrucción sensible a la fase (espesor ≤ 3 mm). También se recomiendan las secuencias axiales T2 o T2* y axial T1 tras aplicación de contraste, ya que una secuencia sagital T1 nos limitaría a la hora de caracterizar enfermedad intramedular. Cuando la adquisición de imagen de la médula se lleva a cabo a la vez que la adquisición de la cerebral con contraste, no se necesitaría la aplicación de una nueva dosis.

Capítulo 1. Introducción

Tabla 4: Protocolo para imagen de RM en médula espinal (Traboulee et al., 2016)

Parámetros	Descripción
Intensidad de campo	Los escaneos deben ser de buena calidad, con una buena SNR y resolución (resolución de pixel $\leq 1 \times 1$ mm)
Cobertura	Médula espinal
Secuencias principales	Sagital T2 Sagital con atenuación de protones STIR o PST-IR Axial T2
Espesor y espaciado	En sagital: ≤ 3 mm, sin espaciado Axial: 5mm
Secuencias opcionales	Axial T2 a través de toda la médula espinal Sagital T1 Sagital T1 sin Gadolinio y con Gadolinio (al menos 5 minutos después de la administración del contraste)

1.2.4.2 Pautas clínicas: diagnóstico de la EM por imagen

Aunque durante muchos años el diagnóstico clínico de la EM se ha llevado a cabo sin el apoyo en la imagen de RM, en la actualidad, nos aporta información muy relevante y es una prueba de imagen prácticamente imprescindible en la práctica clínica para el manejo de los pacientes con esta enfermedad. En los casos de los pacientes que son dudosos de padecer la enfermedad o que tienen un diagnóstico establecido de EM, pero no tienen estudios previos, se recomienda adquirir una imagen de RM cerebral con y sin Gadolinio, además de recomendarse también la adquisición de imagen de RM en la médula espinal cervical para mejorar la evaluación diagnóstica de pacientes con o sin mielitis transversa, reduciendo así el número de pacientes que requerirían otra imagen de RM. Entre los pacientes sospechosos de padecer EM se incluyen aquellos con síndrome clínicamente aislado (CIS) de neuritis óptica, de mielitis transversa parcial o de tronco cerebral.

Los pacientes con CIS con una imagen de RM con más de 2 lesiones características (≥ 3 mm en diámetro) tienen un alto riesgo de sufrir EM. Un tercio de los pacientes con CIS (que no han sido tratados con corticoides) tendrán lesiones asintomáticas realzadas por Gadolinio y cumplirán el criterio diagnóstico de McDonald de 2010 (Kang et al., 2014). La detección de nuevas lesiones en secuencias potenciadas T1 realzadas por Gadolinio o en secuencias potenciadas T2 en un seguimiento de RM cerebral puede generar evidencia suficiente para cumplir criterios de difusión en el tiempo o espacio. El tiempo recomendado para el seguimiento de imagen cerebral es de 6 a 12 meses. La proporción de pacientes con CIS de alto riesgo (más jóvenes de 50 años) que desarrollan nuevas lesiones en 3, 6 y 12 meses es 22%, 51%, y 74%, respectivamente. Entre

Capítulo 1. Introducción

los 12-18 meses, la mayoría de los pacientes con CIS de alto riesgo cumplen el criterio diagnóstico de la EM y desarrollan nuevas lesiones en RM o nuevos síntomas clínicos.

Se pueden adquirir imágenes de RM adicionales según el criterio clínico. Algunas investigaciones (Brex et al., 2002; Brodsky et al., 2008; Fisniku et al., 2008) recomiendan una imagen de RM cerebral anual en pacientes con CIS que presentan resultados normales de imágenes de RM cerebral durante 1-2 años o en el momento del desarrollo de nuevos síntomas que coinciden con desmielinización. Además de otros tests como los potenciales evocados y el análisis del LCR, el seguimiento de imágenes de RM nos puede dar apoyo en el diagnóstico de EM demostrando la aparición de nuevas lesiones en pacientes sospechosos de padecer la enfermedad.

En pacientes con resultados ambiguos en la imagen de RM cerebral (no cumplen el criterio de diseminación en el espacio dentro del cerebro), la detección de lesiones en imágenes de médula espinal nos puede dar evidencia adicional en cuanto a la diseminación de la enfermedad en el espacio (a otra región extra cerebral). Las lesiones en la médula espinal cervical tienen una mayor especificidad en el caso de la enfermedad desmielinizante. Las lesiones en la sustancia blanca son muy poco comunes en la médula espinal, mientras que en el cerebro son frecuentes. Un seguimiento mediante imágenes de RM nos daría evidencia de una mejora en la resolución de la lesión, lo que nos indicaría la presencia de un proceso inflamatorio o la detección de nuevas lesiones que podrían llevarnos a un diagnóstico diferencial.

Las imágenes de médula espinal, junto a las cerebrales, se recomiendan si los síntomas involucran a la médula espinal como la mielitis transversa parcial, mielitis transversa completa o una miopatía progresiva sospechosa de ser EMPP. De manera infrecuente, algunos pacientes con EM progresiva pueden tener una señal de médula espinal crónica, difusa y anormalmente extensa. La detección de una lesión longitudinalmente extensa en la médula espinal (≥ 3 segmentos) con predominancia en la zona central y efecto de masa en un paciente con mielitis transversa aguda es propicia a ser neuromielitis óptica.

1.2.4.3 Atrofia de la médula espinal en EM

Los estudios de imagen tanto en la práctica como en los ensayos clínicos están centrados tradicionalmente en los cambios en el cerebro, sobre todo en términos de carga lesional, sin embargo, la médula espinal es una estructura usualmente afectada en EM y representa un papel muy importante en la determinación de la discapacidad clínica. Tal como se refleja en la declaración de consenso de MAGNIMS (Rovira et al., 2015), se ha comprobado la utilidad de incluir como medida rutinaria la pérdida de área de la médula espinal cervical en EM ya que la

Capítulo 1. Introducción

atrofia tiene un papel muy importante en el seguimiento de la enfermedad, pues es un fenómeno común, especialmente en los fenotipos progresivos (Azodi et al., 2017; J. C. J. Bot et al., 2004) y se ha demostrado que tiene una relación directa con la discapacidad clínica (Bjartmar et al., 2000; Bonati et al., 2011; Cohen et al., 2012a; Daams et al., 2015; Oh et al., 2015; Song et al., 2020; Ukkonen et al., 2003) de forma independiente a otras mediciones de imagen (Cohen et al., 2012b; Daams et al., 2015; Oh et al., 2015) y su relación con la discapacidad es mayor que la relación que presenta el número de lesiones focales en médula espinal (Furby et al., 2008).

La atrofia en la médula espinal es consecuencia de diferentes procesos patológicos, incluyendo la transección axonal y la pérdida neuroaxonal asociada, desmielinización, gliosis, y, en última instancia, lesión tisular difusa (Andelova et al., 2019). Aunque estas anomalías patológicas ocurren dentro de las lesiones focales, también hay presentes extensas anomalías de tejido en la médula espinal de pacientes de EM, hecho que puede explicar por qué la atrofia de la médula espinal se da independientemente de las lesiones en ella. Además de esto, la atrofia en médula espinal también tiene lugar, al menos en parte, independientemente de la patología cerebral (Losseff et al., 1996a). Es por todo ello que, en la presente tesis, estudiamos la atrofia de médula espinal medida como volumen (mm^3) o sección media (mm^2) de dicha estructura dando lugar a un biomarcador cuantitativo de imagen, tal como explicamos en la sección 3.7.

1.2.4.3.1 Atrofia de la médula espinal en los distintos fenotipos

La atrofia en médula espinal está presente incluso en las fases más tempranas de la EM y se ha detectado en pacientes con CIS (Biberacher et al., 2015; Daams et al., 2014; Hagström et al., 2017; Rocca et al., 2011). En estos estudios, los pacientes recibieron seguimiento durante 5 años, la menor tasa de atrofia (-0.1% por año) se observó en aquellos que permanecían con CIS, mientras que la mayor ratio (-1.4% por año) se detectó en pacientes que acabaron desarrollando EM. En general, se observa una mayor tasa de atrofia en las formas progresivas de la enfermedad, especialmente en EMSP (-2.2% por año). Mediante un estudio multicentro de MAGNIMS se ha detectado una tasa de -1.22% por año en pacientes con EM estable y -2.01% en pacientes que sufrieron deterioro temporal (Rocca et al., 2019). Se encontró una evolución significativa de la atrofia en pacientes en las etapas tempranas de EMPP en comparación con pacientes de control sanos, pero no en pacientes con EMSP estable, los cuales tienen una discapacidad mayor y más atrofia que los pacientes de EMPP en fases tempranas (Cawley et al., 2018). Aunque la tasa de atrofia puede variar ligeramente entre estudios por razones como la diferencia de cohortes y métodos, está firmemente demostrado que la ocurrencia de atrofia

Capítulo 1. Introducción

medular es mayor que la cerebral, la cual es de aproximadamente -0.5% por año en pacientes de EM (Eshaghi et al., 2018).

1.2.4.3.2 Atrofia de la médula espinal y su relación con la discapacidad

Se han realizado estudios que muestran asociaciones entre: el alcance de la atrofia de la médula en un momento temporal y la discapacidad simultánea (Kearney et al., 2014) y la evolución temporal de la tasa de atrofia en médula espinal y la progresión de la discapacidad (Brownlee et al., 2017; Cohen et al., 2012a; Daams et al., 2014; Hagström et al., 2017; Lukas et al., 2015). Recientemente se ha demostrado que cada aumento del 1% en la tasa anual de pérdida de volumen medular está asociado con un riesgo del 28% de desarrollar progresión de la discapacidad al año siguiente (Tsagkas et al., 2018). Además, una mayor tasa de atrofia medular se asocia con un empeoramiento de medidas más específicas de discapacidad motora como, por ejemplo, *hole peg test (9HPT)* y el *25-foot walking test (25FWT)* (Kearney et al., 2014; Tsagkas et al., 2018). Finalmente, las relaciones entre la atrofia medular y la progresión de la discapacidad son particularmente más fuertes en EMPP (Tsagkas et al., 2019).

1.2.4.3.3 Atrofia de la médula espinal en ensayos clínicos

Debido a que las tasas de atrofia en médula espinal son entre 2 y 3 veces mayores que en cerebro (-1.78% frente a -0.5% por año), en particular en EM progresiva (Bonati et al., 2011; Casserly et al., 2018), y a que la médula espinal es una estructura muy significativa en la patología de EM, su atrofia se considera como una medida exploratoria en ensayos clínicos de fase II y III, especialmente en los pacientes progresivos de EM, aunque todavía de forma mucho menos frecuente que en la atrofia cerebral (Tur et al., 2018). Un ejemplo es el de ensayos clínicos terapéuticos que han incorporado la medida de la atrofia de la médula espinal pero no han demostrado beneficios provenientes de la administración de nuevos medicamentos en estudio (Dupuy et al., 2016; Kapoor et al., 2010; Singhal et al., 2017). Aparte de que exista la posibilidad de que los medicamentos probados no fuesen efectivos, existen otras razones para justificar la falta de éxito de estudios clínicos, relacionadas con las dificultades metodológicas que implica calcular la atrofia de la médula espinal, donde se incluye: artefactos de movimiento y el ruido de imagen; la resolución espacial limitada de los equipos de imagen de RM, factor importante debido al pequeño tamaño de la médula; los estudios multicentro, que presentan variabilidad entre sitios proveniente al uso de distintos equipos con diferentes parámetros de ajuste; y la variabilidad entre estudios debido al uso de diferentes métodos para calcular las dimensiones de la médula espinal (Durand-Dubief et al., 2012; Papinutto et al., 2018).

Capítulo 1. Introducción

Todos estos estudios e información de la significancia de la atrofia de la médula espinal en el diagnóstico y seguimiento de la enfermedad de EM, destaca todavía más la importancia del cálculo de esta característica en el ámbito clínico y nos motiva a desarrollar el presente método para salvar las dificultades técnicas que presenta el complejo estudio en imagen de una estructura anatómica tan sensible como la médula espinal y dar lugar a un biomarcador de imagen que represente esta relación entre atrofia de la médula, fenotipos y discapacidad en EM.

1.2.5 Biomarcadores de imagen

El término “medicina de precisión”, el cual, en los últimos años, está siendo caracterizado por su presencia, se refiere al hecho de innovar en técnicas diagnósticas que permitan estratificar adecuadamente diferentes genotipos y fenotipos de enfermedad, de tal manera que se administre el tratamiento más adecuado, al paciente adecuado, en el momento adecuado. Esto redundará en una mejora en la salud del paciente factorizando el estudio en variables como genomas, entorno y estilo de vida de este (Joyner & Paneth, 2019). La imagen médica representa un papel importante en la medicina de precisión, pues nos puede dar información suplementaria para estratificar pacientes que han sido diagnosticados a partir de síntomas o características de la enfermedad similares y que comparten respuestas al tratamiento (Giardino et al., 2017).

Durante las últimas décadas, la inspección visual y la interpretación subjetiva de las imágenes médicas por expertos clínicos ha sido el método tradicional seguido para propósitos tanto diagnósticos como terapéuticos. Aunque este procedimiento continúa presente en la rutina clínica, existen cada vez más enfoques alternativos que facilitan o permiten interpretar automáticamente los datos que contienen las imágenes mediante algoritmos de IA. Estas mediciones visuales de la imagen, además de ser subjetivas también son métodos exigentes y propensos a errores, son tareas que involucran mucho tiempo, provocando variabilidad inter/intraobservador, lo que puede degradar el resultado clínico esperado. Una forma de reducir la incertidumbre que rodea a esta interpretación es convertir los datos de imagen en características cuantitativas de las cuales se puedan deducir resultados cuantitativos.

Gracias a los rápidos avances en el campo de la adquisición de imagen médica y en el desarrollo de hardware y software en las últimas décadas, contamos con modalidades de imagen médica que permiten capturar información de alta resolución cuantitativa del cuerpo humano tanto anatómica y fisiológica, como funcional y metabólica. Estos aspectos cuantitativos de las modalidades de imagen actuales pueden proporcionarnos características que pueden llegar a ser comparables o complementarias a los biomarcadores de laboratorio.

Capítulo 1. Introducción

Como definición general, con biomarcador nos referimos a un indicador objetivamente medible de alguna condición biológica para evaluar procesos biológicos o respuestas a una intervención terapéutica. Por lo tanto, la medición de características anatómicas, fisiológicas, metabólicas o funcionales de los órganos y tejidos a través de imágenes médicas se reconoce como biomarcador de imagen. Estas características se pueden utilizar para una gran variedad de aplicaciones médicas, como la evaluación de la funcionalidad de tejidos, la cuantificación de la gravedad o el grado de un cambio y la especificación del estadio de la enfermedad. En nuestro estudio, tal como mostramos en la sección 3.7, nuestro biomarcador de imagen extraído para ayudar en el diagnóstico de la EM, serán las dimensiones de médula espinal cervical, ya sea en medidas de volumen (mm^3) o sección media (mm^2).

1.2.5.1 Desarrollo de los biomarcadores de imagen

Como hemos señalado anteriormente, la imagen médica tiene un lugar importante en la medicina de precisión a través del concepto de biomarcadores cuantitativos de imagen (*Quantitative Imaging Biomarkers, QIBs*). Aunque las imágenes médicas siempre se han interpretado de manera cualitativa, objetivamente, una imagen médica es esencialmente adquirida de manera cuantitativa, ya que es una matriz de valores. Estas intensidades de vóxel se pueden considerar biopsias virtuales que representan varias características anatómicas o patológicas. Por lo tanto, los atributos fisiológicos y anatómicos específicos de un paciente pueden derivarse de dichas imágenes cuantitativas, tal como procedemos en nuestro estudio con la medición de las dimensiones de la médula espinal en nuestros pacientes.

Es importante señalar que, aunque estas características cuantitativas extraídas de la imagen pueden tener una gran correlación con algunos procesos clínicos, no siempre tienen por qué revelar la asociación exacta entre una causa y su efecto. Los biomarcadores, generalmente se diferencian entre (a) predictivos, si dan información sobre los efectos de una intervención terapéutica, (b) pronósticos si mejoran la precisión del diagnóstico o pronóstico del paciente, (c) de respuesta si muestra resultados ventajosos después del tratamiento, y (d) de seguimiento si detectan recaídas, recurrencias o toxicidad de un fármaco (Martí-Bonmatí & Alberich-Bayarri, 2016; Oldenhuis et al., 2008).

1.2.5.2 Biomarcadores de imagen y análisis médico de imagen

Los métodos de análisis de imagen normalmente comienzan con una etapa de preprocesado de imagen. La reducción de ruido y los métodos de mejoran de imagen son cruciales para mejorar la calidad de la imagen, además, los artefactos deberían ser eliminados siempre que sea posible, ya que introducen datos no reales y causan sesgos en el análisis. Con frecuencia, existe una

Capítulo 1. Introducción

notable discrepancia entre los ajustes de adquisición de incluso las mismas modalidades de imagen, de manera que, en estas ocasiones, las imágenes se adquieren con diferentes resoluciones espaciales. Dado que la extracción de los *QIBs* puede ser sensitiva a la resolución espacial, como es nuestro caso, es necesario redimensionar las imágenes a un tamaño fijo común como parte de la etapa de preprocesado. En nuestro estudio, y tal como explicamos en la sección 3.4, nuestro preprocesado consiste en la normalización y re-escalado a un tamaño común para evitar heterogeneidades en las imágenes tanto de intensidad como de tamaño.

Siguiendo con la metodología de análisis, las imágenes pueden ser adquiridas del mismo órgano o sujeto, pero bajo diferentes condiciones como tiempo, ángulo o incluso modalidad. El registro de imágenes médicas es la tarea de alinear espacialmente dos o más imágenes para tener una correspondencia espacial del contenido en común. Algunas de las aplicaciones más comunes dentro del registro de imágenes médicas son el alineamiento de secuencias de imágenes adquiridas en distintos momentos temporales, detectar cambios antes y después de un tratamiento o combinar información de distintas modalidades (Valsecchi et al., 2018). Para la elaboración de nuestro método de segmentación automática no fue necesario el registro de imágenes, pero si lo es para el propósito contenido en el Anexo I sobre detección automática de lesiones intramedulares, tal como indicamos en la sección 5.2 de dicho anexo.

Por otra parte, los *QIBs* deberían ser extraídos de *ROIs* bien diferenciadas del resto de estructuras adyacentes. Esta tarea, como ya sabemos, se conoce como segmentación y es uno de los pasos más desafiantes en los estudios de análisis de imagen, ya que, en muchas ocasiones, separar las imágenes en “fondo” y “objetivo” resulta difícil debido a la similitud tanto en la distribución de texturas como de intensidades. Es por ello que, en la presente tesis, se desarrolla un método preciso de segmentación automática de médula espinal cervical, el cual presenta satisfactorios resultados y habilita la extracción de *QIBs*, tal como se recoge en el capítulo 4.

Finalmente, tras el preprocesado y la segmentación de las imágenes, el objeto a estudio, en nuestro caso la médula espinal cervical, debe de ser representado como valores cuantitativos mediante la extracción de *QIBs*. En general, estas características extraídas como *QIBs* suelen representar atributos anatómicos estáticos o biológicos dinámicos. Mientras que los primeros estiman propiedades relacionadas con el volumen, topología, morfología o textura de la estructura bajo estudio, los biomarcadores biológicos cuantifican los diferentes procesos físicos, químicos y biológicos.

Capítulo 1. Introducción

1.2.5.3 Validación de los biomarcadores de imagen

Se considera que los *QIBs* son fiables cuando se puede demostrar su precisión, estabilidad y reproducibilidad. Una etapa importante en el desarrollo de los biomarcadores de imagen es determinar estadísticamente la reproducibilidad y el rango biológico de las características individuales. Las características más reproducibles serán las más aceptables para distinguir las pequeñas variaciones en términos de tiempo, cambios fisiopatológicos o respuesta al tratamiento que buscamos.

Para que los *QIBs* sean útiles tanto en ensayos clínicos como en la práctica real, su investigación debe realizarse y validarse siguiendo unos criterios estandarizados bien establecidos. En concreto, la cualificación, validación y estandarización son partes críticas de su desarrollo. La cualificación se refiere al uso de un biomarcador en un contenido concreto y se relaciona con la aprobación clínica, mientras que la validación se refiere al rendimiento del biomarcador en la fase de prueba. En nuestro estudio, recogemos la cualificación de las dimensiones de médula espinal cervical como biomarcador de la EM en la sección 1.2.4.3.

La calidad del diagnóstico está relacionada con la exactitud del biomarcador, mientras que la calidad del seguimiento de la enfermedad o de la respuesta al tratamiento se basa en su precisión. En este campo, exactitud se refiere a cómo de parecidas son las medidas a los valores reales, y precisión a cómo de consistentes son los resultados cuando las medidas se repiten en varias ocasiones. Por lo tanto, los conceptos de reproducibilidad y repetibilidad se pueden considerar como dos tipos de precisión. El estudio de la repetibilidad normalmente se lleva a cabo en un solo centro con una modalidad de imagen específica. Después de las pruebas en un solo centro, la fase de test debería de completarse en distintos centros bajo diferentes condiciones. Por otra parte, los estudios de reproducibilidad están diseñados para evaluar el impacto de variables como la diferencia de centros o de protocolos de adquisición en la precisión del biomarcador. En la presente investigación, realizamos el estudio de repetibilidad, el cual se recoge en el capítulo 3.

Las validaciones biológicas se basan en la correlación entre los estudios en humanos y otros puntos de referencia como histopatología o pruebas invasivas. Estas validaciones se llevan a cabo como estudios experimentales para seleccionar el criterio que mejor identifica los cambios biológicos que se pretenden evaluar. En estos estudios experimentales se examinan los efectos de los datos epidemiológicos y las variaciones biológicas de las mediciones. En la sección 1.2.4.3, estudiamos cómo la atrofia de la médula espinal cervical, medida como una reducción de las

Capítulo 1. Introducción

dimensiones de esta, tiene relación con los distintos fenotipos y grados de discapacidad en la EM.

Por otra parte, las validaciones clínicas comienzan con estudios observacionales en un solo centro en pacientes que participan en experimentos de biomarcadores de imagen para comprobar cómo de bien funcionan bajo una situación controlada. Si los resultados muestran ser precisos se podría pasar a estudios observacionales o ensayos clínicos multicentro (Martí-Bonmatí & Alberich-Bayarri, 2016). Esta validación clínica es realizada en la presente investigación, encontrando resultados entre el biomarcador propuesto y los distintos fenotipos y valores de EDSS de los pacientes, tal como se recoge en la sección 4.6.

1.3 Segmentación en imagen médica

La segmentación tiene un papel muy importante en el análisis de imagen médica como una etapa de preprocesado para adquirir datos cuantitativos de valor para la interpretación médica, ya que nos permite obtener valiosa información de la anatomía de las estructuras a estudiar de una manera no invasiva. El seguimiento de progresión en enfermedades, la localización de patologías, el modelado biomecánico, la simulación de procesos biológicos y la evaluación para intervenciones quirúrgicas y terapias (Ecabert et al., 2008; El-Baz et al., 2007; Grau et al., 2004; Greenspan et al., 2006; Prastawa et al., 2009) son posibles gracias en parte a la información obtenida de la segmentación (Elnakib et al., 2011).

Esta etapa de procesado clasifica los píxeles de una imagen en aquellos que pertenecen a una clase determinada. La segmentación genera una máscara binaria, donde los píxeles a '1' corresponden con aquellos pertenecientes al tejido de interés y el '0', al resto de la imagen. La segmentación multiclase permite generar diferentes máscaras para cada uno de los tejidos segmentados.

Dentro del proceso de segmentación hay dos niveles: 1) Segmentación semántica, el proceso de asignar una etiqueta categórica a cada píxel de la imagen y 2) Segmentación por instancias (instance segmentation), el proceso de localizar y delinear distintivamente cada objeto de interés en una imagen.

Las herramientas de procesado de imagen para tareas de segmentación utilizadas ocupan tanto una gran amplitud de regiones anatómicas del cuerpo como a nivel celular. Estas técnicas de segmentación se pueden aplicar en cualquier tipo de imagen médica (TC, RM, ultrasonidos, PET...). Algunas de las prácticas de segmentación más antiguas son la umbralización ("Local

Capítulo 1. Introducción

Adaptive Thresholding,” 2009; Norouzi et al., 2014; Sonawane & Dhawale, 2015), la detección de bordes (Vincent & Folorunso, 2009; Acharjya et al., 2012; Aquib Ansari et al., 2017; Dey & Polley, 2016; KumarSingh et al., 2014) y los métodos de agrupación (clustering) (Sonawane & Dhawale, 2015).

A pesar de que estos métodos han resultado útiles a lo largo de los años para tareas de segmentación como la extracción de tejidos sin la necesidad de algoritmos de aprendizaje máquina, como en el caso del cerebro, la utilización de técnicas de IA basadas en redes neuronales convolucionales ha ido en aumento debido a su elevado potencial para la realización de tareas de segmentación, con resultados muy superiores a las técnicas anteriores.

1.3.1 *Machine learning en segmentación de imagen médica*

El desarrollo en tecnologías de la información ha tenido un gran impacto en los avances tecnológicos del cuidado de la salud, especialmente con la integración de la IA. El crecimiento continuo de los datos y su complejidad ha ocasionado que los algoritmos de Machine Learning (ML) sean aplicados cada vez con más frecuencia en el ámbito médico (Davenport & Kalakota, 2019).

El ML es el estudio de algoritmos y modelos estadísticos empleados para desarrollar una tarea específica sin dar instrucciones específicas (Cielen et al., 2016). Estos algoritmos construyen un modelo matemático a partir de un conjunto de datos que contiene unas entradas en forma de información y unas salidas que se quieren obtener como resultado (Bishop, 2006). Hay tres categorías principales en ML: aprendizaje supervisado, no supervisado y reforzado.

- **Supervisado:** los algoritmos de aprendizaje supervisado tradicionales son entrenados con datos que contienen etiquetas designando la salida que se desea para cierta entrada. Este tipo de aprendizaje es el más utilizado y establecido en el ámbito de la segmentación. Hay dos clases de aprendizaje supervisado, la clasificación y la regresión. La tarea de clasificación determina si un objeto está presente en una imagen, la salida es una predicción de valores Booleanos de verdadero y falso. Las tareas de regresión tienen como salida valores numéricos continuos dentro de un rango, como puede ser una medida o el precio de un objeto. En los casos en los que no se puede tener un conjunto de datos etiquetados en su totalidad, se emplean alternativas para superar esta dificultad de tener pocos datos etiquetados frente a una mayoría sin etiquetar. Los conjuntos de datos etiquetados completamente de manera manual son muy costosos de obtener en términos de tiempo y esfuerzo, sobre todo en los casos en los que se

Capítulo 1. Introducción

requiere la ayuda de un experto cómo es en el ámbito médico. Es por ello por lo que han surgido otros tipos de aprendizaje dentro del supervisado: semi-supervisado, débilmente supervisado y aprendizaje activo.

- Los algoritmos de aprendizaje semi-supervisado son entrenados con datos incompletos o etiquetados de manera parcial. La dinámica en el ámbito de la segmentación de imagen médica sería utilizar un conjunto relativamente pequeño de imágenes con anotaciones precisas que indican la ubicación exacta de las regiones de interés. Después, utilizar un conjunto más grande de imágenes sin anotaciones detalladas que pueden contener o no la región de interés, pero no se dispone de máscaras precisas que indiquen su ubicación exacta. Finalmente, utilizar una combinación de datos etiquetados y no etiquetados para entrenar el modelo de segmentación. Los datos etiquetados se utilizan para guiar al modelo y aprender características relevantes, mientras que los datos no etiquetados se utilizan para aprender patrones más generales y representaciones útiles.
- Los algoritmos de aprendizaje débilmente supervisado aprenden de una cantidad limitada de datos etiquetados o etiquetas con ruido. La hipótesis que tenemos que asumir aquí es que las etiquetas pueden ser inexactas (Z. H. Zhou, 2018). Al utilizar anotaciones menos precisas, se puede ampliar la cantidad de datos disponibles para el entrenamiento del modelo y, al mismo tiempo, reducir los costos y el esfuerzo necesario para recopilar etiquetas precisas. En el contexto de la segmentación de imágenes médicas, las anotaciones débilmente supervisadas pueden involucrar, por ejemplo: en lugar de anotar píxeles individuales o regiones precisas en cada imagen, indicar si la imagen contiene la región de interés o dar coordenadas o puntos que representan la presencia o ubicación aproximada de esta.
- El aprendizaje activo es una estrategia que busca mejorar la eficiencia y la calidad del modelo de segmentación al seleccionar de manera inteligente las muestras más informativas o difíciles para ser etiquetadas por expertos, y que, una vez etiquetadas, pueden mejorar significativamente el rendimiento del modelo, lo que optimiza el método de conseguir etiquetas de entrenamiento en términos de eficiencia y costo. El aprendizaje activo se basa en la idea de que el modelo de segmentación puede mejorar más rápidamente si se le permite seleccionar las muestras en las que necesita más información. Al centrarse en las muestras más informativas o difíciles, se puede lograr una segmentación más

Capítulo 1. Introducción

precisa con una cantidad menor de datos etiquetados en comparación con un enfoque de aprendizaje supervisado estándar.

- **No supervisado:** aprenden la salida deseada de los datos sin ninguna etiqueta. En lugar de depender de datos etiquetados, este enfoque se basa en el descubrimiento de patrones y estructuras intrínsecas en los datos no etiquetados. Un ejemplo serían los algoritmos de clustering como los mencionados en la introducción de ésta sección, los cuales agrupan las imágenes en diferentes conjuntos en función de sus características comunes. Algunos ejemplos de estos algoritmos son los autocodificadores (Baldi, 2012), que aprenden representaciones de alta dimensión de las imágenes sin etiquetas pudiendo utilizarse para reconstruir la imagen original y también para identificar características relevantes que ayuden en la segmentación, redes *Deep belief* (Hua et al., 2015) o redes adversarias generativas (*Generative Adversarial Networks, GANs*)(Goodfellow et al., 2014), entre otros.
- **Reforzado:** son los algoritmos de aprendizaje que manejan la toma de decisiones de forma secuencial (Francois-Lavet et al., 2018). Son entrenados de manera que se va aplicando una retroalimentación negativa o positiva de manera dinámica. Estos algoritmos son penalizados cuando la decisión tomada ha sido errónea y premiados cuando ha sido correcta. No se basan en datos etiquetados como el aprendizaje plenamente supervisado, sino más bien en agentes de aprendizaje que interactúan con el entorno.

Durante este trabajo empleamos el método de aprendizaje supervisado mediante la arquitectura que se estudia más en profundidad en la sección 1.3.3.1.5. Este método de aprendizaje resultaba óptimo para nuestra tarea, pues mediante él y, al tener etiquetas precisas y confiables de la región de interés (*Region of Interest, ROI*), podemos entrenar modelos confiables y de alta precisión, características cruciales en el campo de la medicina, siendo capaces así de aprender patrones y características sutiles en nuestras imágenes de RM. Pese a que el proceso de etiquetar manualmente las imágenes es costoso y requiere la intervención de expertos, una vez que se tiene un conjunto de datos etiquetados, el entrenamiento de modelos supervisados puede ser relativamente eficiente y, una vez entrenado, el modelo es capaz de procesar nuevas imágenes rápidamente, lo que es crucial en el entorno médico donde el tiempo es esencial.

Capítulo 1. Introducción

1.3.2 Aprendizaje profundo en segmentación de imagen

1.3.2.1 Introducción a las redes neuronales

Las redes neuronales fueron desarrolladas para simular el funcionamiento del cerebro humano y su complejidad en el procesamiento de información a través de redes neuronales interconectadas, creando así IA. Se desarrollaron poco después de la llegada de los ordenadores en los años cincuenta y sesenta. El algoritmo de perceptrón de Rosenblatt (Jamshidi & Zilouchian, 2001) fue visto como un punto clave en el campo, ya que provocó una gran expectativa entorno a la IA. Sin embargo, tras esta euforia inicial, llegó un periodo de desencanto en el que tanto la necesidad de un gran volumen de datos como la necesidad de una gran potencia computacional fueron vistas como limitaciones para su usabilidad. Un hito importante ocurrió en la década de 1980 con el desarrollo del algoritmo de retropropagación (*backpropagation*), que permitió el entrenamiento eficiente de redes neuronales profundas mediante la optimización de sus pesos y conexiones. Sin embargo, a pesar de este avance, el aprendizaje profundo seguía enfrentando dificultades para superar el llamado "problema de la maldición de la dimensionalidad" y prevenir el sobreajuste en modelos complejos.

Con los años, a principios de este siglo, estos déficits fueron menguando y dando paso a una etapa de evolución para las redes neuronales de la mano de las con unidades de procesamiento gráfico (*Graphical Processing Units, GPUs*), las cuales han desempeñado un papel crucial en el progreso, acelerando el entrenamiento y ejecución de los modelos, pues han sido fundamentales para acelerar los cálculos intensivos requeridos y manejar los altos volúmenes de datos. Esta materia renació como *Deep learning (DL)*, aprendizaje profundo. En la década de 2000, estos avances en la capacidad de cómputo y la disponibilidad de grandes cantidades de datos comenzaron a cambiar el panorama del aprendizaje profundo. La comunidad científica pudo entrenar redes neuronales más grandes y profundas con mejores resultados. Además, se introdujeron nuevas arquitecturas, como las redes neuronales convolucionales (*Convolutional Neural Network, CNN*) para el procesamiento de imágenes y las redes neuronales recurrentes (*Recurrent Neural Network, RNN*) para el procesamiento de secuencias, lo que permitió aplicaciones en campos como la visión por computadora y el procesamiento del lenguaje natural (Dhruv & Naskar, 2020; Yin et al., 2017). El primer entrenamiento de CNN con GPUs en el *desafío ImageNet Large Scale Visual Recognition Challenge (ILSVRC) en 2012*, marcó un hito crucial en la historia del aprendizaje profundo y la visión por computadora al demostrar cómo las GPUs, combinadas con las CNN, podían acelerar el entrenamiento de modelos complejos y mejorar drásticamente el rendimiento en tareas de clasificación de imágenes. El ILSVRC es una competencia anual de visión por computadora siendo uno de los eventos más influyentes en el

Capítulo 1. Introducción

campo del DL y la IA. Esta competencia se ha centrado en la tarea de clasificación de imágenes a gran escala, lo que significa que los participantes deben desarrollar algoritmos y modelos capaces de asignar categorías correctas a una amplia variedad de imágenes. Fue introducido por primera vez en 2010 por el laboratorio de visión por computadora de la Universidad de Stanford y se originó como una forma de impulsar la investigación en el campo de la visión por computadora y, en particular, en el desarrollo de algoritmos y modelos que pudieran abordar la clasificación de imágenes en conjuntos de datos extremadamente grandes y diversos. El equipo que lideró esta revolución en el uso de GPUs para entrenar CNNs en el ILSVRC 2012 fue el equipo de investigadores de la Universidad de Toronto, dirigido por Alex Krizhevsky, Ilya Sutskever y Geoffrey Hinton, quienes utilizaron una red neuronal profunda (*Deep Neural Network, DNN*) llamada "AlexNet" (por el nombre del líder del grupo) que tenía una profundidad significativamente mayor en comparación con los enfoques anteriores. El equipo de Toronto implementó su sistema de entrenamiento en dos GPUs NVIDIA GeForce GTX 580, lo que les permitió entrenar AlexNet mucho más rápido que las configuraciones tradicionales basadas en unidades centrales de procesamiento (*Central Processing Units, CPUs*) y lograron una tasa de error de alrededor del 16.4%, superando por un margen significativo a los enfoques anteriores. Es por esto por lo que AlexNet marcó un punto de inflexión en la investigación de visión por computadora y el campo del aprendizaje profundo, demostrando la importancia de las DNNs y la eficacia de las GPUs en el entrenamiento de modelos complejos.

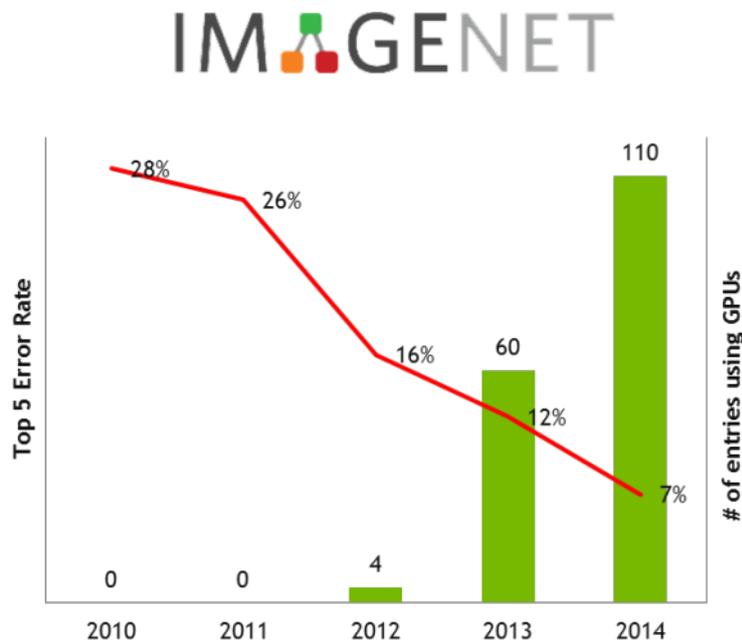


Figura 10. Se muestra la rápida decaída de las tasas de error ganadoras del ILSVRC (línea roja) y el creciente número de participaciones que emplean GPU cada año (barras en verde) (NVIDIA and IBM Cloud Support ImageNet Large Scale Visual Recognition Challenge | NVIDIA Technical Blog, n.d.).

Capítulo 1. Introducción

Desde 2012, el uso de las GPUs para el entrenamiento de redes neuronales se ha generalizado tanto en el ILSVRC como en el mundo de la IA en general.

Si bien la IA ha superado a los humanos en ciertos dominios específicos, todavía hay algunas áreas en las que los humanos superan a la IA en términos de comprensión emocional, creatividad, adaptabilidad y sentido común. La IA ha demostrado un rendimiento sobresaliente en tareas específicas, por ejemplo, en juegos como ajedrez y Go, arquitecturas (como AlphaZero) (Mnih et al., 2015; Zahavy et al., 2023) ha superado a los mejores jugadores humanos. En el reconocimiento de patrones, la IA ha superado la precisión humana en tareas como el reconocimiento de imágenes y la detección de objetos en imágenes médicas (Gonem et al., 2020). Además, puede procesar grandes cantidades de datos y realizar cálculos complejos en un tiempo mucho más corto que los humanos, lo que la hace valiosa en aplicaciones que requieren análisis rápido y eficiente, como el procesamiento de datos en tiempo real y el análisis de grandes conjuntos de datos. Por otra parte, existen otras tareas en las que, pese a que la IA ha avanzado y lo sigue haciendo a ritmo acelerado en la generación de contenido creativo, como música y arte, todavía no puede igualar la profundidad de la creatividad humana (Cope et al., 2021; *What DALL-E Reveals About Human Creativity*, n.d.). Además, características como la empatía, la comprensión contextual y la toma de decisiones éticas son áreas en las que, pese a que se han demostrado avances como en el procesamiento del lenguaje natural, la IA todavía carece de un conocimiento profundo (Jarrahi, 2018; Klein & Martinez, 2022; Liu-Thompkins et al., 2022) y un sentido común sobre el mundo que los humanos adquirimos a lo largo de nuestras experiencias y educación. La relación entre el rendimiento de la IA y el rendimiento humano es, en definitiva, un campo de estudio en constante evolución.

1.3.2.2 Redes Neuronales Convolucionales (CNNs)

Las Redes Neuronales Convolucionales (CNNs) han emergido como una arquitectura en el campo del aprendizaje profundo con ventajas en exactitud y rendimiento, especialmente en el procesamiento y análisis de imágenes (Litjens et al., 2017). Estas redes están diseñadas para capturar de manera eficiente y automática características y patrones complejos presentes en imágenes, superando las limitaciones de los métodos tradicionales de procesamiento de imágenes.

En imagen médica, tanto la adquisición de estructuras 2D como 3D, son cruciales para la identificación y delimitación de regiones o estructuras de interés como órganos, tumores o estructuras anatómicas, entre otros. Manteniendo las relaciones espaciales, las CNNs pueden ser capaces de desarrollar tareas de reconocimiento de imagen como puede ser la clasificación,

Capítulo 1. Introducción

localización, detección, segmentación y registro. Son una de las evoluciones de ML más populares en este ámbito que engloba tareas de aprendizaje visual, ya que tienen como característica especial la capacidad de mantener relaciones locales en la imagen mientras lleva a cabo su aprendizaje.

Esto hace que se capturen relaciones de características importantes en la imagen a la vez que se va reduciendo el número de parámetros que el algoritmo tiene que calcular, aumentando así, de paso, la eficiencia computacional. Las CNNs son capaces de recibir a su entrada y procesar tanto imágenes 2D como 3D, lo que resulta una ventaja muy útil en el diseño de un sistema de uso hospitalario, pues algunas modalidades de imagen como los rayos-X son 2D mientras que otras como los TC o las RM son volúmenes 3D.

1.3.2.2.1 Arquitectura

A diferencia de las redes neuronales convencionales, las CNNs incorporan dos ideas clave: la convolución y el *pooling*.

La convolución es una operación matemática que involucra una multiplicación puntual entre una máscara o filtro (kernel) y una sección de la imagen. Esta operación se realiza en varias ubicaciones de la imagen y produce un mapa de características, donde cada valor representa la respuesta de activación de un filtro en una posición específica. Los filtros pueden detectar características simples, como bordes o esquinas, en las capas iniciales, y características más complejas y abstractas, como objetos o texturas, en las capas posteriores.

Como se muestra en la Figura 11, la CNN recibe una imagen de entrada y transforma la información de los píxeles en características mediante dichas capas. Al final del proceso, esta información en forma de características llega a la capa *Fully Connected* que asigna probabilidades o puntuaciones, clasificando así a la entrada en la clase con la probabilidad más alta.

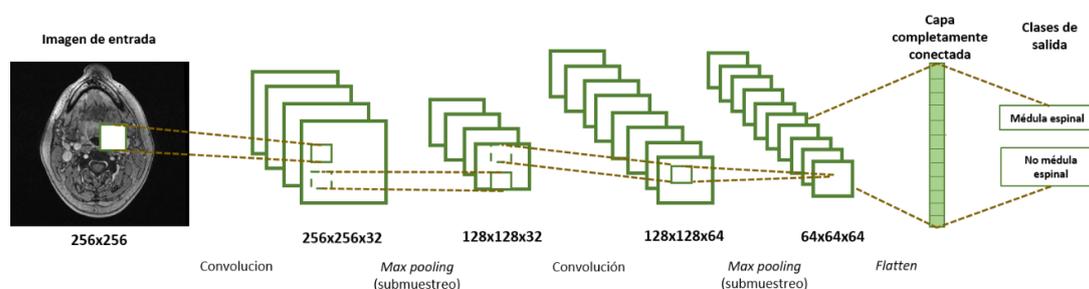


Figura 11. Ejemplo de clasificación de médula espinal. Una imagen de entrada 2D axial se procesa a través de una CNN en forma esquemática. La extracción de características es llevada a cabo por las capas de convolución para después pasar la capa de *max-pooling* y aplicar un submuestreo de factor 2. Este proceso es repetido en la segunda capa de convolución y *max-pooling*. Por último, la salida de la segunda capa de *pooling* es pasada a un vector unidimensional (*flatten*) y se genera la probabilidad de clasificación mediante la capa totalmente conectada.

Capítulo 1. Introducción

Capa convolucional

Una convolución se define como una operación entre dos componentes. En análisis de imagen, estaría compuesta por valores de entrada (por ejemplo, píxeles) que corresponden a una posición de la imagen, y un filtro (o kernel), ambos representados como vectores numéricos. Esta convolución da como resultado una salida. Dicho proceso es iterativo, de manera que el filtro recorre la imagen y cada vez va procesando unos valores de entrada distintos hasta que la recorre de manera completa y se produce un mapa de características (o activaciones) en el que se puede observar dónde el filtro se ha activado con más fuerza y, por lo tanto, es indicador de que se haya detectado alguna característica como podría ser una línea recta o un borde curvo. Este mapa de características pasaría a ser la entrada de la siguiente capa de la arquitectura.

Este proceso de convolución tiene intrínsecas 3 características para ejecutar el aprendizaje de manera más eficiente: conexiones dispersas (*sparse connections*), intercambio de parámetros (o de pesos) y representación invariante (Heaton, 2018).

- La propiedad de conexiones dispersas significa que, a diferencia de algunas redes neuronales en las que cada neurona de entrada está conectada a cada neurona de salida de la capa anterior, las conexiones de las CNNs son diseminadas, es decir, que solo algunas neuronas están conectadas a la siguiente capa. Esto se consigue haciendo el kernel más pequeño que la entrada. Por ejemplo, cuando procesamos una imagen de entrada, ésta podría tener miles o millones de píxeles, aunque se pueden detectar pequeñas pero importantes características (como bordes) con kernels que cubran tan sólo decenas o centenas de píxeles, por lo que necesitaríamos almacenar menos parámetros y por lo tanto reducir los requerimientos de memoria del modelo y mejorar su eficiencia estadística.
- La idea principal detrás del intercambio de parámetros es que ciertas características, como bordes, texturas o patrones simples, son relevantes en diferentes partes de una imagen. En lugar de aprender estos patrones independientemente en cada ubicación, las CNN permiten compartir los mismos pesos de las neuronas en diferentes regiones de la imagen, lo que reduce drásticamente la cantidad de parámetros que se deben aprender y mejora la capacidad de generalización del modelo. La capa convolucional consta de un conjunto de filtros (kernels) que son pequeñas ventanas deslizantes. Cada filtro se aplica repetidamente en diferentes ubicaciones de la imagen de entrada y realiza una operación de convolución para extraer características locales. En lugar de

Capítulo 1. Introducción

tener un filtro único para cada ubicación en la imagen, el intercambio de parámetros permite que un mismo filtro se utilice en varias posiciones. Durante el entrenamiento, el mismo conjunto de pesos del filtro se aplica a diferentes partes de la imagen, y a medida que el modelo se ajusta a los datos, los pesos se adaptan para capturar patrones relevantes en toda la imagen. Este intercambio de parámetros, además de hacer que la CNN sea mucho más eficiente en términos de uso de memoria y capacidad de cómputo, ayuda a generalizar mejor a nuevas imágenes que nunca antes se han visto, ya que ha aprendido patrones comunes que son útiles en múltiples contextos.

- En la convolución, esta particular forma de intercambiar parámetros hace que esta capa tenga también la propiedad conocida como representación invariante. Esta propiedad establece que las capas convolucionales son capaces de reconocer y extraer características en una imagen de manera invariante a las traslaciones espaciales, es decir, son eficaces en el reconocimiento de patrones y características locales en una imagen, independientemente de su ubicación exacta. Por ejemplo, cuando un filtro convolucional pequeño que se utiliza para detectar un borde vertical en una imagen se aplica a una posición específica de la imagen y detecta un borde vertical, si la imagen se traslada horizontalmente hacia la izquierda o hacia la derecha, el mismo filtro aplicado a la nueva posición también detectará un borde vertical. La capa convolucional es “equivariante a la traslación” porque las características que extrae son consistentes y desplazadas de manera proporcional a los desplazamientos en la entrada.

El producto de convolución está definido por el símbolo $*$. En las siguientes líneas se muestra la expresión analítica de cómo se define una salida (o mapa de características) $s(t)$ cuando una entrada $I(t)$ es convolucionada con un filtro o kernel $K(a)$.

$$s(t) = I(t) * K(a) \quad (1)$$

Si t sólo puede tomar valores discretos n , la convolución discreta se define como:

$$s(n) = \sum_a I(a) \cdot K(n - a) \quad (2)$$

La anterior expresión corresponde a una operación convolucional en una dimensión. La misma operación pero en dos dimensiones con entrada $I(m, n)$ y un kernel $K(a, b)$ sería:

$$s(m, n) = \sum_a \sum_b I(a, b) \cdot K(m - a, n - b) \quad (3)$$

Capítulo 1. Introducción

Mediante la ley conmutativa, el kernel se puede invertir y la expresión anterior sería equivalente a:

$$s(m, n) = \sum_a \sum_b I(m - a, n - b) \cdot K(a, b) \quad (4)$$

Finalmente, las redes neuronales implementan la función de correlación cruzada, que matemáticamente es la misma operación de convolución, pero sin invertir el kernel:

$$s(m, n) = \sum_a \sum_b I(m + a, n + b) \cdot K(a, b) \quad (5)$$

Capa RELU

Otro aspecto importante en las CNNs es el uso de funciones de activación no lineales, como ReLU (*Rectified Linear Unit*) o sus variantes. Estas funciones introducen no linealidades en la red, permitiendo capturar relaciones más complejas entre las características y aumentando la capacidad de representación.

La capa ReLU es una función de activación que transforma los valores de entrada negativos a cero, lo que simplifica y acelera los cálculos y el entrenamiento, además de evitar el problema de desvanecimiento de gradiente. Matemáticamente se define como:

$$f(x) = \max(0, x) \quad (6)$$

Siendo x la entrada a la neurona. Esta capa suele seguir a la de convolución. Vale la pena señalar que la función de activación ReLU se ha añadido de manera relativamente reciente a la arquitectura de las redes neuronales, aunque antes de su implantación se solían utilizar funciones de activación como *tanh* y *sigmoid*.

En términos de velocidad y precisión, la ReLU tiene una importante ventaja sobre estas funciones de activación, pues se podrían aplicar modelos más complejos y entrenarlos durante más tiempo. La función de activación ReLU ha ido progresivamente relevando a otras en las arquitecturas de las CNN, hasta el punto de que ahora es utilizada como la función de activación por defecto (Kallenberg et al., 2016). Algunas variantes de ReLU pueden ser *leaky RELUs*, *Randomized RELUs* y *parametric RELUs*.

Capítulo 1. Introducción

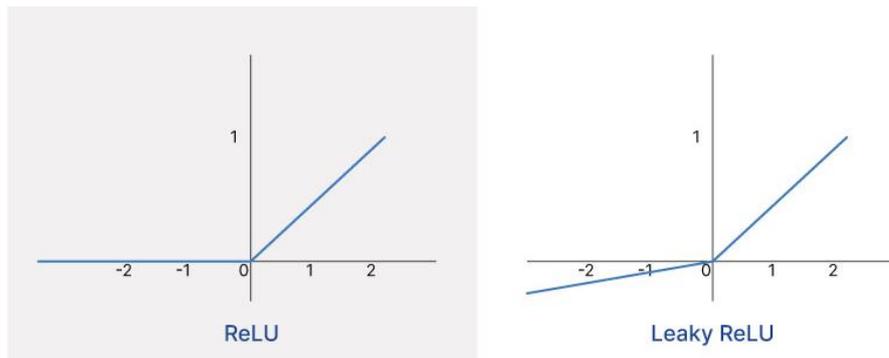


Figura 12. Comparación de las funciones de activación. ReLU (izquierda). Leaky ReLU (derecha). Siendo el eje x la entrada a la red y el eje y la salida. (Activation Functions: With Real-Life Analogy and Python Code - Shiksha Online, n.d.)

Los otros tipos de funciones de activación que suelen ser utilizadas tras la última capa con el fin de hacer que la salida este en un rango deseado son:

- Sigmoide: devuelve una salida con un rango de [0,1] como podemos observar en la Figura 13. La función de activación sigmoide es utilizada cuando queremos interpretar la salida como una probabilidad en los casos de clasificación binaria. La ecuación 10 recoge su expresión matemática:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (7)$$

- Tangente hiperbólica: devuelve una salida con un rango [-1,1] como podemos observar en la Figura 13. Esta función tiene una forma similar a la de la sigmoide, pero la tangente hiperbólica o tanh, presenta simetría sobre el origen del sistema de coordenadas. Esta función puede ocasionar desvanecimiento de gradiente.

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (8)$$

- Softmax: devuelve un número de probabilidades que suman 1, cumpliendo la expresión:

$$\text{Softmax}(x)_i = \frac{e^{x_i}}{\sum_C e^{x_c}}, \quad \text{con } i = 1, \dots, C, \quad (9)$$

Donde C es el número de clases. Esta función es particularmente útil en clasificación o segmentación de imagen cuando hay más de dos clases y se quiere conocer la probabilidad de cada una.

Capítulo 1. Introducción

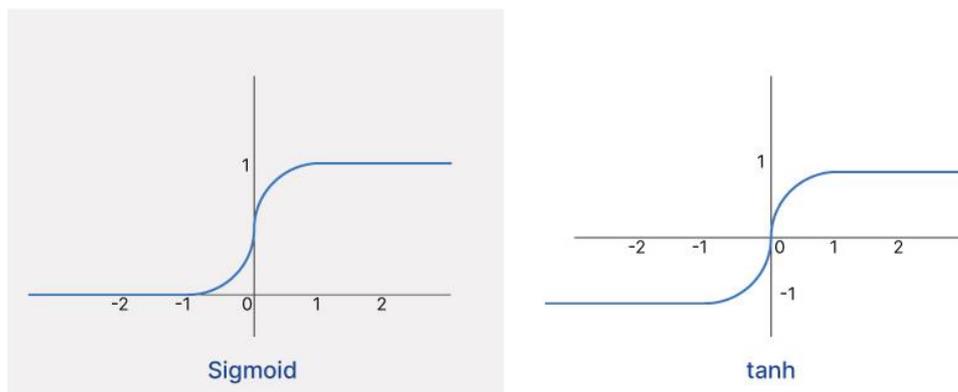


Figura 13. Comparación de las funciones de activación. Sigmoid (izquierda). Tangente hiperbólica (derecha). Siendo el eje x la entrada a la red y el eje y la salida. (Activation Functions: With Real-Life Analogy and Python Code - Shiksha Online, n.d.)

En nuestro trabajo, se emplea tanto la función de activación ReLU como la sigmoide. La ReLU la empleamos en el bloque residual del mecanismo de aprendizaje *attention-aware* empleado en nuestra arquitectura final para introducir no linealidad en el bloque residual y ser capaces de aprender relaciones más complejas en los datos. La función de activación sigmoide es utilizada en la parte decodificadora de la arquitectura, de manera que nos da una probabilidad en la capa de salida de que los datos de entrada pertenezcan a una clase (región perteneciente a médula espinal cervical) o no (resto de la imagen).

Pooling

El *pooling*, por su parte, es una técnica de reducción de dimensionalidad que se aplica después de las capas convolucionales. Su objetivo principal es disminuir la cantidad de información mientras se conservan las características más relevantes. Comúnmente se utiliza el *max-pooling*, que selecciona el valor máximo en una vecindad local, o el *average-pooling*, que calcula el promedio de los valores en la vecindad. Esto reduce la resolución espacial y proporciona una invariancia parcial a las pequeñas variaciones en la posición de las características.

El proceso de *pooling*, a diferencia de los filtros, se aplica en pequeños *patches* de tamaño $P_q \times P_q$ de manera que se genera un solo número que representa la información de esta pequeña región en la que nos encontramos.

El tamaño más común, P_q , de la región en la que se hace *pooling* es 2×2 , ya que con “saltos” de 2 no habría superposición entre zonas. Este tipo de medida/convenio es muy frecuente. Sin embargo, hay argumentos que señalan que podría ser ventajoso tener al menos cierta superposición entre los distintos *patches* donde se lleva a cabo el *pooling* ya que reduce la posibilidad de sobreajuste o *overfitting* (Aggarwal, 2018).

Capítulo 1. Introducción

Capa fully connected

Una característica distintiva de las CNNs es su capacidad de aprendizaje jerárquico. A medida que la información fluye a través de las capas de la red, las características se vuelven cada vez más complejas y abstractas. Esto se logra mediante el apilamiento de múltiples capas convolucionales y de pooling, seguidas de capas completamente conectadas (*fully connected*). Las capas completamente conectadas están compuestas por neuronas que reciben información de todas las características extraídas en las capas anteriores y generan salidas que representan las clases o categorías finales.

Esta es la última capa de la red. Al igual que en el caso de las capas de convolución, ReLU y *pooling*, en la arquitectura puede haber una o más capas *fully connected*, dependiendo del nivel de abstracción de características que queramos conseguir. La entrada a la capa *fully connected* viene de la salida de la última capa convolucional o de *pooling*.

Esta entrada es un vector que se ha creado a partir del mapa de características tras su conversión a una dimensión (*flatten*) y a partir del cual se calcula un valor de probabilidad de clasificación de entre las diferentes opciones que haya en estudio. Principalmente, la capa *fully connected* analiza la combinación de las características que están más fuertemente activadas y que indicarían que la imagen pertenece a una clase en particular.

La capacidad de las CNNs para aprender automáticamente características y patrones ha sido fundamental en el campo de la visión artificial. Además, se han desarrollado diversas técnicas para mejorar su rendimiento, como la utilización de arquitecturas más profundas, el uso de conexiones residuales para aliviar el problema del desvanecimiento del gradiente y la aplicación de técnicas de regularización, como *dropout* o *batch normalization*, para evitar el sobreajuste.

En resumen, las CNNs son una arquitectura clave en el campo del aprendizaje profundo y la visión artificial. Su capacidad para aprender automáticamente características y patrones complejos en imágenes ha permitido grandes avances en tareas como la clasificación de objetos, la detección de objetos, el reconocimiento facial y la segmentación.

1.3.2.2.2 Métodos de aprendizaje

Las redes neuronales y las redes neuronales profundas pueden ser rápidas y precisas, pero también difíciles de entrenar. El entrenamiento involucra un proceso iterativo mediante la propagación de información hacia delante y hacia atrás, como mostramos en la Figura 14. Durante la propagación hacia delante, la información inicial se propaga desde la entrada a través de cada una de las unidades ocultas de cada capa. Cada una de estas unidades ocultas (neurona) lleva a cabo transformaciones en los datos recibidos de la capa anterior y los envía a la siguiente

Capítulo 1. Introducción

capa hasta que se llega a la última capa y se obtiene una salida final para una muestra de entrada inicial.

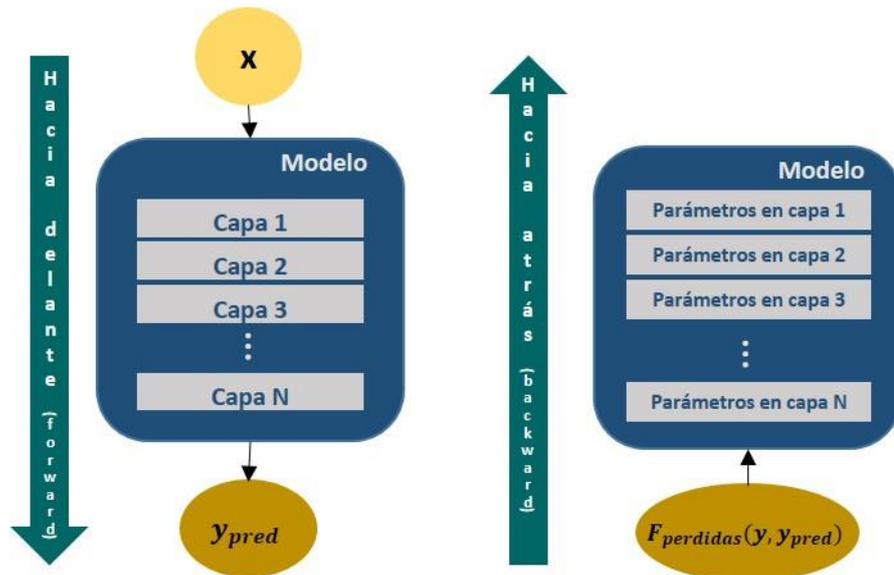


Figura 14. Propagación hacia delante (izquierda) y hacia atrás (derecha). La primera calcula la salida del modelo a través de todas las capas, la segunda actualiza los pesos de las capas para minimizar al máximo las pérdidas en la siguiente iteración.

El principal objetivo del algoritmo de entrenamiento de la red es minimizar el error de la función de pérdidas. La propagación hacia atrás permite que el flujo de información pasada a la función de pérdidas se propague e itere por las neuronas de manera inversa a como se ha hecho para calcular la salida del nodo final. Durante este proceso iterativo, los pesos de cada capa de neuronas se ajustan buscando reducir el error de la función de pérdidas.

1.3.2.2.3 Funciones de pérdidas

La capacidad de una red de aproximar lo máximo posible la salida que deseamos (*ground truth*) a las imágenes de entrenamiento de entrada se cuantifica definiendo una función de pérdidas que se calcula mediante muestras del conjunto de entrenamiento, los pesos y los sesgos. El objetivo de la función de pérdidas se puede observar en la Figura 15. Durante el entrenamiento, la red trata de minimizar la función de pérdidas a un valor lo más cercano posible a cero. Si el valor de las pérdidas es alto, la función de pérdidas “penalizará” a la red cambiando notablemente los pesos, mientras que, si el valor es bajo, los pesos sólo cambiarán de manera muy suave ya que la red está funcionando bien.

Capítulo 1. Introducción

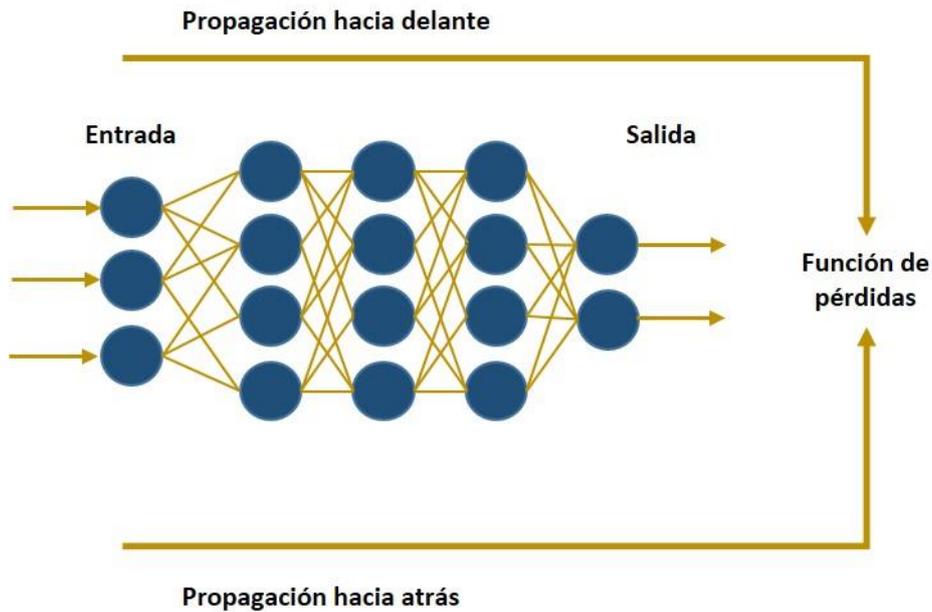


Figura 15. Proceso de entrenamiento de una red neuronal. Las redes neuronales aprenden en un proceso iterativo propagando la información hacia adelante y hacia atrás mediante una función de pérdidas. La red se centra en minimizar la función de pérdidas actualizando sus pesos durante la propagación hacia atrás, de manera que la nueva información de salida se retroalimenta a la pérdida por propagación directa.

Vamos a ver a continuación un breve repaso de las funciones de pérdidas más comunes que existen para la segmentación de imagen en DL:

- Pérdida de entropía cruzada (Cross-Entropy Loss): Esta es una de las funciones de pérdida más utilizadas en tareas de clasificación y segmentación. La pérdida de entropía cruzada mide la discrepancia entre la distribución de probabilidad predicha por el modelo y la distribución de probabilidad real de las etiquetas. En la segmentación de imágenes, la pérdida de entropía cruzada se aplica píxel a píxel, calculando la entropía cruzada entre la máscara de segmentación predicha y la máscara de segmentación real.

$$L_{CE} = - \sum_{i=1}^n y_{true_i} \log(y_{pred_i}) \quad (10)$$

Donde n es el número de clases, y_{true} es la etiqueta real e y_{pred} es la probabilidad predicha para la clase i^{th} .

- Pérdida de Dice (Dice Loss): La pérdida de Dice se basa en el coeficiente de Dice, que es una métrica de similitud ampliamente utilizada en la segmentación de imágenes. Esta función de pérdida mide la superposición o solapamiento entre las máscaras de segmentación predicha y real. La pérdida de Dice se calcula como 1 menos el coeficiente de Dice, y su objetivo es maximizar la similitud entre las máscaras.

Capítulo 1. Introducción

$$L_{DICE} = 1 - \frac{2y_{pred_i} + y_{true_i} + \gamma}{y_{pred_i}^2 + y_{true_i}^2 + \gamma} \quad (11)$$

Donde γ es un valor muy pequeño para evitar cocientes indeterminados.

- Pérdida de Jaccard (Jaccard Loss): La pérdida de Jaccard, también conocida como índice de similitud de Jaccard o coeficiente de Intersección sobre Unión (*Intersection over Union, IoU*), es otra métrica utilizada en la segmentación de imágenes. La pérdida de Jaccard se calcula como 1 menos el coeficiente de Jaccard y busca maximizar la similitud entre las máscaras de segmentación predicha y real.

$$L_{jaccard} = 1 - \frac{(y_{true} \cdot y_{pred}) + \gamma}{(y_{true} + y_{pred} - y_{true} \cdot y_{pred}) + \gamma} \quad (12)$$

- Pérdida focal (Focal Loss): La pérdida focal es una función de pérdida diseñada específicamente para abordar el desequilibrio de clases en la segmentación de imágenes. En lugar de tratar todas las clases de manera uniforme, la pérdida focal se enfoca en los píxeles mal clasificados que pertenecen a las clases minoritarias. Esto se logra mediante la asignación de pesos diferenciales a los píxeles en función de su dificultad de clasificación.

$$L_{focal} = - \sum_{i=1}^n (i - y_{pred_i}) \log(y_{pred_i}) \quad (13)$$

La elección de la función de pérdida adecuada depende del problema específico que se esté abordando y de la naturaleza de los datos. En general, el objetivo es encontrar una función de pérdida que sea apropiada para medir la discrepancia entre las predicciones del modelo y las etiquetas reales, y que se pueda optimizar eficientemente durante el entrenamiento para mejorar el rendimiento del modelo. En nuestro estudio, empleamos la función de pérdidas denominada *Focal Tversky Loss*, la cual es explicada en la sección 1.3.4.1 del presente capítulo.

1.3.2.2.4 Algoritmos de optimización

La optimización involucra a los métodos utilizados para actualizar los pesos de la red y la tasa de aprendizaje con el fin de minimizar la función de pérdidas escogida. Hay muchos tipos de estrategias de optimización u optimizadores, que intentan alcanzar el mínimo en la función de pérdidas en el menor número de iteraciones:

Capítulo 1. Introducción

- Descenso de gradiente estocástico (*stochastic gradient descent, SGD*): SGD es uno de los algoritmos de optimización más básicos y ampliamente utilizados en DL. Funciona mediante la actualización iterativa de los parámetros del modelo en la dirección opuesta al gradiente de la función de pérdidas. En cada iteración, se selecciona un subconjunto aleatorio de muestras de entrenamiento (llamado lote o *batch*) para estimar el gradiente. El SGD puede sufrir de problemas como convergencia lenta y sensibilidad a la elección de la tasa de aprendizaje.
- Descenso de gradiente con momento (*Momentum*): El algoritmo de *Momentum* busca mejorar el SGD al agregar una componente de momento. El momento acumula una fracción de los gradientes anteriores para acelerar el proceso de convergencia. Al utilizar la información de gradientes pasados, el *Momentum* puede ayudar a superar los mínimos locales y converger más rápidamente.
- RMSprop: RMSprop es otro algoritmo popular para optimización en DL. Aborda el problema de la tasa de aprendizaje adaptativa. En lugar de utilizar una tasa de aprendizaje global fija, RMSprop ajusta la tasa de aprendizaje para cada parámetro del modelo según la magnitud de los gradientes recientes. Esto ayuda a adaptar la tasa de aprendizaje a diferentes parámetros y características de la función de pérdidas.
- Algoritmo adaptativo de momento estocástico (Adam): Adam combina las ideas del *Momentum* y RMSprop en un solo algoritmo de optimización. Al igual que el *Momentum*, Adam utiliza una componente de momento para acelerar el proceso de convergencia. Además, utiliza estimaciones adaptativas de los momentos de primer y segundo orden para ajustar la tasa de aprendizaje. Adam ha demostrado ser eficaz en una amplia gama de problemas de DL debido a su capacidad para adaptarse automáticamente a diferentes tipos de funciones de pérdidas. Este será el optimizador que empleamos en el entrenamiento de nuestro modelo.

Además de los algoritmos mencionados anteriormente, también existen enfoques de optimización de segundo orden, como el método de Newton y el método de Levenberg-Marquardt. Estos algoritmos tienen en cuenta la curvatura de la función de pérdidas al calcular las actualizaciones de los parámetros. Sin embargo, los métodos de segundo orden son computacionalmente más costosos en comparación con los métodos de primer orden y pueden no ser prácticos para aplicaciones de DL en grandes conjuntos de datos.

Estos son solo algunos de los algoritmos de optimización utilizados en la segmentación de imágenes en DL. Al igual que en otros componentes durante el diseño, la elección del algoritmo de optimización adecuado depende del problema específico, el tamaño del conjunto de datos y

Capítulo 1. Introducción

la arquitectura del modelo. Además, existen numerosas variantes y extensiones de estos algoritmos, cada una con sus propias características y ventajas para abordar diferentes desafíos en el entrenamiento de modelos de segmentación de imágenes. La elección y el ajuste adecuados del algoritmo de optimización son aspectos importantes en la investigación para lograr mejores resultados en la segmentación de imágenes en DL.

En nuestro estudio, empleamos una novedosa técnica de optimización descrita en la sección 1.3.4.2, la cual nos permite mejorar la convergencia y el rendimiento del entrenamiento variando el valor de la tasa de aprendizaje entre un rango predefinido de valores a lo largo del tiempo.

1.3.2.2.5 Enfoques de regularización

Durante el entrenamiento, la precisión de las redes neuronales profundas puede converger continuamente a la perfección, pero verse degradada en la validación. Esto es debido al *overfitting*, que tiene lugar debido a que la red se ha ajustado demasiado a los datos de entrenamiento y no generaliza bien a nuevos datos. Las técnicas de regularización se utilizan con el fin de prevenir esto y mejorar la generalidad de un modelo.

Los tipos más comunes de regularización son L1 y L2, de manera que ambos métodos añaden un término de regularización a la función de pérdidas del modelo. L1 (regresión Lasso) penaliza el valor absoluto de los pesos, mientras que L2 (regresión rígida), también conocida como decaimiento de peso, fuerza a los pesos a decaer a cero. Otras técnicas de regularización son el aumento de datos, *dropout*, *early stopping* y *batch normalization*.

- Aumento de datos: el aumento de datos es el método utilizado para aumentar el tamaño de las muestras de entrenamiento con el fin de mejorar la capacidad de generalización del modelo de aprendizaje profundo, al exponerlo a una mayor variedad de casos y escenarios. Esto ayuda al modelo a capturar características relevantes y a ser más robusto frente a las variaciones en los datos de prueba. Algunas técnicas comunes de aumento de datos utilizadas en la segmentación de imágenes en DL son: desplazamiento y rotación, escalado y zoom, cambios de contraste y brillo, recorte aleatorio y volteo horizontal y aumento basado en transformaciones geométricas.
- *Dropout*: la idea del dropout es eliminar aleatoriamente cierto porcentaje de neuronas en cada capa durante el entrenamiento con el fin de evitar el *overfitting* en redes neuronales y mejorar la generalización del modelo. Cuando se entrena una red neuronal, es común que algunas neuronas dependan demasiado de las neuronas vecinas. Esto puede llevar a un escenario en el que el modelo se ajuste demasiado a los

Capítulo 1. Introducción

datos de entrenamiento, perdiendo la capacidad de generalizar y obtener buenos resultados en datos nuevos. La idea detrás del dropout es desactivar aleatoriamente un conjunto de neuronas durante el entrenamiento, lo que significa que se excluyen temporalmente de la propagación hacia adelante y hacia atrás de las señales en la red. Estas neuronas desactivadas se "caen" (*dropout*) temporalmente y no contribuyen a la predicción del modelo en ese paso en particular.

Durante cada iteración del entrenamiento, se elige aleatoriamente un conjunto diferente de neuronas para ser apagadas. Al hacerlo, se fuerza a la red a aprender características más robustas y redundantes, ya que no puede depender demasiado de ninguna neurona en particular. En esencia, el dropout evita la especialización excesiva de las neuronas, lo que conduce a una red más general y menos propensa al sobreajuste. Durante la fase de test o validación, no se utiliza el dropout, todas las neuronas están activas y se ponderan de acuerdo con la probabilidad de retención utilizada durante el entrenamiento. Esto es necesario para obtener predicciones consistentes y fiables del modelo.

- *Early stopping* o detención temprana: la técnica de *Early Stopping* es comúnmente utilizada en el entrenamiento de modelos con el fin de evitar el sobreajuste. Durante el entrenamiento de un modelo de DL, se realiza un seguimiento del desempeño del modelo en un pequeño conjunto de datos de test de manera que se evalúa utilizando una métrica relevante para el problema, como la precisión, el error o la pérdida. Mediante *Early Stopping* el entrenamiento del modelo se detiene cuando el desempeño en el conjunto de test deja de mejorar significativamente, lo que se determina comparando el desempeño en el momento actual con el mejor desempeño obtenido hasta el momento.
- *Batch Normalization* o normalización por lotes: es un método para mejorar el rendimiento, rapidez y estabilidad de un modelo de redes neuronales. Fue propuesta originalmente por Sergey Ioffe y Christian Szegedy en 2015 (Ioffe & Szegedy, 2015). En el entrenamiento de una red neuronal, los valores de entrada a cada capa oculta pueden variar ampliamente debido a la propagación de los pesos y los sesgos a lo largo de la red. Esto puede hacer que los valores se vuelvan demasiado grandes o demasiado pequeños, lo que dificulta un entrenamiento efectivo. La normalización por lotes se introduce para abordar este problema, de manera que tiene lugar una etapa de normalización para ajustar las medias y las varianzas de las entradas a cada capa, para

Capítulo 1. Introducción

cada *mini-batch* durante el entrenamiento. La fórmula de la normalización por lotes se puede expresar mediante:

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad (14)$$

Siendo \hat{x} la entrada a la capa normalizada, x la entrada original, μ la media del lote, σ^2 la varianza del lote y ϵ una pequeña constante muy cercana a cero incluida con el fin de evitar la división por cero.

Mediante esta técnica, se consigue la reducción del problema de desvanecimiento y explosión de gradiente, ya que mantiene las activaciones en rangos más estables, permite el uso de tasas de aprendizaje más altas, lo que acelera la convergencia del entrenamiento, y se incluye una forma de regularización, reduciendo el riesgo de sobreajuste de la red.

En nuestra arquitectura final, utilizamos tanto *Batch Normalization* como *Dropout* a la salida del bloque residual, como técnica para controlar la complejidad de la red y prevenir el sobreajuste, y también a la entrada y la salida de la arquitectura final, pues al ser una arquitectura compleja y no contar con un conjunto de datos relativamente grande, nos interesa evitar en gran medida el *overfitting*.

1.3.2.2.6 Métricas de evaluación

La evaluación de las segmentaciones semánticas puede llegar a ser compleja ya que es necesario medir tanto la precisión de la clasificación como el que la localización haya sido correcta. El objetivo es medir la similitud entre la segmentación predicha y la etiquetada manualmente. Durante los últimos 30 años, se han desarrollado una gran variedad de métricas de evaluación. Sin embargo, sólo unas cuantas han demostrado ser lo suficientemente apropiadas y se ha estandarizado su uso (Taha & Hanbury, 2015).

Las métricas que se definen a continuación se calculan a partir de la matriz de confusión resultante del modelo entrenado, la cual permite visualizar y analizar el rendimiento del modelo al clasificar diferentes clases o categorías en un conjunto de datos.

		Predicción	
		Positivos	Negativos
observación	Positivos	Verdaderos positivos (VP)/ True positives (TP)	Falsos negativos (FN)/ False negative (FN)
	Negativos	Falsos positivos (FP)/ False Positive (FP)	Verdaderos Negativos (VN)/ True negatives (TN)

Figura 16. Componentes de matriz de confusión.

Capítulo 1. Introducción

La matriz de confusión tiene cuatro componentes:

- Verdaderos positivos (VP o *True Positive, TP*): representa el número de casos en los que el modelo clasificó correctamente una imagen perteneciente a la categoría.
- Verdaderos negativos (VN o *True Negative, TN*): representa el número de casos en los que el modelo clasificó correctamente una imagen no perteneciente a la categoría.
- Falsos positivos (FP): representa el número de casos en los que el modelo clasificó incorrectamente una imagen como perteneciente a una clase, cuando en realidad no lo era.
- Falsos negativos (FN): representa el número de casos en los que el modelo clasificó incorrectamente una imagen como no perteneciente a una clase, cuando en realidad sí lo era.

El comportamiento de estas métricas se puede ver en la Figura 17 reflejando varios posibles escenarios.

Sensibilidad y especificidad

Especialmente en medicina, la especificidad y la sensibilidad son métricas estándar para la evaluación del rendimiento (Popovic et al., 2007; Taha & Hanbury, 2015). Para la clasificación de los píxeles, la sensibilidad, también conocida como *Recall* o tasa de verdaderos positivos, se centra en la capacidad de detección de los VP, mientras que la especificidad, también llamada tasa de verdaderos negativos, evalúa las capacidades para detectar correctamente las clases verdaderamente negativas (VN), como la clase perteneciente al fondo. En el ámbito de la segmentación de imagen médica, la sensibilidad es una métrica válida y popular, pero aun así menos sensitiva que las basadas en F-score para hacer una evaluación y comparación de métodos más precisa. Sin embargo, la especificidad nos puede llevar a una métrica de segmentación incorrecta si no se sabe entender de manera correcta. En tareas de segmentación de imagen médica, la especificidad indica la capacidad del modelo para detectar la clase perteneciente al fondo de una imagen. Debido a la gran cantidad de píxeles etiquetados como fondo a comparación de la ROI, los rangos de especificidad cercanos a 1 son el estándar y los esperados. Es por ello por lo que, la especificidad, puede ser una buena métrica para asegurar la funcionalidad del modelo, pero no para su rendimiento.

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (15)$$

Capítulo 1. Introducción

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (16)$$

Precisión

La precisión es una de las métricas más conocidas, por no decir la más, en estadística (Taha & Hanbury, 2015). Se define como el número de predicciones correctas (VP y VN) de entre el número total de predicciones. Sin embargo, no se recomienda utilizar esta métrica en segmentación de imagen médica, ya que es un ámbito en el que se presenta un gran desbalanceo de clases. Debido a la importancia que se les da a los verdaderos negativos en su definición, esta métrica nos dará siempre una puntuación alta errónea que no se corresponde con la realidad. Por lo tanto, la precisión se trata de una métrica engañosa, no apropiada para el ámbito médico.

$$\text{Precisión} = \frac{VP + VN}{VP + VN + FN + FP} \quad (17)$$

Métricas basadas en F-score

Las métricas basadas en F-score son unas de las más extendidas para medir el rendimiento en visión artificial, así como en la segmentación de imagen médica (Asgari Taghanaki et al., 2019; X. Liu et al., 2021; Popovic et al., 2007; Taha & Hanbury, 2015). Se calcula a partir de la sensibilidad y precisión de una predicción de manera que puntúa la coincidencia entre la segmentación predicha y la manual. Al incluir la precisión, se penaliza los FP, lo que resulta importante en conjuntos de datos altamente desbalanceados como los que nos solemos encontrar en imagen médica (Popovic et al., 2007; Taha & Hanbury, 2015). Entre estas métricas hay dos popularmente utilizadas:

- La IoU, también conocida como índice Jaccard o coeficiente de similitud Jaccard.

$$\text{IoU} = \frac{VP}{VP + FP + FN} \quad (18)$$

- El coeficiente de similitud Dice (DSC de sus siglas en inglés), también conocido como F1 score o índice Sørensen-Dice. El DSC se define como la media entre la sensibilidad y precisión y es una métrica ampliamente utilizada que calcula el grado de superposición entre la máscara de segmentación predicha y la máscara de referencia.

$$\text{DSC} = \frac{2VP}{2VP + FP + FN} \quad (19)$$

Capítulo 1. Introducción

Tanto el valor del DSC como el de IoU varía entre 0 y 1, donde 1 indica una segmentación perfecta y 0 indica una segmentación completamente incorrecta. La diferencia entre las dos métricas es que la IoU penaliza las sub- o sobre-segmentación más que el DSC. Aun así, ambas métricas son buenas, siendo el DSC la más usada en la mayoría de las publicaciones científicas para la evaluación de las segmentaciones en imagen médica (X. Liu et al., 2021; Popovic et al., 2007; Taha & Hanbury, 2015).

Señalar que, en nuestro estudio, como métrica de entrenamiento, empleamos el coeficiente de correlación de Matthews (Chicco & Jurman, n.d.) (*Matthews Correlation Coefficient, MCC*), que se encuentra definido en la sección 1.3.4.3 del presente capítulo.

Visualización de la muestra

Además de buscar una métrica de evaluación, aunque sea obvio, siempre se recomienda inspeccionar visualmente los resultados que hemos obtenido de la segmentación. Comparar visualmente las segmentaciones manuales y las predichas nos permite la mejor de las evaluaciones. Esta visualización puede ser binaria o utilizando distintos colores y niveles de transparencia basándonos en las clases de los píxeles de la imagen original. La ventaja más fuerte de la visualización es evitamos cualquier influencia estadística, sobreestimación o poder predictivo ocasionado por una métrica inapropiada o mal calculada.

Médula espinal – Pequeña región de interés: 0.2% de la imagen

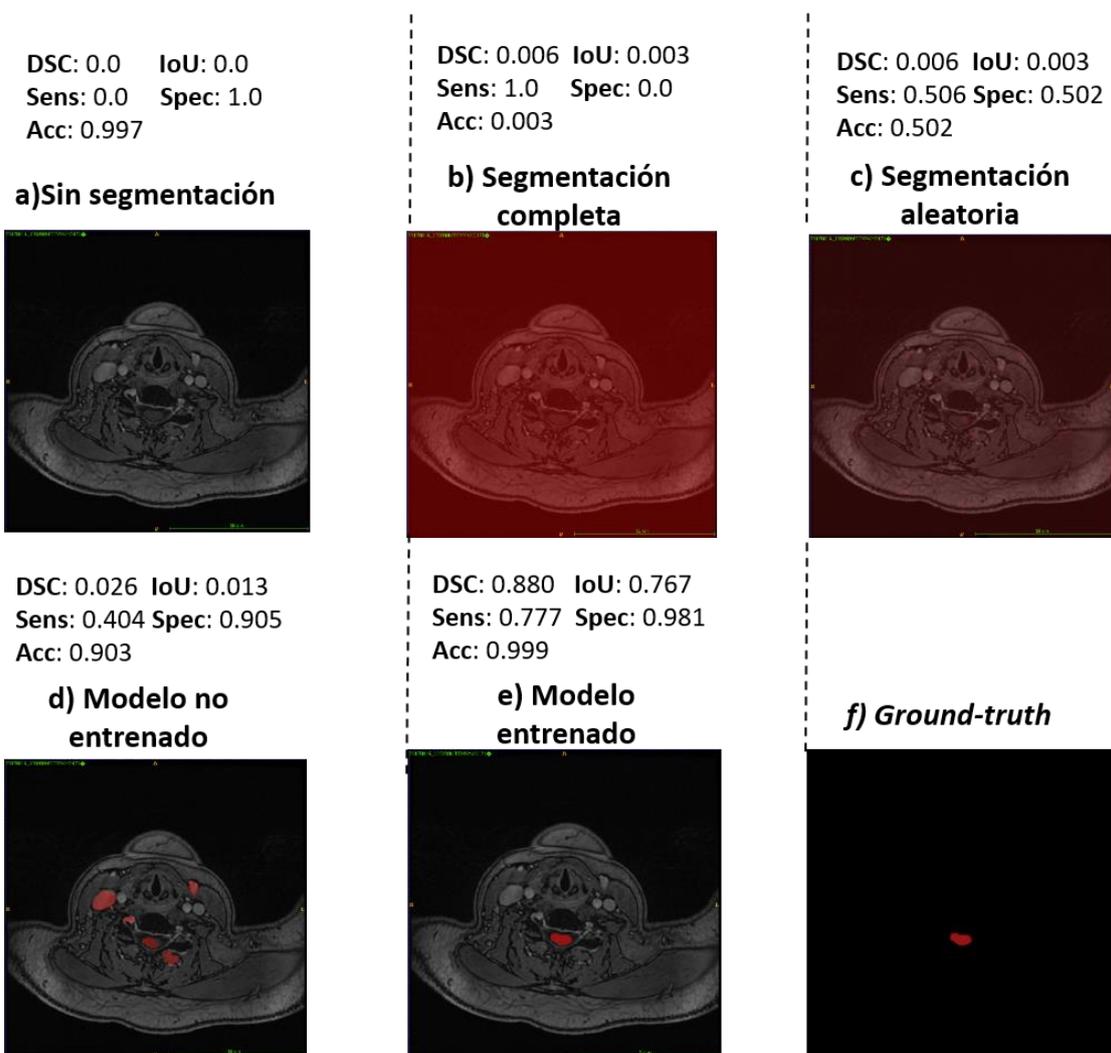


Figura 17. Demostración del comportamiento de las métricas en términos de distintos tamaños de ROI en comparación a la imagen completa. Podemos observar varios escenarios en el ámbito de la segmentación de imagen médica. Escenarios: a) Sin segmentar (no hay píxeles etiquetados en la ROI en color rojo), b) segmentación completa (todos los píxeles etiquetados como ROI, todos en color rojo), c) segmentación aleatoria (etiquetado de la ROI completamente aleatorio), d) sin entrenar (después de pocas épocas de entrenamiento), e) entrenado (modelo completamente ajustado, solo en color rojo la lesión buscada), f) *Ground-truth*.

1.3.2.3 Relevancia clínica de las técnicas de segmentación de imagen

Los algoritmos de segmentación de imagen médica automáticos son herramientas muy importantes en el ámbito clínico debido a su capacidad para identificar y delimitar estructuras anatómicas o patológicas en imágenes médicas, como RM, tomografías computarizadas (TC) y ecografías.

La segmentación de imágenes médicas es un paso fundamental en el procesamiento de imágenes, ya que proporciona información cuantitativa y cualitativa sobre las estructuras de interés presentes en la imagen. Estos algoritmos permiten extraer y cuantificar características

Capítulo 1. Introducción

específicas, como volúmenes, áreas, densidades, texturas y características morfológicas, que son cruciales para el diagnóstico, pronóstico y planificación de tratamientos en diversas especialidades médicas, como radiología, oncología, neurología y cardiología, entre otras.

La relevancia clínica de los algoritmos de segmentación automática se puede resumir en los siguientes puntos:

1. **Diagnóstico y evaluación de enfermedades:** La segmentación precisa de las estructuras de interés permite a los médicos identificar y analizar anomalías, como tumores, lesiones, aneurismas, áreas de isquemia, deformidades y malformaciones congénitas. Esto ayuda en el diagnóstico temprano, la clasificación y la evaluación de la gravedad de las enfermedades, lo que a su vez puede influir en el plan de tratamiento y en el seguimiento de la progresión de la enfermedad.
2. **Planificación de tratamientos:** La segmentación automática facilita la planificación precisa de los tratamientos médicos. Por ejemplo, en radioterapia, los algoritmos de segmentación permiten definir con precisión el volumen objetivo a irradiar y los órganos sanos a preservar, mejorando así la eficacia del tratamiento y minimizando los efectos secundarios en los tejidos sanos circundantes.
3. **Guía quirúrgica:** En cirugía, la segmentación automática puede proporcionar información preoperatoria vital sobre la ubicación y la relación de las estructuras anatómicas, lo que ayuda a los cirujanos a planificar y realizar procedimientos quirúrgicos más seguros y precisos. Además, puede permitir la integración de imágenes preoperatorias con imágenes intraoperatorias en tiempo real, lo que facilita la navegación y la orientación durante la cirugía.
4. **Monitorización y seguimiento:** Los algoritmos de segmentación automática pueden usarse para cuantificar y rastrear cambios en las estructuras anatómicas o patológicas a lo largo del tiempo. Esto es especialmente útil para evaluar la eficacia de los tratamientos, monitorear la progresión de la enfermedad y realizar un seguimiento de la respuesta del paciente.

Los algoritmos de segmentación automática son, pues, herramientas valiosas en la práctica clínica, ya que permiten una evaluación objetiva y cuantitativa de las imágenes médicas, facilitando el diagnóstico preciso, la planificación de tratamientos, la guía quirúrgica y el seguimiento de los pacientes. Estos algoritmos ayudan a los médicos a tomar decisiones informadas y mejorar la atención médica en diferentes especialidades.

Capítulo 1. Introducción

1.3.2.4 Principales desafíos de la segmentación en imagen médica

La segmentación automática en imagen médica mediante *Deep Learning* es una aplicación altamente útil y valiosa, la cual, durante su desarrollo y para su progreso, se encuentra con algunas dificultades comunes que tenemos que superar a la hora de crear un modelo preciso y con un buen rendimiento. Un problema común que nos encontramos a la hora de aplicar técnicas de *Deep Learning* en imagen médica, es el de la escasez de datos etiquetados, pues resulta una tarea costosa tanto en tiempo como en esfuerzo de profesionales. Las técnicas de *Deep Learning* han incrementado notablemente la precisión de las segmentaciones gracias a su capacidad de manejar diferentes y complejos escenarios, pero para ganar esta capacidad, las redes requieren esta gran cantidad de muestras etiquetadas que mencionamos. Por ejemplo, a la hora de entrenar un sistema de aprendizaje profundo que segmente imagen de RM en 3D, el etiquetado imagen a imagen de un conjunto de datos lo suficientemente grande para conseguir un buen entrenamiento, involucraría una gran cantidad de tiempo y esfuerzo. Conseguir tal conjunto de datos de casos etiquetados en el campo de la imagen médica es normalmente una tarea complicada dado la laboriosidad y complejidad que implica el realizar el etiquetado de imágenes nuevas, siendo además necesaria la intervención de expertos en la materia. Para solventar esta problemática existen diferentes soluciones como la técnica de aumento de datos, la cual consiste en aplicar un conjunto de transformaciones a la imagen como rotación, volteo, reflejo o la adición de una pequeña cantidad de ruido aleatorio. Otra posible solución es la transferencia de aprendizaje (*Transfer Learning*) de modelos que ya desempeñan un buen rendimiento en la misma área (o incluso en otra diferente), de manera que, el nuevo modelo sea capaz de reconocer y utilizar conocimientos aprendidos para otras tareas. En otras ocasiones, el etiquetado completo de los datos no siempre es posible, por lo que sólo disponemos de datos etiquetados de manera dispersa. Para ser capaces de aprender de este tipo de conjunto de datos, se utilizan funciones de pérdidas ponderadas que nos ayudan a aprender sólo de los píxeles etiquetados. Finalmente, y también en línea a solventar el problema en el ámbito de la escasez de datos, mencionamos las redes adversarias generativas (*Generative Adversarial networks, GANs*), muy populares dada su habilidad para imitar las distribuciones de las imágenes muestreadas. Entre otras aplicaciones, las GANs pueden ser utilizadas para generar imágenes sintéticas realistas para poder llegar a entrenar modelos más profundos y obtener mejores resultados.

Otro desafío común en el procesamiento de imagen médica, el cual, a su vez, enfrentamos en nuestro propósito de segmentación de médula espinal cervical, es el del desbalanceo de clases, donde la región de interés ocupa una parte muy pequeña de la imagen, por ejemplo, el área

Capítulo 1. Introducción

tumoral significativamente más pequeña que el resto de tejidos sanos o en el caso que nos aplica, la porción de médula espinal pequeña en comparación al resto de la imagen. En estos casos, la mayoría de los parches extraídos pertenecerán a regiones que no son de interés y el entrenamiento estará sesgado. Una solución común es re-ponderar las muestras, de manera que, durante el entrenamiento, se aplica un peso mayor a las porciones de imagen pertenecientes a la región de interés. Otra forma de mitigar este problema es utilizar funciones de pérdidas muestreadas en las que el cálculo de las pérdidas no es efectuado en toda la imagen, sino que sólo es calculado en algunos píxeles aleatorios. En problemas como este, resulta más sencillo centrarse en dichas pequeñas regiones cambiando la función de pérdidas que cambiando la estructura de la red. Aun así, el diseño de la función de pérdidas debe de ser específico para cada tarea, por lo que primero hay que analizar con cuidado las necesidades y luego emplear la función de pérdidas en base a ello.

La imagen médica, en ocasiones, puede ir acompañada por artefactos de ruido, lo que, junto a la problemática de lo costoso de la obtención de datos etiquetados, nos lleva a considerar la utilización de modelos de *Deep Learning* pre-entrenados (*Transfer Learning*) en imágenes comunes. Mediante *Transfer Learning* podremos usar los conocimientos ya adquiridos por un modelo para aprender nuevos conocimientos, y procuraremos encontrar similitudes entre este conocimiento existente y el nuevo, las cuales pueden pertenecer a diferentes niveles de abstracción. Estas similitudes nos permitirán utilizar los parámetros del modelo (conocimiento aprendido por el modelo) con el nuevo modelo, de manera que se agiliza la eficiencia del aprendizaje y se solventa a la vez el problema de la escasez de datos etiquetados, tal como explicamos anteriormente.

1.3.3 Arquitectura empleada: Residual Attention-UNet

La red U-Net, explicada a continuación en la sección 1.3.3.1.4, es un tipo de arquitectura de red neuronal convolucional ampliamente utilizado en tareas de segmentación de imágenes. Se compone de una estructura de codificador-decodificador que permite capturar características de diferentes niveles de abstracción en la imagen de entrada. Sin embargo, U-Net puede tener dificultades para enfocarse en las regiones más relevantes de la imagen durante el proceso de segmentación.

Para abordar esta limitación, RA-UNet incorpora mecanismos de atención residual en la arquitectura U-Net. La atención residual se refiere a la capacidad de la red para asignar

Capítulo 1. Introducción

importancia a diferentes partes de la imagen y enfocarse en las regiones más relevantes mediante la introducción de bloques de atención residual en la estructura U-Net.

Los bloques de atención residual se utilizan para calcular mapas de atención, que indican la importancia relativa de cada píxel en función de su relación con los demás píxeles de la imagen. Estos mapas de atención se combinan con las características extraídas en diferentes niveles de la red para resaltar las regiones más relevantes y suprimir las regiones menos importantes.

RA-UNet también utiliza conexiones residuales en la estructura U-Net para mejorar la propagación de la información y facilitar el flujo de gradiente durante el entrenamiento. Estas conexiones permiten que la información se transmita directamente de las capas de codificación a las capas de decodificación, lo que ayuda a preservar los detalles de baja y alta resolución.

En definitiva, RA-UNet es una mejora de U-Net mediante la incorporación de bloques de atención residual y conexiones residuales. Estos mecanismos mejoran la capacidad del modelo para enfocarse en regiones relevantes de la imagen, en nuestro caso, la médula espinal, y preservar los detalles durante el proceso de segmentación.

1.3.3.1.1 Mecanismo de aprendizaje residual

El mecanismo de aprendizaje residual en las redes neuronales, también conocido como "*residual learning*" en inglés, se introdujo en el marco de las Redes Neuronales Profundas (DNNs) por el artículo "*Deep Residual Learning for Image Recognition*" de Kaiming He, et al., en 2015. Este enfoque revolucionó el campo de la visión por computadora y se ha convertido en una técnica fundamental en el diseño de arquitecturas de redes neuronales.

Anteriormente, ya hemos comentado la importancia de la profundidad en las redes neuronales convolucionales, pero muchas veces surge la siguiente pregunta: ¿es tan fácil mejorar las redes con el simple hecho de añadir más capas? En las DNNs tradicionales, cada capa de la red neuronal intenta aprender una representación de las características de entrada. Sin embargo, a medida que estamos en una capa más profunda, puede volverse difícil para las capas posteriores aprender directamente una representación significativa de las características originales. Este problema se conoce como el problema de degradación de la red, donde agregar más capas puede llevar a un empeoramiento del rendimiento en lugar de mejorarlo.

El mecanismo de aprendizaje residual aborda este problema introduciendo conexiones de "salto" o "*skip connections*" que permiten que las capas de la red aprendan las diferencias o residuos entre las características de entrada y la salida esperada. En lugar de intentar aprender una representación completa de las características, las capas de la red se enfocan en aprender

Capítulo 1. Introducción

solo los cambios o residuos necesarios para transformar las características de entrada en la salida deseada. Estas conexiones de salto permiten que la información fluya directamente a través de la red sin pasar por tantas capas, evitando así el problema de degradación.

Matemáticamente, supongamos que tenemos una función de activación no lineal en una capa de la red neuronal, que representa la transformación de las características de entrada X a las características de salida H . En lugar de aprender directamente $H(x)$, la capa aprende los residuos $F(x) := H(x) - x$. Luego, la salida de la capa se convierte en $F(x) + x$. Esta adición de la entrada original x a los residuos aprendidos $F(x)$ se denomina conexión residual. El uso de estas conexiones residuales permite que la red aprenda las diferencias o cambios necesarios en lugar de tener que aprender toda la transformación de una vez.

La expresión del nuevo mapeo $F(x) + x$ se puede llevar a cabo mediante redes neuronales feedforward (de avance) mediante las conexiones de "salto" o "*skip connections*" anteriormente mencionadas. En nuestro caso, las conexiones de salto simplemente llevan a cabo el mapeo de identidad y sus salidas se añaden a las salidas de las capas que están apiladas, no añaden ni parámetros extra ni complejidad computacional.

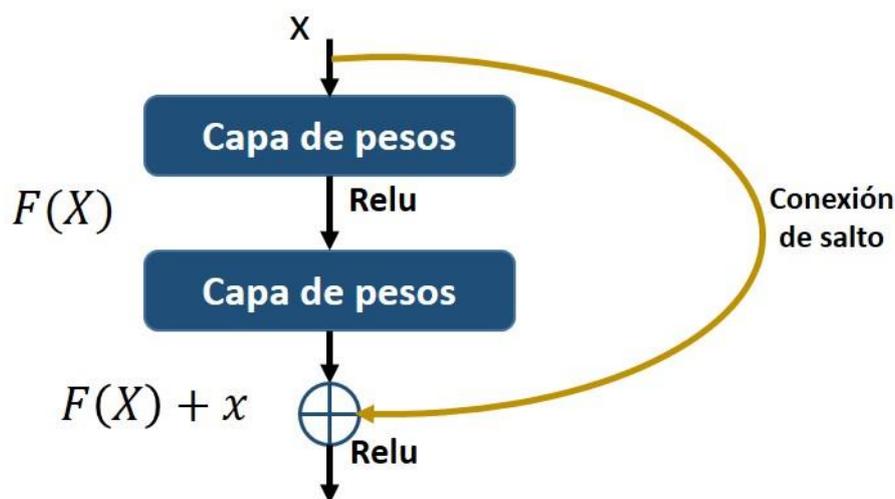


Figura 18. Diagrama representativo de la conexión de salto.

En nuestra arquitectura, los bloques residuales aparecen apilados en todas las capas excepto en la primera y la última para potenciar la capacidad de las redes neuronales profundas. Estos bloques residuales solucionan el problema del desvanecimiento de gradiente a nivel estructural de la red neuronal, usando mapeo de identidad como conexiones de salto. Las unidades residuales propagan características directamente desde la primera convolución a la última y como resultado, mejora el rendimiento del modelo.

Capítulo 1. Introducción

Es importante señalar que un modelo con conexiones residuales tiene el mismo número de parámetros que un modelo sin ellas, pero es más fácil de entrenar. La razón es que el gradiente puede ir directamente desde la salida hasta las capas iniciales, como se puede ver en la Figura 19.

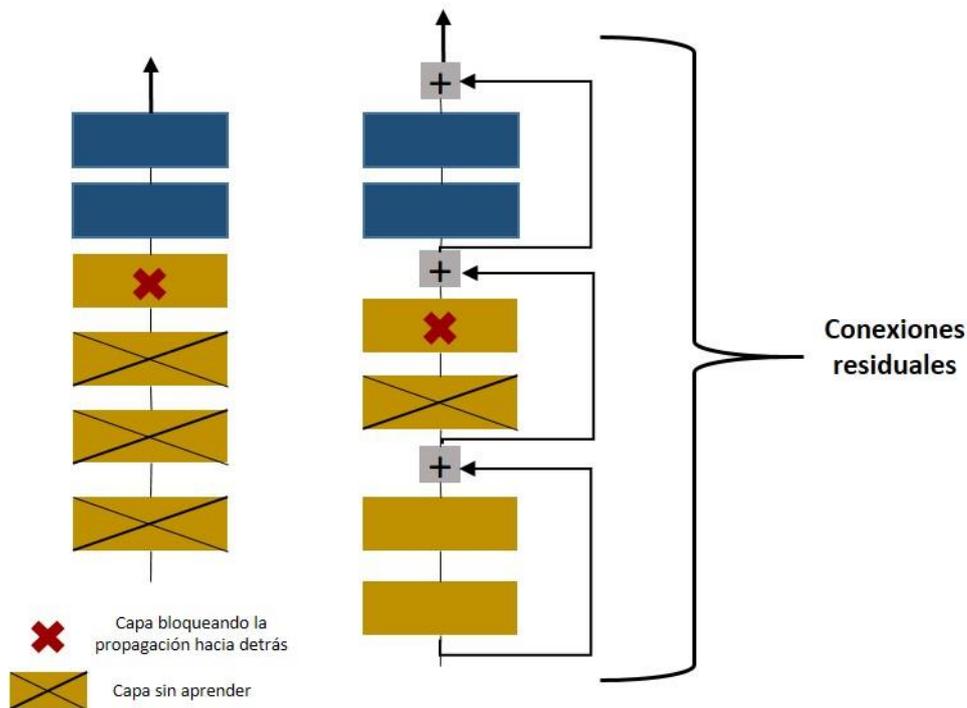


Figura 19. Visualización de por qué añadir conexiones residuales puede ayudar cuando se entrena un modelo muy profundo.

Estas conexiones directas tampoco introducen complejidad computacional, lo que es importante en la práctica, pero también cuando nos interesa comparar el rendimiento con redes poco profundas, ya que podemos hacerlo de forma imparcial, pues ambas arquitecturas tendrán el mismo número de parámetros, profundidad, ancho y coste computacional (excepto por la insignificante adición de la suma de elementos).

El aprendizaje residual tiene varias ventajas. Primero, facilita el entrenamiento de redes neuronales más profundas al resolver el problema de degradación de la red. Segundo, ayuda a evitar el estancamiento en el aprendizaje y la desaparición del gradiente, lo que permite una propagación más suave del gradiente a través de las capas de la red. Además, las conexiones residuales también pueden permitir la reutilización de capas previamente entrenadas en nuevas arquitecturas de red, lo que facilita el desarrollo de modelos más complejos.

En definitiva, el mecanismo de aprendizaje residual en las redes neuronales es una técnica que utiliza conexiones residuales para aprender las diferencias o cambios necesarios en lugar de aprender directamente una representación completa. Este enfoque ha demostrado ser

Capítulo 1. Introducción

altamente efectivo en el entrenamiento de redes neuronales profundas, permitiendo la construcción de modelos más profundos, más rápidos y con un mejor rendimiento en una variedad de tareas de visión por computadora y aprendizaje automático.

1.3.3.1.2 Mecanismo de atención

El mecanismo de atención en redes neuronales es una técnica utilizada para mejorar el rendimiento de modelos de aprendizaje automático en tareas de procesamiento de secuencias, como la traducción automática, el reconocimiento de voz y la generación de texto. Esta técnica se inspira en el sistema de atención humana, que permite a los seres humanos enfocar su atención en partes específicas de una entrada para procesarla de manera más efectiva. Inspirados por estos mecanismos de atención visual, los investigadores han intentado diseñar modelos para imitar el proceso de percepción visual del ser humano, de forma que se pueda modelizar la distribución de atención cuando se observan imágenes.

En un nivel alto, el mecanismo de atención en redes neuronales funciona al asignar pesos a diferentes partes de una secuencia de entrada en función de su relevancia para la tarea en cuestión, en nuestro caso, detectar la región perteneciente a médula espinal. Estos pesos se utilizan para calcular una representación ponderada de la secuencia, donde las partes más relevantes tienen un peso mayor y, por lo tanto, una mayor contribución al resultado final.

El mecanismo de atención se compone generalmente de tres pasos principales:

1. Cálculo de la similitud: Se calcula una medida de similitud entre un vector de consulta y cada uno de los vectores de la secuencia de entrada. Esto se puede lograr utilizando diferentes métodos, como el producto escalar o el producto interno ponderado por una matriz de pesos aprendida.
2. Cálculo de los pesos de atención: Las medidas de similitud obtenidas en el paso anterior se transforman en pesos de atención a través de una función de softmax. La función softmax asegura que los pesos sumen uno y asigna una probabilidad a cada elemento de la secuencia de entrada.
3. Cálculo de la representación ponderada: Se calcula una representación ponderada de la secuencia de entrada multiplicando los pesos de atención por los vectores de la secuencia. Esto enfatiza las partes más relevantes de la secuencia y reduce la importancia de las partes menos relevantes.

El resultado final de este proceso de atención ponderada es una representación contextualizada de la secuencia de entrada, que se utiliza como entrada para la siguiente capa del modelo de aprendizaje automático. Esto permite que el modelo se centre en partes específicas de la

Capítulo 1. Introducción

secuencia en cada paso, mejorando así su capacidad para capturar dependencias a largo plazo y realizar predicciones más precisas.

Se ha demostrado que los mecanismos de atención pueden mejorar el rendimiento de los modelos. Utilizando como ejemplo el dominio de la visión artificial, la mayoría de las investigaciones que combinan Deep learning y mecanismos de atención visual se centran en el uso de la máscara, por lo que, mediante el entrenamiento y aprendizaje, las redes neuronales profundas pueden aprender cuáles son las áreas donde merece la pena poner atención en cada imagen.

1.3.3.1.3 Mecanismo de aprendizaje de atención residual

Durante nuestro trabajo, se utilizó un mecanismo de aprendizaje de atención residual en el que combinamos los dos métodos anteriormente expuestos: la atención y las conexiones residuales. La atención se refiere a la capacidad de una red neuronal de enfocarse en partes específicas de una entrada, lo que permite capturar relaciones y dependencias entre diferentes elementos. Las conexiones residuales, por otro lado, son un tipo de arquitectura de red neuronal que facilita el flujo de información a través de capas mediante la adición de atajos o conexiones directas.

Este mecanismo, divide el módulo de atención en una rama “fuerte” (*trunk branch*) y una “débil” (*soft branch*), donde la fuerte es utilizada para procesar las características originales y la débil para construir el mapeo de identidad. La salida OA del módulo de atención en el aprendizaje de atención residual es:

$$OA_{i,c}(x) = (1 + S_{i,c}(x))F_{i,c}(x) \quad (20)$$

Donde $S(x)$ tiene valores entre $[0, 1]$. Si $S(x)$ está cerca de 0, $OA(x)$ se aproximará al mapa de características original $F(x)$. La rama débil $S(x)$, la cual se encarga de construir el mapeo de identidad, realzar las características de interés y reducir el ruido de la rama principal, tiene el papel más importante en el mecanismo de atención residual. En nuestra aplicación, esta rama débil es la que se centra en las características procedentes de la región de la médula espinal.

En términos generales, este mecanismo de aprendizaje de atención residual puede mantener la información de las características originales a través de la rama fuerte y centrarse en las características procedentes de la médula espinal mediante la rama débil, a la vez que nos permite diseñar una arquitectura profunda.

La rama de “débil” tiene una estructura codificador-decodificador ya típica en aplicaciones de segmentación de imagen médica. En el mecanismo de atención residual, esta estructura está

Capítulo 1. Introducción

diseñada para realzar las características de interés y eliminar el ruido procedente de la rama principal. El codificador en la rama “débil” contiene una operación de max-pooling, un bloque residual y otro bloque residual de largo alcance conectado al decodificador correspondiente, donde tiene lugar una suma de elementos seguida de un bloque residual y una operación de sobremuestreo. Tras la parte del codificador y el decodificador, la rama débil cuenta con dos capas convolucionales y una capa Sigmoid para normalizar la salida. En la Figura 20 podemos hacernos una idea esquemática de ello:

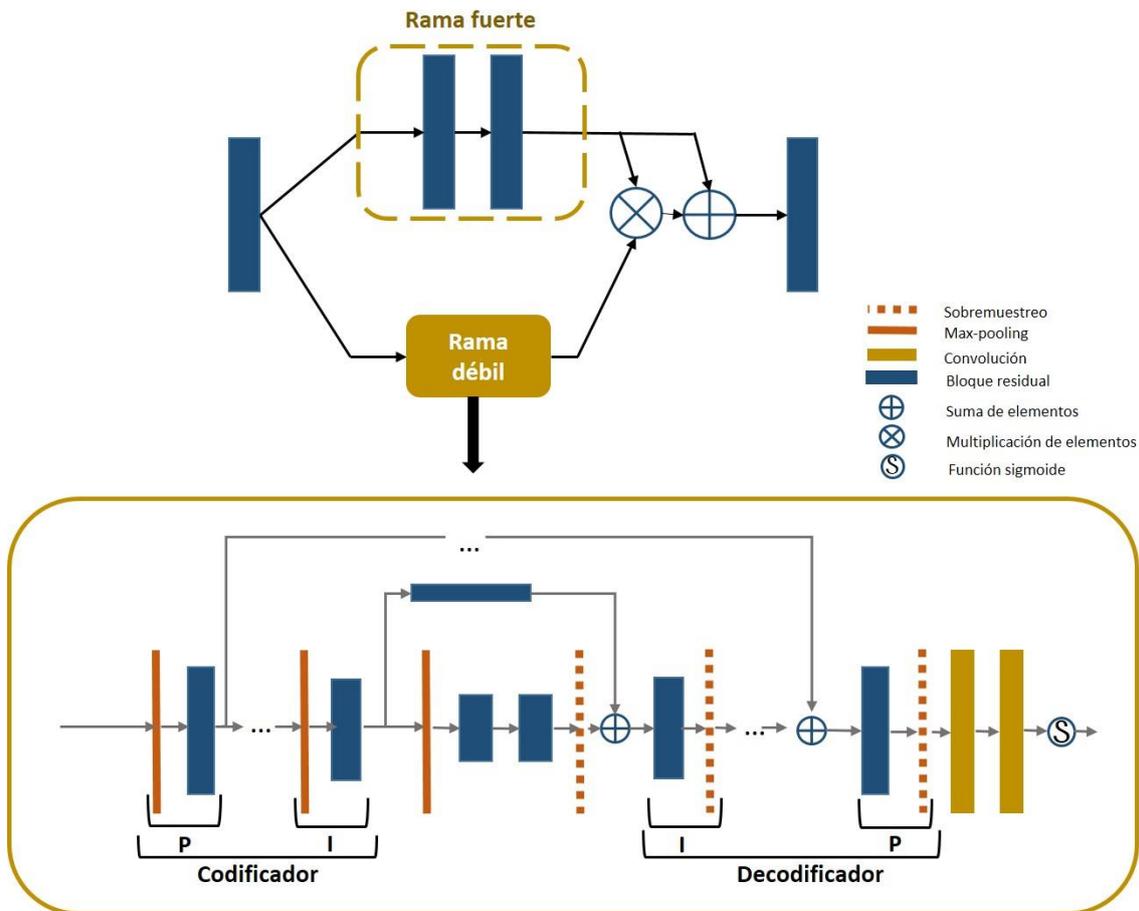


Figura 20. Arquitectura del módulo de atención. En la parte superior, el módulo de atención compuesto por la rama fuerte y la rama débil. La rama fuerte aprende características originales de la imagen mientras que la rama débil se centra en reducir ruido y mejorar las características que señala la máscara. En la parte inferior, la rama débil compuesta por una agrupación de bloques codificador-decodificador. La P indica la profundidad de las conexiones de salto.

La combinación de la atención y las conexiones residuales en el aprendizaje de atención residual proporciona varias ventajas: permite que la red se centre en características específicas y relaciones importantes, lo que mejora la capacidad de la red para capturar patrones complejos, por otra parte, también ayuda a aliviar el problema de degradación de la red, que ocurre cuando agregar más capas a la red no mejora su rendimiento. Las conexiones residuales permiten que

Capítulo 1. Introducción

la información fluya más fácilmente a través de las capas, evitando que se degraden los gradientes durante el entrenamiento

1.3.3.1.4 Red U-Net

La red U-Net es una CNN que se utiliza comúnmente en tareas de segmentación y procesamiento de imágenes médicas. Fue propuesta por primera vez en 2015 por Olaf Ronneberger, Philipp Fischer y Thomas Brox en un artículo titulado "*U-Net: Convolutional Networks for Biomedical Image Segmentation*" (Ronneberger et al., 2015).

La U-Net se diseñó específicamente para abordar el desafío de la segmentación semántica en imágenes, donde el objetivo es clasificar cada píxel en la imagen en una categoría particular. Esta tarea es esencial en aplicaciones médicas, como la segmentación de órganos o estructuras específicas en imágenes de RM o TC.

La arquitectura U-Net se basa en una estructura de codificador-decodificador. El codificador, también conocido como la ruta de compresión (parte izquierda), sigue una arquitectura típica de red convolucional y captura las características de la imagen a diferentes niveles de abstracción a través de capas de convolución y agrupación (*pooling*). Esta ruta de compresión está compuesta por la aplicación de varios pares de convoluciones 3x3 (*unpadded convolutions*), cada una seguida de una ReLU y una operación de *max pooling* en parches de 2x2 para la reducción de dimensionalidad. A medida que se profundiza en la ruta de compresión, las características espaciales se reducen mientras que las características semánticas se incrementan.

La ruta de decodificación, también conocida como la ruta de expansión (parte derecha), reconstruye una representación espacial precisa de la imagen segmentada a partir de las características extraídas por el codificador. Cada capa de la ruta de expansión está conectada a una capa correspondiente en la ruta de compresión, y utiliza técnicas de convolución transpuesta o *upsampling* para aumentar el tamaño espacial de las características. Cada nivel de la ruta de expansión realiza un aumento de dimensionalidad del mapa de características junto a una convolución 2x2 (*up-convolution*) que reduce a la mitad el número de características espaciales, una concatenación con el mapa de características recortado correspondiente de la ruta de compresión y dos convoluciones 3x3, cada una seguida de una ReLU. El recorte es necesario debido a la pérdida de los píxeles de los bordes de la imagen en cada convolución. En la capa final se utiliza una convolución 1x1 para mapear cada vector de características de 64 componentes al número de clases deseado. Finalmente, la arquitectura tendrá en total 23 capas de convolución.

Capítulo 1. Introducción

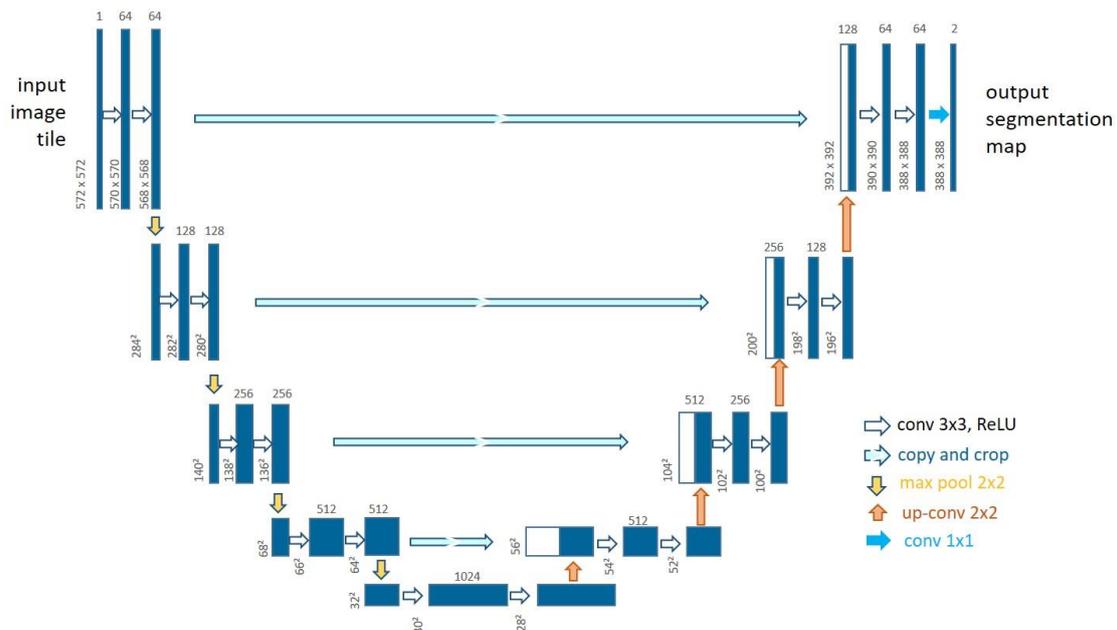


Figura 21. Arquitectura U-Net (representación de ejemplo para la menor resolución, 32x32 píxeles). Cada caja azul corresponde a un mapa de características multicanal. El número de canales está indicado encima de cada caja y el tamaño x-y en la parte inferior izquierda. Las cajas blancas representan mapas de características copiados y las flechas representan las diferentes operaciones.

Las conexiones laterales entre las capas correspondientes en la ruta de compresión y la ruta de expansión permiten que la información de alta resolución de las capas de la ruta de compresión se transmita a las capas de la ruta de expansión, lo que ayuda a mejorar la localización precisa de los objetos en la imagen segmentada.

Una característica distintiva de la red U-Net es que utiliza mapas de características (*feature maps*) simétricos en cada nivel de la ruta de compresión y de expansión. Esto significa que la información espacial y semántica se conserva durante el proceso de codificación y decodificación, lo que facilita la generación de mapas segmentados de alta calidad.

La red U-Net es una arquitectura de red neuronal convolucional que ha demostrado ser efectiva en tareas de segmentación de imágenes médicas. Su estructura de codificador-decodificador, junto con las conexiones laterales y el uso de mapas de características simétricos, permite una segmentación precisa y detallada de objetos en imágenes, lo que la convierte en una opción popular para aplicaciones de investigación y práctica médica avanzada.

1.3.3.1.5 Arquitectura Residual Attention-Aware 2D U-Net

Con el propósito de solventar una de las problemáticas encontradas en nuestro objetivo, que es el gran desbalanceo de clases al que nos enfrentamos (pequeña proporción de médula espinal con respecto al resto de la imagen) en nuestro dataset, y dada su habilidad para mejorar el

Capítulo 1. Introducción

rendimiento de los modelos, elegimos utilizar las ventajas de la arquitectura *Residual Attention-Aware U-Net*.

En la arquitectura de redes residuales, se introducen conexiones directas conocidas como conexiones residuales, que permiten que la información fluya directamente de una capa a otra, saltando capas intermedias. Esto ayuda a mitigar el problema de la desaparición del gradiente y permite que la red aprenda representaciones más profundas y precisas.

La idea principal detrás de la arquitectura *Attention Aware U-Net*, como hemos mencionado anteriormente, es incorporar mecanismos de atención en las conexiones residuales. La atención permite que la red se enfoque en partes específicas de la imagen que son más relevantes para la tarea de segmentación. Esto mejora la capacidad del modelo para enfocarse en regiones importantes y filtrar el ruido o las características irrelevantes.

En la arquitectura *Attention Aware U-Net*, se agregan bloques de atención en las conexiones residuales. Estos bloques de atención se componen típicamente de capas convolucionales y operaciones de agrupamiento que permiten modelar la relación espacial entre las características. Al aprender a asignar pesos a las características en función de su importancia relativa, los bloques de atención pueden resaltar regiones importantes y suprimir las menos relevantes.

Al combinar las conexiones residuales con los bloques de atención, la arquitectura *Attention Aware U-Net* logra una mayor capacidad de aprendizaje y una mejor representación de las características relevantes en la tarea de segmentación. Esto puede conducir a una segmentación más precisa y mejorada de las imágenes médicas.

En nuestra arquitectura, esquematizada en la Figura 22, empleamos bloques residuales en cada una de las capas de la red U-Net, excepto en la primera y en la última, para solucionar el problema de desvanecimiento de gradiente utilizando mapeos de identidad como conexiones de salto. Los bloques de atención se incluyen en los dos primeros bloques residuales.

Como las unidades residuales propagan directamente las características desde la primera a la última convolución, se consigue mejorar el rendimiento del modelo.

Capítulo 1. Introducción

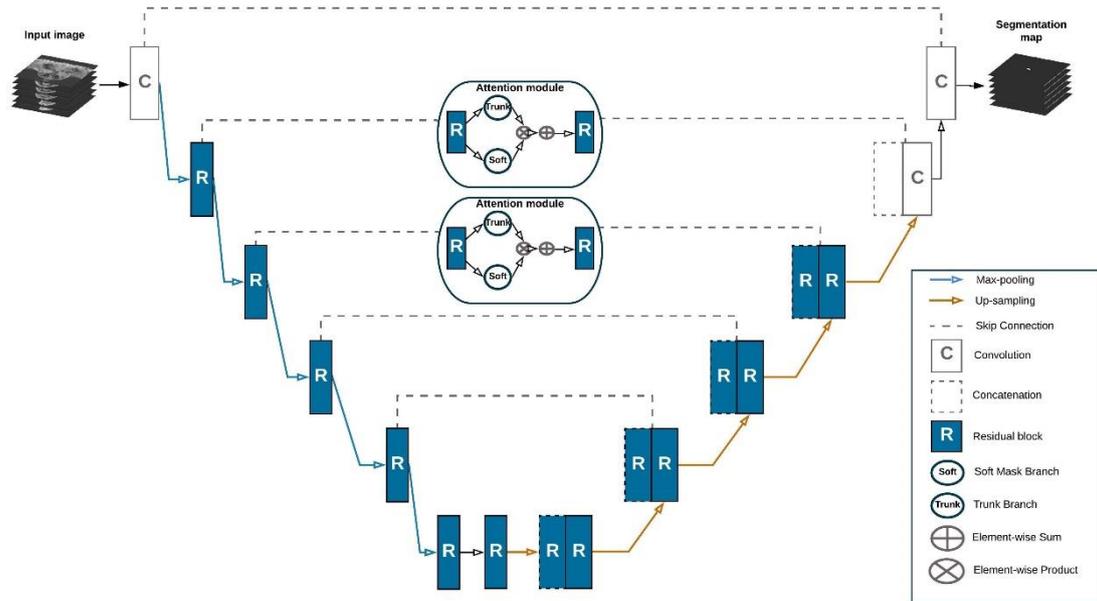


Figura 22. Arquitectura general basada en bloques residuales y mecanismos de atención. Los bloques residuales permiten el diseño de una red más profunda solventando el problema de desvanecimiento de gradiente. El mecanismo de atención resuelve el problema de desbalanceo de datos aprendiendo a focalizar el aprendizaje en las localizaciones relevantes.

El bloque residual está compuesto por tres bloques idénticos combinados cada uno por una capa de *batch normalization* (BN), una capa de activación (ReLU) y una capa convolucional. Se utiliza una conexión de mapeo de identidad convolucional para asegurar la precisión conforme la red se adentra en las capas más profundas.

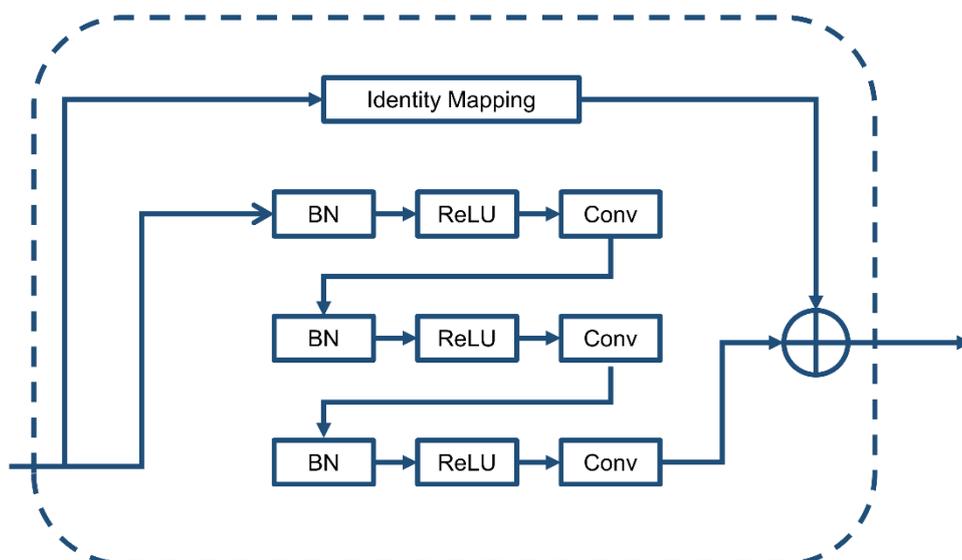


Figura 23. Diseño del bloque residual. El mapeo de identidad funciona como conexión de salto, la salida del mapeo de identidad será añadida a la salida de los bloques apilados.

Capítulo 1. Introducción

La capa de BN se emplea para cada capa convolucional y produce la media y la desviación estándar para normalizar la activación de la capa actual antes de avanzar a la siguiente. La conexión de mapeo de identidad convolucional es utilizada para asegurar la precisión cuando la red se va haciendo más profunda (He et al., 2016).

En lo que a nuestra aplicación aplica, utilizando este mecanismo de atención residual, la información original de las características se conserva a través de la rama principal y nos centramos en las localizaciones correspondientes a la médula espinal mediante la rama “débil”, mitigando así el problema de desbalanceo de clases (información perteneciente a la médula espinal vs información del resto de la imagen) y consiguiendo que nuestra arquitectura de red aprenda a segmentar los píxeles pertenecientes a la médula espinal.

1.3.4 Hiperparámetros en entrenamiento y test

Las configuraciones de hiperparámetros deben establecerse antes de comenzar el proceso de aprendizaje ya que afectan directamente al proceso de entrenamiento y la arquitectura de la red. La elección adecuada de hiperparámetros varía según la naturaleza del problema y los datos y es crucial para lograr un buen rendimiento del modelo y una convergencia efectiva durante el entrenamiento. Vemos a continuación los hiperparámetros ajustados durante nuestro método.

1.3.4.1 Funciones de pérdidas

Función de pérdidas Tversky

La función de pérdida Tversky es una métrica de evaluación que se utiliza comúnmente en problemas de aprendizaje automático y en particular en tareas de segmentación de imágenes y detección de objetos. Fue propuesta por primera vez por Tversky en 1977 y ha demostrado ser útil en diversos escenarios. Es una generalización del coeficiente DICE y el índice Jaccard.

A diferencia de otras funciones de pérdida más comunes, como la entropía cruzada o el error cuadrático medio, la función de pérdida Tversky está diseñada específicamente para abordar desequilibrios de clase y tener en cuenta tanto los falsos positivos como los falsos negativos.

La función de pérdida Tversky se basa en el índice de similitud de Tversky, que cuantifica la similitud entre dos conjuntos. En el contexto de la segmentación de imágenes, los conjuntos pueden representar las regiones segmentadas por el modelo y las regiones verdaderas definidas por las etiquetas de referencia. Esta función de pérdida calcula un puntaje de similitud que tiene en cuenta tanto los FP como los FN, utilizando dos parámetros adicionales: el coeficiente de penalización de falsos positivos (α) y el coeficiente de penalización de falsos negativos (β).

Capítulo 1. Introducción

La fórmula general de la función de pérdida Tversky es la siguiente:

$$TL = 1 - TI \quad (21)$$

Donde el índice de Tversky (TI) se define como:

$$TI = \frac{VP}{VP + \alpha \cdot FN + \beta \cdot FP} \quad (22)$$

Donde VP representa los verdaderos positivos, es decir, las regiones correctamente segmentadas por el modelo, FP representa los falsos positivos, es decir, las regiones que el modelo predijo como positivas pero que no corresponden a las regiones verdaderas y FN representa los falsos negativos, es decir, las regiones que el modelo no identificó como positivas pero que deberían haber sido detectadas.

Los parámetros α y β permiten ajustar el énfasis en los falsos negativos y falsos positivos, respectivamente. Si se da más importancia a los falsos negativos, se puede aumentar el valor de α , y si se da más importancia a los falsos positivos, se puede aumentar el valor de β .

$\alpha + \beta = 1$, en los casos en los que $\alpha = \beta = 0.5$ la expresión de TI se simplifica a la del coeficiente DICE, y si $\alpha = \beta = 1$ daría lugar a la del índice Jaccard.

Eligiendo un valor de $\alpha > \beta$, se penaliza más a los falsos negativos, lo que resulta útil en conjuntos de datos desbalanceados en los que este nivel adicional de control sobre la función de pérdidas produce mejores segmentaciones a pequeña escala que el coeficiente DICE normal.

Se trata de una medida de discrepancia utilizada en problemas de segmentación y clasificación de imágenes para penalizar tanto los falsos positivos como los falsos negativos. Es especialmente útil cuando se trata de desequilibrios de clase y se necesita tener en cuenta ambos tipos de errores. En medicina, la segmentación precisa de estructuras de interés en imágenes médicas es crucial para el diagnóstico y el tratamiento por lo que la función de pérdida Tversky puede ser utilizada para entrenar modelos de aprendizaje automático que segmenten con precisión estructuras anatómicas o regiones de interés en imágenes médicas, como tumores o en nuestro caso, la médula espinal.

Capítulo 1. Introducción

Función de pérdidas Focal Tversky Loss

La función Focal Tversky Loss (FTL) es una generación de las pérdidas Tversky mediante la cual, a través de la naturaleza no lineal de la función podemos controlar cómo se comporta ésta para diferentes valores del índice Tversky que tenemos en ese momento.

Su definición es la siguiente:

$$FTL = (1 - TI)^\gamma \quad (23)$$

Siendo γ el parámetro que controla la no linealidad de las pérdidas. Cuando γ tiende a infinito, el gradiente de las pérdidas tenderá también a infinito ya que TI tiende a 1. Cuando γ tiende a 0, el gradiente de las pérdidas tiende a 0 ya que TI tiende a 1.

Con un valor de $\gamma < 1$, el gradiente de las pérdidas es mayor para casos en los que $TI > 0.5$, forzando así al modelo a centrarse en esos ejemplos. Este comportamiento puede ser útil hacia el final del entrenamiento ya que el modelo todavía tiene incentivos para aprender, aunque TI este cerca de converger. Sin embargo, durante las primeras etapas del entrenamiento dará mayor peso a los ejemplos más fáciles, lo que nos puede llevar a un mal aprendizaje.

En el caso del desbalanceo de clases, FTL resulta útil cuando $\gamma > 1$. Esto nos lleva a mayores gradientes de pérdidas para los casos en los que $TI < 0.5$, lo que fuerza al modelo a centrarse en las partes más difíciles, especialmente en pequeñas segmentaciones, las cuales, suelen tener TI bajos.

Capítulo 1. Introducción

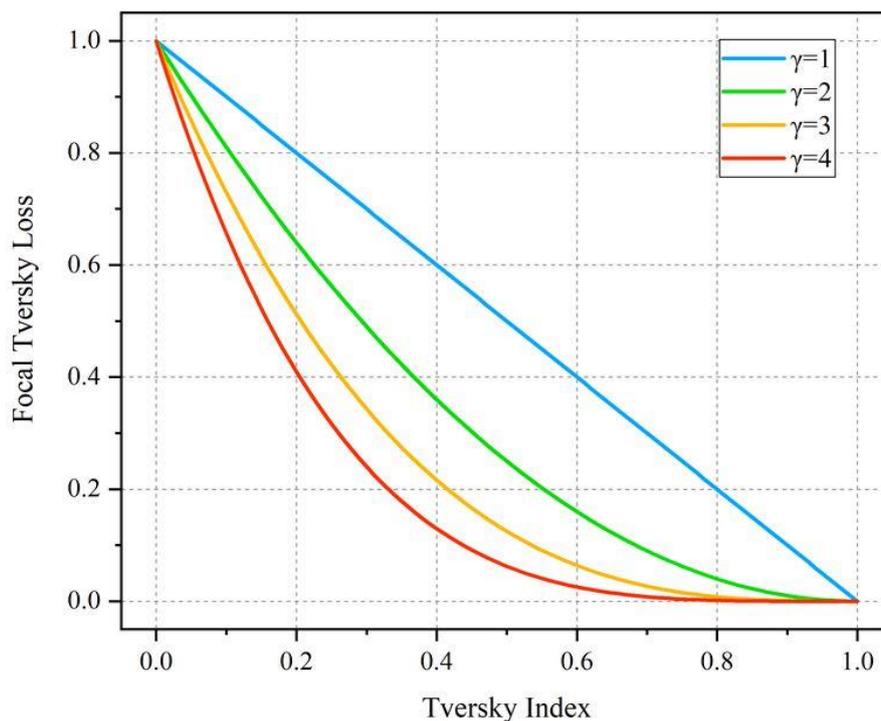


Figura 24. Representación de FTL variando los valores γ . Cuando $\gamma = 1$, tenemos las pérdidas Tversky estándar. (Xu et al., 2020)

En la sección 3.6 explicamos en detalle los parámetros elegidos para controlar la función de pérdidas FTL en el desarrollo de nuestra metodología.

1.3.4.2 Control de tasa de aprendizaje: Cyclical Learning Rate

La tasa de aprendizaje es uno de los hiperparámetros a ajustar más importantes a la hora de entrenar redes neuronales profundas. Se empleó un método conocido como *Cyclic Learning Rate (CLR)*, Tasa de Aprendizaje Cíclica, una técnica utilizada en el entrenamiento de redes neuronales para mejorar la convergencia y el rendimiento del modelo. Esta técnica se basa en la idea de ajustar la tasa de aprendizaje en ciclos durante el entrenamiento en lugar de mantenerla constante. De esta manera eliminamos prácticamente la necesidad de encontrar de manera experimental los mejores valores y elegir una tasa de aprendizaje fija, de manera que el CLR propone variar la tasa de aprendizaje dentro de un rango predefinido durante el entrenamiento, por lo que se define una tasa de aprendizaje mínima y máxima, y luego se va oscilando la tasa de aprendizaje entre estos límites a lo largo del entrenamiento.

El uso del CLR tiene varios beneficios. En primer lugar, permite que el modelo explore una amplia gama de tasas de aprendizaje durante el entrenamiento, lo que puede ayudar a encontrar mínimos locales óptimos. Además, el cambio cíclico de la tasa de aprendizaje puede ayudar a

Capítulo 1. Introducción

evitar el estancamiento en los mínimos locales y permitir saltos más grandes en el espacio de búsqueda de la función de pérdida.

El CLR también puede acelerar el proceso de entrenamiento, ya que permite un ajuste más rápido de los pesos de la red en las primeras etapas, seguido de una refinación más lenta a medida que el ciclo de tasa de aprendizaje progresa. Esto puede llevar a una convergencia más rápida y una mejor generalización del modelo.

El ciclo de la tasa de aprendizaje puede seguir diferentes patrones, como un ciclo triangular o un ciclo triangular modificado. En un ciclo triangular, la tasa de aprendizaje aumenta linealmente desde el valor mínimo hasta el máximo en la primera mitad del ciclo, y luego disminuye linealmente desde el valor máximo hasta el mínimo en la segunda mitad del ciclo. En un ciclo triangular modificado, se utilizan funciones no lineales para ajustar la tasa de aprendizaje en cada iteración. En nuestro caso se empleó el ciclo triangular, ya que con ambas posibilidades conseguíamos resultados similares y éste presentaba una carga computacional algo más ligera.

La tasa de aprendizaje es un hiperparámetro crítico en el entrenamiento de redes neuronales, pues determina la magnitud de los ajustes realizados en los pesos de la red en cada iteración. Una tasa de aprendizaje demasiado alta puede hacer que el modelo diverja o salte por encima de los mínimos locales, mientras que una tasa de aprendizaje demasiado baja puede hacer que el modelo converja lentamente o quede atrapado en mínimos locales subóptimos.

Los métodos convencionales nos indican que la tasa de aprendizaje debe de ser un valor único que decrece monótonamente durante el entrenamiento, pero mediante la técnica que empleamos se percibe una mejora general del modelo, dejando que la tasa varíe cíclicamente entre unos valores en vez de quedarse en un valor fijo.

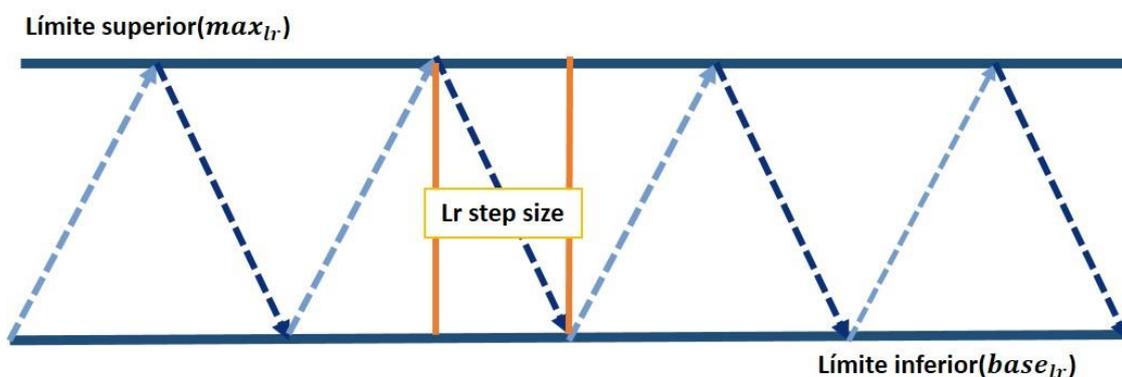


Figura 25. Técnica de tasa de aprendizaje triangular. Las líneas discontinuas representan la tasa de aprendizaje variando entre los valores límite. El parámetro de entrada stepsize es el número de iteraciones en medio ciclo.

Capítulo 1. Introducción

El cálculo de esta nueva técnica seguiría la siguiente dinámica:

$$\text{cycle} = \text{floor} \left(1 + \frac{\text{epochs}}{2 \cdot \text{stepsize}} \right) \quad (24)$$

$$x = \text{abs} \left(\frac{\text{epochs}}{\text{stepsize}} - 2 \cdot \text{cycle} + 1 \right) \quad (25)$$

$$lr = \text{base}_{lr} + (\text{max}_{lr} - \text{base}_{lr}) \cdot \max(0, (1 - x)) \quad (26)$$

Siendo *cycle* la longitud de ciclo, *base_{lr}* el límite inferior, *epochs* el número de épocas de entrenamiento y *lr* la tasa de aprendizaje calculada. En el caso de la técnica *triangular*, también se define *stepsize* (la mitad del periodo o longitud del ciclo) y *max_{lr}* (el límite superior de la tasa de aprendizaje).

En la sección 3.6, explicamos detalladamente el procedimiento para realizar correctamente el ajuste del método de tasa de aprendizaje cíclica.

1.3.4.3 Métrica en entrenamiento: Coeficiente de Correlación de Matthew

Para evaluar problemas de clasificación binaria como al que nos enfrentamos (vóxeles pertenecientes a la médula cervical o no), se pueden utilizar varias medidas estadísticas según el objetivo de la investigación que se está llevando a cabo. A pesar de ser un punto crucial en los algoritmos de aprendizaje automático, todavía no existe en el área un consenso generalizado sobre la elección de una medida unificada. Tanto la precisión como el *F1-score* han sido y siguen siendo las métricas más populares en las tareas de clasificación binaria, aunque este tipo de métricas nos pueden llevar a resultados demasiado optimistas que pongan en peligro la efectividad de nuestro modelo, sobre todo en conjuntos de datos de carácter desbalanceado.

El MCC (Chicco & Jurman, n.d.) es un método de matriz de contingencia del cálculo del coeficiente de correlación producto-momento de Pearson (Powers & Ailab, 2020) entre los valores actuales y los predichos. En términos de matriz de confusión, la definición de MCC es:

$$\text{MCC} = \frac{\text{VP} \cdot \text{VN} - \text{FP} \cdot \text{FN}}{\sqrt{(\text{VP} + \text{FP}) \cdot (\text{VP} + \text{FN}) \cdot (\text{VN} + \text{FP}) \cdot (\text{VN} + \text{FN})}} \quad (27)$$

El MCC es una métrica de clasificación binaria que sólo genera una puntuación alta si el predictor binario es capaz de predecir de manera correcta la mayoría de las instancias de datos positivos

Capítulo 1. Introducción

y la mayoría de datos negativos (Chicco, 2017; Jurman et al., 2012). Su valor está en el rango de $[-1,1]$, siendo -1 una clasificación totalmente errónea y $+1$ una clasificación perfecta, mientras que $MCC=0$ correspondería a un clasificador aleatorio.

Es por ello que el MCC se utilizó como métrica en este modelo, al ser una métrica óptima y sólida para evaluar el rendimiento de modelos de clasificación binaria que se enfrentan a desequilibrios en los datos.

Capítulo 2. Hipótesis y objetivos

Capítulo 2. Hipótesis y objetivos

La atrofia de la médula espinal en la enfermedad de EM es un proceso degenerativo que puede ocurrir como resultado de la inflamación crónica y la desmielinización en esta área del SNC, por lo que el cálculo preciso de la zona atrofiada de la médula espinal es fundamental para proporcionar información valiosa sobre la progresión de la enfermedad y sus consecuencias clínicas.

2.1 Hipótesis

El uso de técnicas basadas en redes neuronales convolucionales para la segmentación de la médula espinal cervical en imágenes de RM de pacientes con EM puede ser una herramienta eficaz para la valoración de la atrofia. Además, esta metodología puede tener un gran impacto tanto en el cuidado del paciente y el manejo y tratamiento de la enfermedad, como en la realización de investigaciones que nos permitan desarrollar nuevas terapias en el futuro.

La atrofia de la médula espinal es una complicación común de la EM que puede afectar significativamente la calidad de vida de los pacientes. La atrofia es un indicador directo del daño neuronal y de la pérdida de tejido no únicamente en el cerebro, sino también en la médula espinal, con el consiguiente aumento del riesgo de discapacidad en la EM. Al calcular de forma precisa la atrofia y evitando la variabilidad inter-observador, los clínicos pueden evaluar el grado de daño neurológico y seguir su evolución a lo largo del tiempo evitando la laboriosa segmentación manual toda la estructura medular, proporcionando una medida cuantitativa objetiva del impacto de la enfermedad en el SNC, permitiendo realizar estimaciones más precisas del pronóstico de discapacidad a largo plazo, lo que ayuda a los pacientes a comprender mejor su situación y al clínico a planificar su tratamiento y manejo de manera adecuada. La EM, al ser una enfermedad crónica y progresiva, necesita de la máxima información posible sobre la velocidad de progresión de la enfermedad y la eficacia de los tratamientos en la prevención de la atrofia.

2.2 Objetivos

El objetivo principal de esta tesis es investigar y desarrollar un nuevo método de segmentación automático de la médula espinal cervical en pacientes de EM, así como una metodología de extracción de biomarcadores de imagen de atrofia, a partir de información de imágenes de RM, que permita la evaluación automatizada y precisa de la atrofia de la médula espinal, ayudando así al diagnóstico y seguimiento de la enfermedad.

Capítulo 2. Hipótesis y objetivos

Este método se implementa a partir datos de pacientes del mundo real (*Real-world data, RWD*) a los que se le aplica el mínimo preprocesado posible con fines de generalización y reproducibilidad del modelo y en busca de la máxima automaticidad, de manera que la segmentación pueda ser extraída de datos comunes adquiridos en diferentes equipos de RM utilizados en la práctica clínica habitual (*standard-of-care*).

Asimismo, se han planteado los siguientes objetivos secundarios:

- Evaluar los protocolos de adquisición de médula cervical en pacientes con esclerosis múltiple en un hospital de referencia y valorar el aprovechamiento de los datos adquiridos en el mundo real (RWD) para la creación de un modelo de IA que permita la segmentación de la médula.
- Extracción del conjunto de datos (*Dataset*), compuesto por imágenes de RM anonimizadas de pacientes de EM.
- Aplicar técnicas de normalización de los datos, tanto de tamaño como de nivel de intensidad para minimizar la variabilidad entre las adquisiciones de RM.
- Generación del “*ground truth*” mediante la segmentación manual de la médula espinal por radiólogos expertos.
- Desarrollar técnicas originales de segmentación basadas en la combinación de las técnicas actuales y en la implementación de nuevos algoritmos avanzados de procesamiento digital de la señal.
- Validación del modelo final mediante la evaluación de su rendimiento en nuevos casos de EM correspondientes a adquisiciones de imagen posteriores a las utilizadas para la creación del modelo.
- Extracción de información sobre la atrofia medular como biomarcadores de imagen.
- Estudio de la relación entre los biomarcadores extraídos, volumen y sección media de la médula espinal cervical, con la distribución de las distintas formas clínicas que componen nuestro grupo de estudio.
- Estudio de relación entre los niveles en la EDSS de cada paciente de nuestro grupo de estudio con su valor de volumetría de médula espinal cervical calculado por nuestro método.

Capítulo 3. Materiales y métodos

Capítulo 3. Materiales y métodos

La metodología seguida para alcanzar el objetivo de esta tesis consiste en un proceso de análisis que se divide en diferentes subtarefas previas a la obtención del resultado final, la segmentación de la médula espinal cervical. Este proceso se describe gráficamente en la Figura 26 y fue dividido en 4 pasos principales:

- Adquisición: adquisición de las imágenes para formar la base de datos
- Normalización y re-escalado a un tamaño común para evitar heterogeneidades en las imágenes tanto de intensidad como de tamaño.
- Etiquetado y generación de *ground truth*: delineación por expertos de los píxeles pertenecientes a la médula espinal cervical
- Entrenamiento y validación.

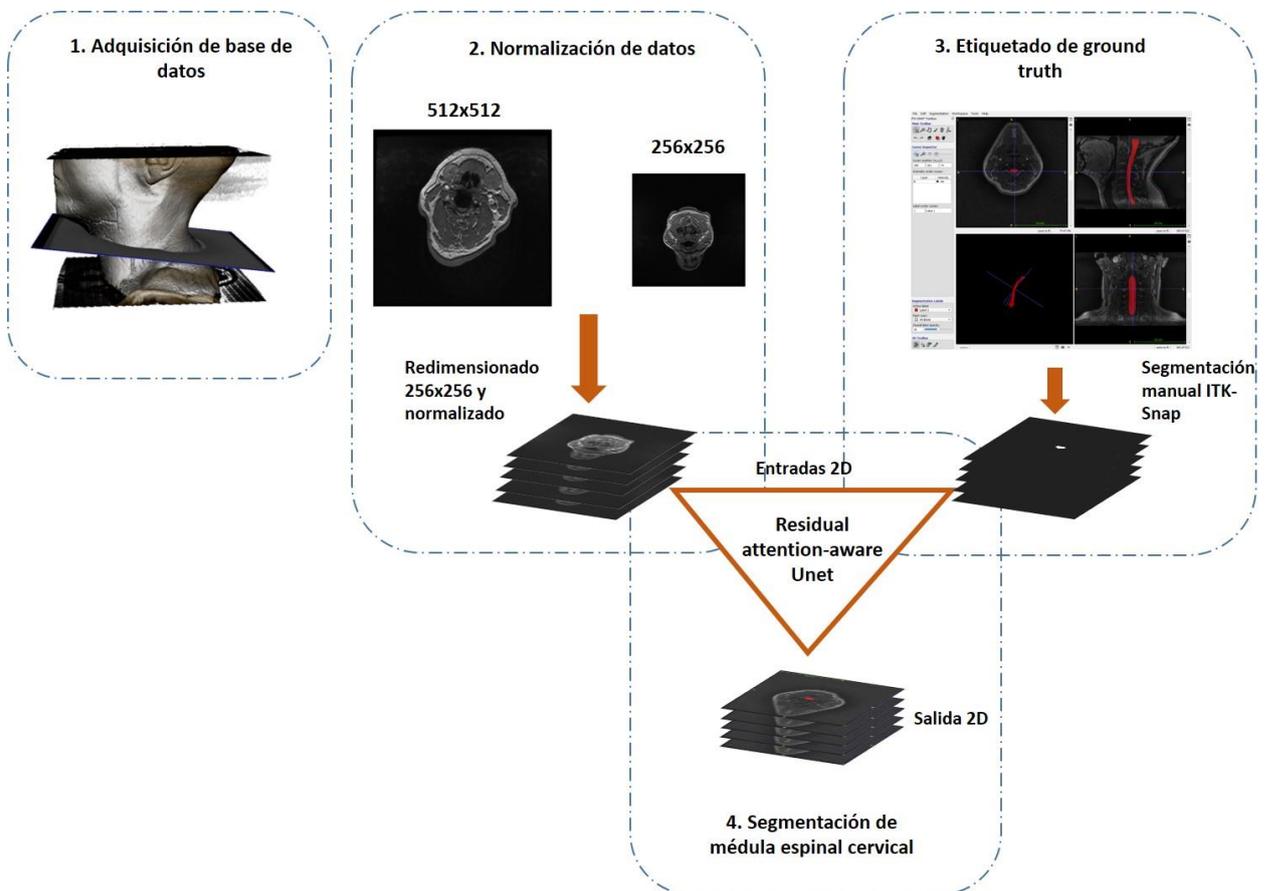


Figura 26. Diagrama de flujo utilizado para desarrollar la herramienta de segmentación automática de médula cervical propuesta. Partiendo de la generación del conjunto de datos (paso 1) y pasando por su normalización (paso 2) y etiquetado (paso 3) hasta llegar al entrenamiento y validación de la arquitectura propuesta (paso 4).

Capítulo 3. Materiales y métodos

3.1 Bases de datos

Este estudio de carácter observacional sin intervención fue aprobado por el Comité de Ética del Hospital Universitario y Politécnico La Fe para garantizar que la investigación se realice de manera ética y segura, protegiendo los derechos, la seguridad y el bienestar de los pacientes involucrados en el estudio.

Para el desarrollo del algoritmo se descargaron y anonimizaron 121 exploraciones de RM en formato DICOM del PACS del Hospital, cuyas fechas de adquisición comprenden los años 2015 y 2016.

La serie de población estaba compuesta por 34 hombres y 87 mujeres de entre 23 y 74 años de edad, siendo la media de 45 años. En cuanto al curso clínico de la EM, el 68% de los pacientes fueron diagnosticados como EMRR el 25% como EMSP y el 7% como EMPP. Las imágenes RM empleadas en el algoritmo fueron secuencias transversales 3D potenciadas T1 adquiridas en varios equipos 3T: Signa HDxt y Optima MR360 de GE Healthcare. Las adquisiciones se realizaron con Tiempos de Eco (TE) de entre 1.7-2.7ms, Tiempos de Relajación (TR) de entre 5.6-8.2 ms y un ángulo de giro (*Flip Angle, FA*) de 12°. Todas las imágenes RM tenían un espesor de corte de 1 mm y una resolución isotrópica en el plano que iba desde 0.43 x 0.43 mm a 1.02 x 1.02 mm.

Tabla 5: características de las secuencias adquiridas para entrenamiento y prueba.

Equipos	Signa HDxT Optima MR360, GE Healthcare
Tiempo de eco, TE (ms)	1.7 – 2.7
Tiempo de relajación, TR (ms)	5.6 – 8.2
Ángulo de giro, FA (°)	12
Espesor de corte (mm)	1
Resolución isotrópica en el plano (mm)	0.43x0.43 – 1.02x1.02

El conjunto de datos se dividió en dos subconjuntos para el desarrollo del algoritmo, siguiendo el Principio de Pareto (*The Pareto Principle | Dunford | The Plymouth Student Scientist, n.d.*), también conocido como la regla 80/20, el cual contenía 96 secuencias de RM, lo que se traduce en un total de 15668 imágenes, y el 20% para test, con 25 secuencias de RM y 4260 imágenes, con fines de evaluación del rendimiento y capacidad de generalización del modelo con datos no vistos previamente.

Para la validación de nuestro modelo, se descargaron del mismo PACS del Hospital Universitario y Politécnico de la Fe, 13 secuencias transversales 3D potenciadas T1 procedentes de los dos mismos equipos, de pacientes entre los 25 y 67 años, 9 mujeres y 4 hombres, cuyas fechas de adquisición eran a lo largo del año 2020, aproximadamente 5 años después de cuando fueron

Capítulo 3. Materiales y métodos

adquiridas las secuencias utilizadas para generar el modelo, con el fin de garantizar que el resultado de rendimiento de nuestro modelo se sigue garantizando en datos no vistos, con protocolos que habían cambiado, y no sesgado por la partición aleatoria del conjunto de datos de entrenamiento y test, proporcionando así una evaluación objetiva e imparcial del rendimiento del modelo.

Tabla 6: características de las secuencias adquiridas para validación.

Equipos	Signa HDxT Optima MR360, GE Healthcare
Tiempo de eco, TE (ms)	1.5 – 2.8
Tiempo de relajación, TR (ms)	6.53 – 8.50
Ángulo de giro, FA (°)	12
Espesor de corte (mm)	1
Resolución isotrópica en el plano (mm)	0.43x0.43 – 0.94x0.94

En el estudio se utilizaron imágenes RM adquiridas en la práctica clínica rutinaria, lo que se conoce como “datos del mundo real, (*Real-world data*)”, esto es, datos recopilados en pacientes y poblaciones en condiciones reales en lugar de datos generados en entornos controlados de investigación clínica o ensayos clínicos, lo que favorece a la reproducibilidad y factibilidad mediante la inclusión de diversidad de condiciones médicas, tratamientos, resultados y factores ambientales y sociales. Este hecho limita en cierta forma los resultados ya que se podrían obtener mejores segmentaciones a través de secuencias de imágenes de alta resolución, con sincronismo de movimiento y tiempos de adquisición mayores, pero quedaban fuera de los dominios del término “*Real-world data*” y nos hacían perder aplicabilidad en la clínica rutinaria.

Basándonos en lo reflejado en el Capítulo 1, concretamente en la parte de protocolos en la sección 1.2.4.2, para el estudio utilizamos la sección de médula espinal a nivel cervical, ya que existen numerosas evidencias que respaldan la medición rutinaria de la pérdida de volumen de la columna cervical en la EM (Bischof et al., 2022; Casserly et al., 2018; Mina et al., 2021; Papinutto et al., 2018) especialmente en formas más avanzadas de la enfermedad, en las que las placas y la formación de cicatrices pueden llevar a un estrechamiento de la médula.

3.2 Anotación/etiquetado de la base de datos

El etiquetado manual de los datos se llevó a cabo de manera individual y detallada en cada caso, identificando y delimitando en cada corte de la secuencia de RM los píxeles pertenecientes a la médula cervical, bajo la supervisión de dos expertos radiólogos con más de 15 y 30 años de

Capítulo 3. Materiales y métodos

experiencia, entre los cuales se resolvieron posibles discrepancias con un tercer radiólogo, obteniendo finalmente un conjunto de máscaras de médula espinal cervical como *ground truth*.

El etiquetado manual de los datos se realizó mediante el software ITK-SNAP el cual describimos en la sección 3.3.2, una herramienta que nos permite cargar imágenes médicas, navegar en ellas y delinear manualmente regiones anatómicas de interés.

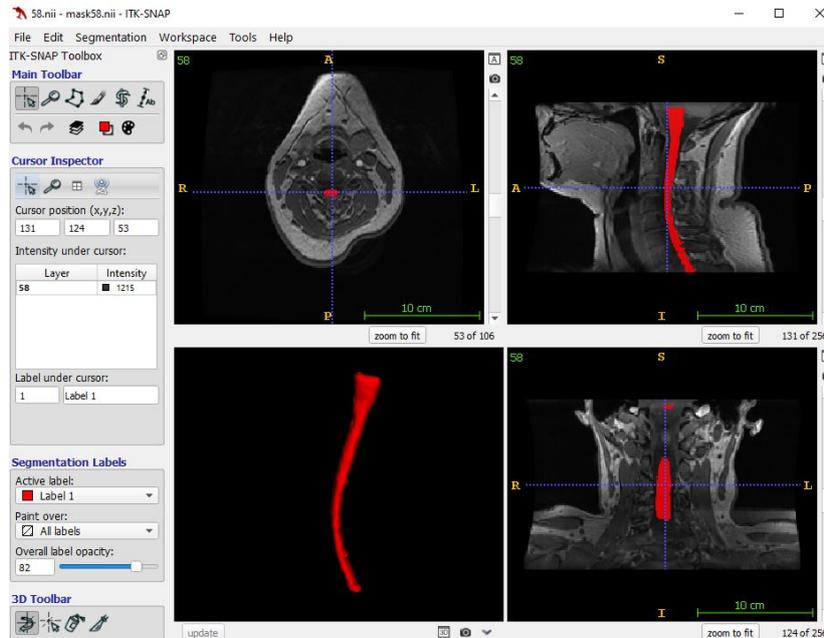


Figura 27. Interfaz gráfica de ITK-SNAP. Ventana dividida por una sección de herramientas disponibles a la izquierda, y una cuadrícula compuesta por 4 imágenes del mismo volumen, 3 vistas en 2D y una en 3D.

3.3 Hardware y software

3.3.1 Anaconda y entorno de GPU

Anaconda es una plataforma de código abierto y gratuita que facilita la gestión de paquetes y ambientes de desarrollo para la ciencia de datos, el análisis y la programación en Python. Fue desarrollado por *Continuum Analytics* (ahora parte de Anaconda, Inc., responsable de mantener y evaluar la herramienta y los paquetes disponibles en el repositorio Conda) y está diseñado para simplificar la instalación y gestión de bibliotecas y herramientas comunes utilizadas en proyectos de análisis de datos y aprendizaje automático Keras y tensorflow. Anaconda es compatible con Windows, macOS y Linux, lo que permite a los usuarios trabajar en una variedad de sistemas operativos.

Entre sus características destaca: una aplicación gráfica que integra las herramientas y versiones de Python orientadas al desarrollo científico, un IDE para desarrollar fácilmente aplicaciones científicas, tanto en código como creación de gráficos, el gestor de paquetes integrado (Conda),

Capítulo 3. Materiales y métodos

la integración avanzada con Jupyter Notebook, el soporte de entornos virtuales de Python y la posibilidad de poder utilizar Python y R en el mismo entorno. La versión de conda utilizada fue la 4.14.0.



Figura 28. Imagen corporativa de Anaconda (Welcome! — Anaconda Documentation, n.d.).

Con anaconda utilizamos librerías como numpy, nibabel, scikit-learn, scipy o matplotlib. También creamos un entorno con TensorFlow y Keras para utilizar la GPU en tareas de aprendizaje profundo. Para el desarrollo del modelo propuesto, creamos un entorno con Python 3.7.4, tensorflow-gpu 1.14.0, keras-gpu 2.2.4, CUDA 10.0 y cuDNN 7.4. Es importante asegurarse previamente de que se cumplen los requerimientos necesarios para el tipo de tarjeta gráfica y compatibilidad entre versiones de bibliotecas, versión de Python y controladores necesarios para el correcto uso mediante GPU. Nuestro modelo de tarjeta gráfica era Nvidia GeForce GTX 2080 de 8Gb de memoria.

3.3.2 Etiquetado manual: ITK-SNAP

ITK-SNAP es un software de código abierto utilizado en la visualización, segmentación y registro de imágenes médicas en 3D de varios formatos. El nombre "ITK" proviene de la "*Insight Segmentation and Registration Toolkit*", que es una biblioteca de procesamiento de imágenes médicas desarrollada por el *National Library of Medicine* (NLM) y el *Center for Imaging Science* de la Universidad de Rochester. La versión utilizada fue ITK-SNAP 3.6.0.

Mediante ITK-SNAP y una interfaz gráfica como la mostrada en la Figura 27. cargamos el fichero Nifti que contiene nuestra imagen axial 3D T1 y delineamos manualmente, imagen a imagen, todos los píxeles que se consideran pertenecientes a la médula espinal cervical, generando así, como *ground truth*, una estructura en 3D, también en formato Nifti, para cada uno de los 121 pacientes de nuestro conjunto de datos.



Figura 29. Imagen corporativa de ITK-SNAP (ITK-SNAP Home, n.d.).

3.3.3 Conversión de formato DICOM a Nifti

Para la conversión de los ficheros DICOM obtenidos directamente del PACS al formato Nifti utilizado para procesar los datos, se utilizó el software dcm2nii, una herramienta gratuita diseñada para ello e incluida en el software MRICroGL, una herramienta de visualización de imágenes cerebrales en 2D y 3D desarrollada por el Dr. Chris Rorden, neurólogo británico. La versión utilizada fue MRICroGL 1.0.20190902.

3.4 Preprocesado de datos

El preprocesado de datos es una etapa crucial en el desarrollo de cualquier algoritmo de IA, ya que se suele buscar que el conjunto de datos esté bien estructurado, limpio de ruidos o artefactos indeseados y normalizado y preparado para alimentar adecuadamente al modelo que va a ser entrenado. Por ejemplo, las capas totalmente conectadas de las CNN requieren que todas las imágenes estén en vectores del mismo tamaño. Aunque en algunas ocasiones es necesario aplicar un preprocesado complejo (por ejemplo, si el conjunto de datos contiene mayoritariamente mucho ruido y se deben aplicar técnicas de filtrado, mejoras de contraste o señal, con el fin de obtener mejores resultados) hay situaciones en las que intentar hacer un preprocesamiento mínimo puede ser beneficioso.

Nuestra tarea tenía como objetivo la aplicación del mínimo preprocesado de modo que se cumpliera con una serie de requerimientos:

- Pérdida mínima de información, ya que el preprocesado podía involucrar la pérdida o transformación de características que podía ayudar al modelo a aprender patrones más precisos y relevantes.
- Mayor generalización, con el propósito de que el modelo fuese capaz de generalizar mejor a datos nuevos y mejorando su rendimiento en situaciones del mundo real, por ejemplo, el uso de nuestra herramienta de segmentación automática de médula cervical con imágenes de RM procedentes de un equipo distinto al del conjunto de datos de entrenamiento.

Capítulo 3. Materiales y métodos

Por ello, el preprocesado de datos que se aplicó a nuestro conjunto de datos se reduce a dos simples etapas: un redimensionamiento a un tamaño común y una normalización de la intensidad de las imágenes, ayudando a evitar efectos no deseados como el efecto escalonado (*aliasing*).

3.4.1 Redimensionado

Todas las imágenes RM en 2D se redimensionaron a un tamaño común de 256 x 256 píxeles empleando un algoritmo de interpolación bicúbica de 4 x 4 píxeles.

En lugar de simplemente interpolar linealmente entre puntos, la interpolación bicúbica utiliza información de los puntos vecinos para crear una superficie más suave y precisa. Este algoritmo se basa en el uso de una función polinómica de tercer grado, que se ajusta a los valores de los puntos vecinos en la cuadrícula para estimar el valor en una posición intermedia. En lugar de solo tomar en cuenta los puntos más cercanos, la interpolación bicúbica utiliza una ventana de 4x4 píxeles (16 píxeles) alrededor del punto de interpolación, para tener en cuenta una mayor cantidad de información local y lograr una estimación más suave.

En el caso de las máscaras etiquetadas manualmente, también se redimensionaron al tamaño común de 256 x 256 píxeles, pero esta vez mediante el algoritmo de interpolación del vecino más próximo. A diferencia de la interpolación bicúbica, este algoritmo es computacionalmente más eficiente, tomando el valor del píxel más cercano, sin considerar los valores de los píxeles vecinos. Aunque no proporciona resultados de la misma calidad, al aplicarse para interpolar imágenes binarias, donde solo hay dos valores de píxeles posibles, este algoritmo es suficiente para lograr la ampliación o reducción de la imagen sin afectar a la interpretación binaria.

3.4.2 Normalización

La normalización consiste en ajustar y escalar las intensidades de los píxeles en las imágenes para que estén dentro de un rango específico y uniforme, lo que facilita el entrenamiento del modelo y mejora su capacidad de generalización, ya que, por ejemplo, en nuestro caso, si trabajamos con imágenes adquiridas con diferentes escáneres o en distintos momentos, es común que las intensidades de los píxeles varíen debido a diversas condiciones de adquisición. Esta variabilidad en la intensidad puede dificultar el proceso de entrenamiento y hacer que el modelo se enfoque más en las diferencias de intensidad en lugar de aprender características relevantes. Por ello, se aplicó la normalización de cada imagen de RM en 2D para conseguir unos nuevos valores de intensidad entre 0 y 1 siguiendo la siguiente expresión:

$$z(i, j) = \frac{(x(i, j) - x_{\min})}{x_{\max} - x_{\min}} \quad (28)$$

Capítulo 3. Materiales y métodos

Siendo $z(i, j)$ el píxel ya normalizado, $x(i, j)$ la intensidad original del píxel y x_{min} , x_{max} los valores de intensidad mínimos y máximos de la imagen, en ese orden.

3.4.3 Conversión de imágenes a tensor

Para nuestra aplicación, un tensor es una estructura de datos que se utiliza para procesar múltiples imágenes a la vez en una red neuronal. Permite que el modelo procese lotes o conjuntos de imágenes en paralelo, lo que mejora significativamente la eficiencia del entrenamiento y la inferencia. El tensor cuenta con 4 dimensiones: (número de imágenes, alto, ancho, número de canales), es decir, un conjunto de imágenes apiladas en la dimensión del lote donde cada imagen se representa mediante una matriz tridimensional (altura, ancho, canales).

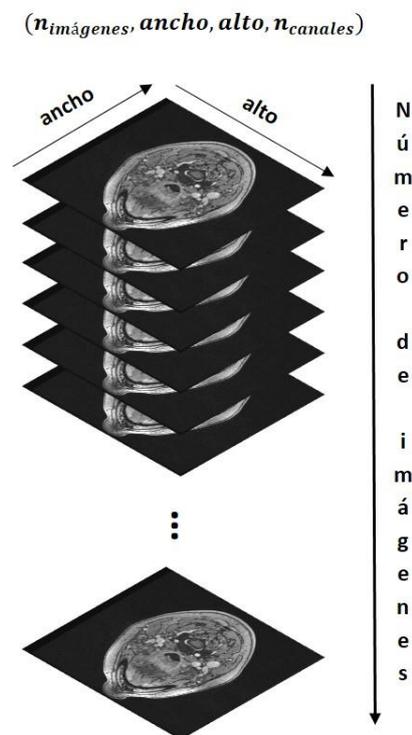


Figura 30. Representación de tensor 3D de imágenes 2D.

Un total de cuatro tensores se crearon en nuestro método siendo almacenados en formato “.npy” de la librería Numpy, dos para el conjunto de entrenamiento, uno con las imágenes de RM y otro con las máscaras, y dos con el conjunto de test, uno con las imágenes de RM y otro con las máscaras. Al ser las imágenes de RM en escala de grises, nuestro número de canales es 1, por lo que las dimensiones de nuestros tensores son: en entrenamiento (15668, 256, 256, 1) y en test (4260, 256, 256, 1)

3.5 Técnicas de inteligencia artificial empleadas

Para conseguir nuestro objetivo, se propuso una arquitectura compuesta por un mecanismo de atención residual y conexiones U-Net, la cual denominamos como *Residual attention-aware 2D U-Net (RA-Unet)* y es explicada en la sección 1.3.3. Este enfoque combinaba los conceptos de la red U-Net y la atención residual para mejorar la precisión y la capacidad de atención del modelo. Antes de llegar a esta decisión, se realizó un estudio de comparación de arquitecturas existentes, las cuales se introducen en la siguiente sección 3.5.1 y cuyos resultados obtenidos durante la evolución de la investigación se muestran en la sección 4.5.

3.5.1 Arquitecturas evaluadas durante la investigación

Hasta llegar a la arquitectura escogida, se realizó una batería de pruebas con diferentes arquitecturas y configuraciones posibles. Cada una de estas pruebas ayudaron a recorrer un proceso en el que las carencias de las redes ayudaban a un mejor reconocimiento de las necesidades específicas de nuestra problemática, como puede ser la constante búsqueda de lidiar con la problemática del desbalanceo de clases. A continuación, se describen brevemente algunas de ellas.

3.5.1.1 Attention U-Net

La arquitectura *Attention U-Net* (Oktay et al., 2018), utiliza la idea de los mecanismos de atención e incorpora conexiones de salto, de manera que aporta información extra sobre en qué región es importante centrarse mientras se aprende la tarea de segmentación. Como podemos ver utiliza una filosofía similar a la arquitectura final que empleamos, *Residual-Attention Aware U-Net*, pero en este caso, una de las mayores diferencias es que no se emplean bloques residuales para permitir una mejor retropropagación de gradiente y mecanismos de atención más suavizados, adición que, como veremos en la sección 1.3.3.1.3, aportan una notable mejora del rendimiento. Los modelos entrenados con mecanismos de atención, como ya hemos visto anteriormente, aprenden implícitamente a suprimir regiones irrelevantes en la imagen de entrada mientras que resaltan las características que son más útiles para la tarea en cuestión.

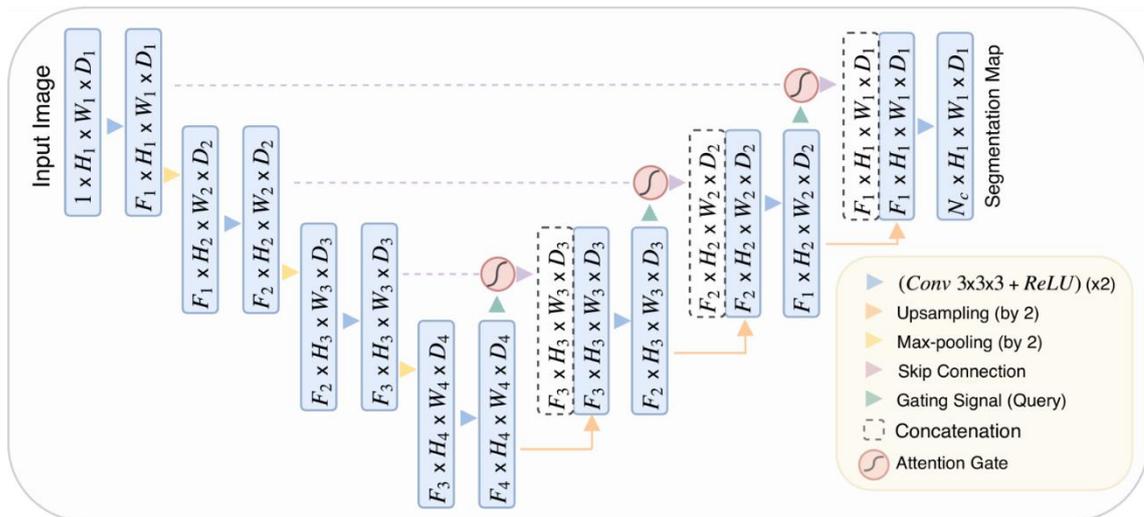


Figura 31. Diagrama de bloques del modelo de segmentación Attention U-Net. Las puertas de atención filtran las características propagadas a través de las conexiones de salto. Imagen extraída del paper original de Attention U-Net (Oktay et al., 2018).

Las puertas de atención aparecen tras el bloque codificador, que es básicamente la mayor aportación de *Attention U-Net* respecto a *U-Net*. Estas puertas de atención se utilizan dentro del bloque decodificador, empleado para resaltar las características en las conexiones de salto y suprimir las que son irrelevantes, lo que ayuda a la mejora del rendimiento del modelo.

3.5.1.2 Deep U-Net

Deep U-Net (Li et al., 2017), es una arquitectura también basada en la original U-Net, la cual mantiene la filosofía de una ruta de compresión y una de expansión que reconstruye una representación espacial precisa de la imagen segmentada. Encontramos diferencias en que, en la ruta de compresión, *Deep U-Net* utiliza “*DownBlocks*” en vez de capas de convolución y “*UpBlocks*” en la ruta de expansión. Estos dos nuevos bloques aportan dos nuevas conexiones conocidas como *U-connection* y *Plus-connection*, las cuales buscan resultados de segmentación más precisos. La parte izquierda está formada por una serie de “*DownBlocks*”, representados en azul en la Figura 32, que están conectados a los “*UpBlocks*” correspondientes de la parte derecha, representados en verde. Estas conexiones se pueden ver en amarillo y son las llamadas “*U-connections*”. En morado están representadas las “*Plus-connections*” entre los sucesivos “*DownBlocks*” y “*UpBlocks*” las cuales se encargan de solucionar el problema del aumento en el error de pérdidas conforme la red se vuelve más profunda.

Capítulo 3. Materiales y métodos

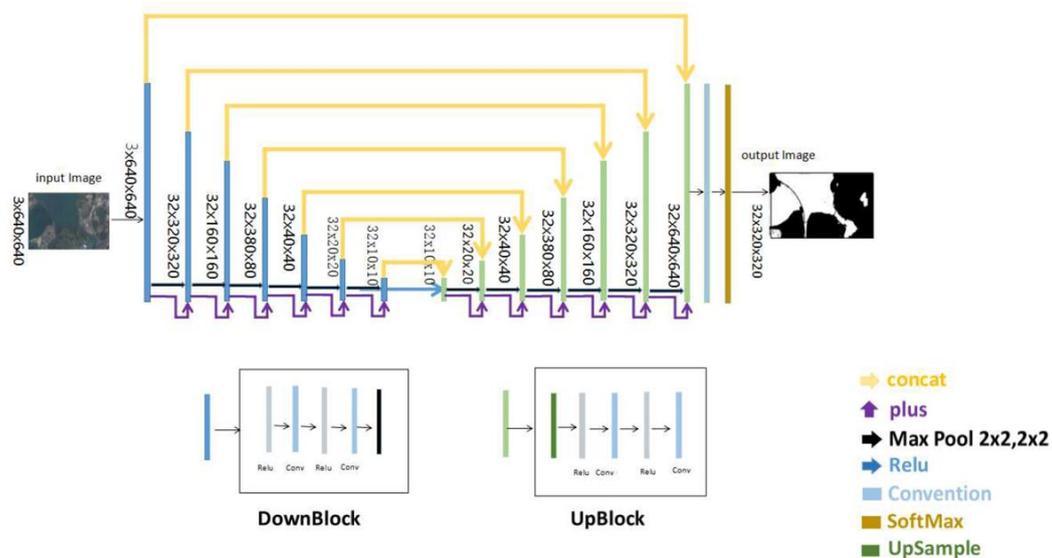


Figura 32. Estructura Deep U-Net. Imagen extraída del paper original (Li et al., n.d.).

La idea principal de *Deep U-Net* es concatenar las capas de la ruta de compresión con sus correspondientes en la capa de expansión, de manera que las características de alta resolución de la ruta de compresión son combinadas con la salida sobremuestreada, de manera que, mediante una capa convolucional sucesiva se pueda aprender a dar lugar a una salida más precisa basada en esta información. Además, con el fin de extraer mejor información semántica con menor error de pérdidas, esta arquitectura optimiza la ruta de compresión y de expansión con la introducción de los ya mencionados “*Downblock*”, “*UpBlock*” y las “*Plus-connections*”. Tanto en el “*DownBlock*” como en el “*UpBlock*”, las características previas y posteriores a las capas de convolución se suman, por lo que, de esta manera, esta estructura se ahorra la operación de convolución y proporciona una arquitectura neuronal convolucional más profunda y eficiente.

3.5.1.3 U-Net ++

Esta arquitectura U-Net++ (Z. Zhou et al., 2018), también proviene de una modificación de U-Net y es una red profunda supervisada con forma codificador-decodificador en la que las distintas partes codificadoras y decodificadora están conectadas por una serie de conexiones de salto anidadas diseñadas con el objetivo de reducir la brecha semántica entre los mapas de características de la parte del codificador y del decodificador.

Se basa en la hipótesis de que el modelo podrá captar detalles más finos del fondo de la imagen cuando los mapas de características de alta resolución de la parte del codificador se mejoran gradualmente antes de que se junten con los mapas de características ya optimizados

Capítulo 3. Materiales y métodos

semánticamente correspondientes a la parte del decodificador, es decir, se defiende la idea de que la red se enfrenta a una tarea de aprendizaje más sencilla cuando los mapas de características de la parte del codificador y del decodificador son semánticamente similares. Esto contrasta con la idea de conexiones de salto planas utilizadas en U-Net, las cuales envían los mapas de características de alta resolución directamente del codificador al decodificador, dando lugar a la fusión de mapas de características poco similares.

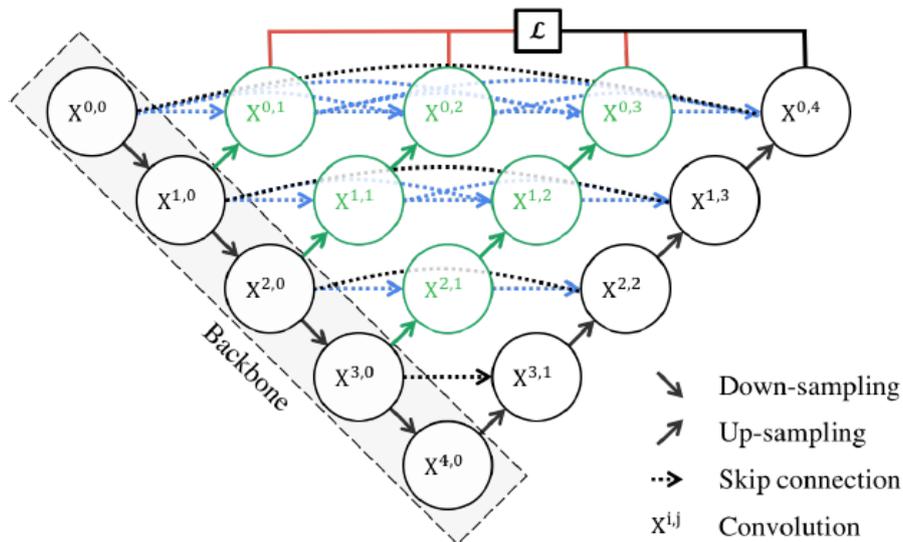


Figura 33. Estructura de U-Net ++. Imagen extraída del paper original (Z. Zhou et al., 2018).

En la Figura 33 en negro se puede observar la red U-Net original, en verde y en azul los bloques densos de convolución en los caminos de salto y el color rojo indica supervisión profunda. Estos componentes señalados en rojo, verde y azul son los distintivos entre U-Net ++ y U-Net.

3.6 Configuración del entrenamiento y métricas en test

Nuestro modelo ha sido entrenado con el 80% de los datos disponibles, descritos en la sección 3.1, con un tamaño de *batch* de 15 imágenes durante 30 *epochs*, empleando la técnica de regularización *Early Stopping*, de manera que nuestro entrenamiento parará en el caso de que durante 10 *epochs* no haya habido mejoras significativas, reflejadas como aumento de la métrica MCC o decremento del valor de la función de pérdidas.

Para configurar el entrenamiento, ajustamos principalmente tres hiperparámetros, tal como definimos en la sección 1.3.4: la función de pérdidas, la tasa de aprendizaje y la métrica de entrenamiento. A continuación, analizamos las elecciones llevadas a cabo.

Función de pérdidas: Focal Tversky Loss

Tal como explicamos en la sección 1.3.4, la función de pérdida Tversky se utiliza comúnmente en problemas de aprendizaje automático y está diseñada específicamente para abordar

Capítulo 3. Materiales y métodos

desequilibrios de clase y tener en cuenta tanto los falsos positivos como los falsos negativos. La función Focal Tversky Loss (FTL) es una generalización de las pérdidas Tversky mediante la cual, a través de la naturaleza no lineal de la función podemos controlar cómo se comporta ésta para diferentes valores del índice Tversky que tenemos en ese momento.

Recordamos su definición:

$$FTL = (1 - TI)^\gamma = \left(1 - \frac{VP}{VP + \alpha \cdot FN + \beta \cdot FP}\right)^\gamma \quad (23)$$

Siendo γ el parámetro que controla la no linealidad de las pérdidas. Cuando γ tiende a infinito, el gradiente de las pérdidas tenderá también a infinito ya que TI tiende a 1. Cuando γ tiende a 0, el gradiente de las pérdidas tiende a 0 ya que TI tiende a 1. VP representa los verdaderos positivos, es decir, las regiones correctamente segmentadas por el modelo, FP representa los falsos positivos, es decir, las regiones que el modelo predijo como positivas pero que no corresponden a las regiones verdaderas y FN representa los falsos negativos, es decir, las regiones que el modelo no identificó como positivas pero que deberían haber sido detectadas.

Recordamos que los parámetros α y β permiten ajustar el énfasis en los falsos negativos y falsos positivos, respectivamente. Si se da más importancia a los falsos negativos, se puede aumentar el valor de α , y si se da más importancia a los falsos positivos, se puede aumentar el valor de β .

En nuestra configuración de entrenamiento, empleamos como función de pérdidas la definida por Focal Tversky Loss, escogiendo como $\gamma = 1.33$, $\alpha = 0.6$ y $\beta = 0.4$. Al ser $\gamma > 1$ tendremos mayores gradientes de pérdidas cuando $TI < 0.5$, lo que fuerza al modelo a centrarse en las partes más difíciles, especialmente en pequeñas segmentaciones, las cuales, suelen tener TI bajos, como la zona correspondiente a la médula espinal. Cuando escogemos $\alpha = 0.6 > \beta$ estamos penalizando más a los falsos negativos, lo que es útil en conjuntos de datos que presentan un gran desbalanceo de clases como el nuestro, ya que el nivel adicional de control sobre la función de pérdida produce mejores segmentaciones a pequeña escala que por ejemplo otras métricas como el coeficiente DICE.

Control de tasa de aprendizaje: Cyclical Learning Rate

En cuanto a la tasa de aprendizaje, utilizamos el método *Cyclic Learning Rate (CLR)* definido en la sección 1.3.4.2, el cual, mejora la convergencia y el rendimiento del modelo. Esta técnica, recordamos, se basa en la idea de ajustar la tasa de aprendizaje en ciclos durante el entrenamiento en lugar de mantenerla constante, cambiando la necesidad de encontrar de

Capítulo 3. Materiales y métodos

manera experimental la mejor tasa de aprendizaje fija, por una tasa de aprendizaje variable dentro de un rango predefinido durante el entrenamiento, por lo que se define una tasa de aprendizaje mínima y máxima, y luego la tasa de aprendizaje oscila entre estos límites a lo largo del entrenamiento.

Vamos ahora a describir los pasos necesarios a seguir para aplicar este método.

Calcular la longitud de ciclo

La longitud de un ciclo y el parámetro *stepsize*, ambos explicados en la sección 1.3.4.2, fueron calculados a partir del número de iteraciones en una época. Una época se calcula dividiendo el número de imágenes de entrenamiento por el tamaño de batch (*batchsize*) utilizado en el entrenamiento, lo que nos da un número de iteraciones. Los resultados finales de precisión son normalmente muy robustos en base a la longitud del ciclo, pero experimentalmente se ha demostrado que es mejor establecer un *stepsize* que sea de 2 a 10 veces el número de iteraciones en una época.

Además, esto aporta cierta elegancia al ritmo de estos ciclos y simplifica la decisión de cuando disminuir la tasa de aprendizaje y cuando parar el entrenamiento. Varios experimentos demuestran que, si reemplazamos cada paso de una tasa de aprendizaje constante con al menos 3 de estos ciclos, ya se entrenan la mayoría de los pesos y ejecutando con 4 o más ciclos nos dará incluso un mejor rendimiento.

Calcular los valores mínimos y máximos de oscilación

Para calcular valores mínimos y máximos razonables entre los que varía nuestra tasa de aprendizaje, se realizó un testeo inicial en el que se ejecutó el modelo durante unas pocas épocas dejando que la tasa de aprendizaje aumentara linealmente entre dos valores de *lr*, uno muy bajo y otro muy alto. Para ello, *stepsize* y *max_iter* fueron fijados al mismo número de iteraciones. De este modo, la tasa de aprendizaje aumentaba linealmente del mínimo al máximo en esta corta ejecución. Posteriormente, se representaron las pérdidas en función de la tasa de aprendizaje y se seleccionaron los valores en los que la precisión empieza a aumentar y cuando empieza a bajar. Estos valores son los que se fijaron como *base_{lr}* y *max_{lr}*, respectivamente.

Capítulo 3. Materiales y métodos

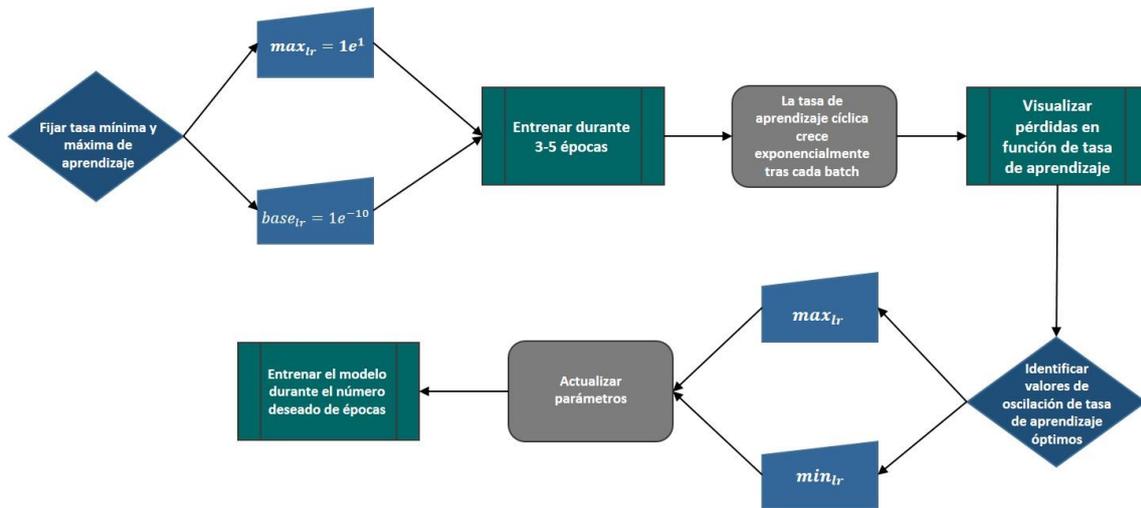


Figura 34. Procedimiento a seguir en la técnica de Learning rate finder (buscador de tasa de aprendizaje).

En nuestro caso, se comenzó entrenando durante 2 *epochs* para ver el comportamiento de nuestro modelo en términos de la función de pérdidas. Para esta prueba inicial elegimos un valor límite muy pequeño de tasa de aprendizaje ($1e^{-10}$) y uno muy alto ($1e^1$), lo suficiente para que a nuestro modelo le resultara difícil aprender.

En la Figura 35 podemos ver el comportamiento de nuestro modelo en términos de pérdidas y determinar los límites óptimos para realizar el entrenamiento completo con la técnica de CLR. En el momento en que observábamos que nuestras pérdidas empezaban a disminuir, podíamos considerar que era un valor óptimo para fijar nuestro rango mínimo de oscilación de tasa de aprendizaje ($base_{lr}$), mientras que el valor de tasa de aprendizaje en el que las pérdidas aumentan, era un valor óptimo para el valor que fija el rango superior (max_{lr}).

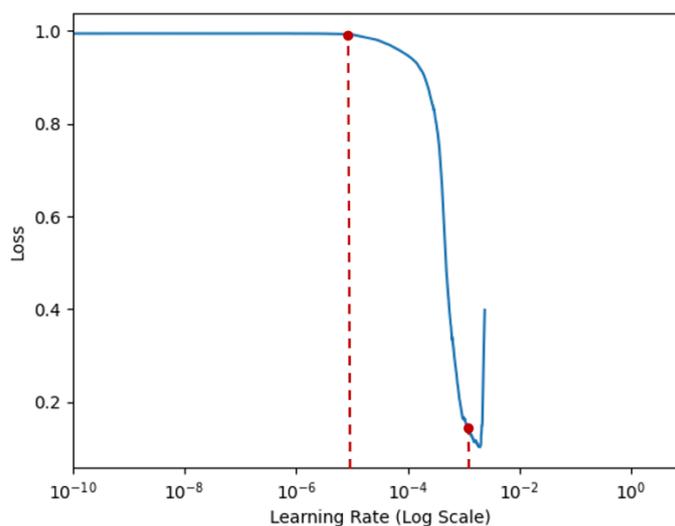


Figura 35. Pérdidas en función de la tasa de aprendizaje para localizar los límites óptimos de CLR. max_{lr} : las pérdidas empiezan a aumentar, $base_{lr}$: las pérdidas comienzan a bajar.

Capítulo 3. Materiales y métodos

En nuestro caso, fijamos $max_{lr} = 1e^{-3}$ (las pérdidas empiezan a aumentar) y $base_{lr} = 1e^{-5}$ (las pérdidas comienzan a bajar).

Una vez que los límites de la técnica de CLR fueron definidos y fijados, se hizo el entrenamiento completo de la red variando la tasa de aprendizaje de manera lineal entre el mínimo ($base_{lr}$) y el máximo (max_{lr}), como refleja la Figura 36. De este modo, se reducía tanto el trabajo como el tiempo que supone ajustar una tasa de aprendizaje óptima para el modelo a la vez que se mejoraba el rendimiento de este.

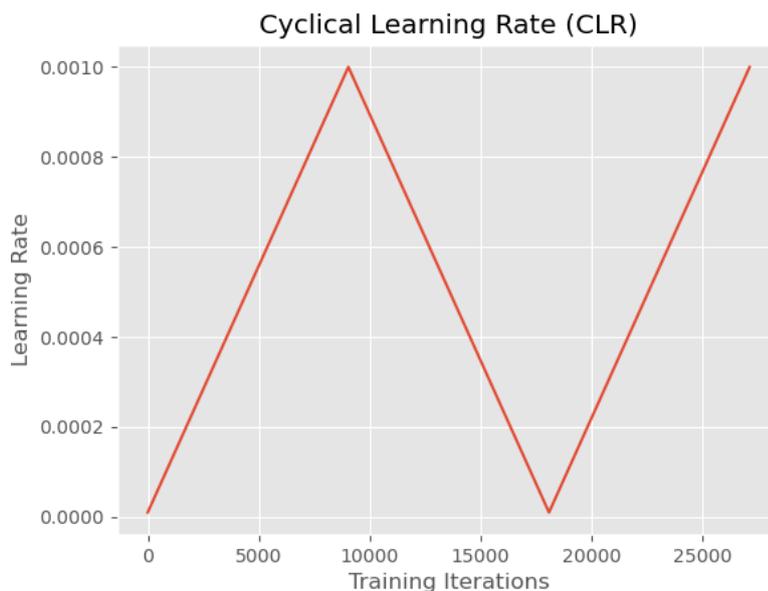


Figura 36. Tasa de aprendizaje variando entre los límites establecidos. 'Training iterations' es el número de actualizaciones de pesos por época, es decir, el número total de muestras de entrenamiento dividido por el tamaño de batch.

En definitiva, mediante CLR variábamos la tasa de aprendizaje en ciclos durante el entrenamiento lo que nos permitió una exploración más amplia del espacio de búsqueda, evitaba el estancamiento en mínimos locales y aceleraba la convergencia del modelo. Esta técnica ha demostrado ser efectiva en una variedad de problemas de aprendizaje profundo y ha mejorado el rendimiento de los modelos en términos de precisión y eficiencia de entrenamiento.

Métrica en entrenamiento: Coeficiente de Correlación de Matthew

Para evaluar problemas de clasificación binaria como al que nos enfrentamos (vóxeles pertenecientes a la médula cervical o no), en esta tesis, se empleó el MCC, descrito en la sección

Capítulo 3. Materiales y métodos

1.3.4.3, ya que es una medida más segura, que no se ve afectada por el problema de desbalanceo de datos, pues sólo produce una calificación alta si se tienen buenos resultados en la predicción en las cuatro categorías de las que está compuesta la matriz de confusión (verdaderos positivos, falsos negativos, verdaderos negativos y falsos positivos) y proporcionalmente tanto al tamaño de los elementos positivos como al tamaño de los elementos negativos del conjunto de datos.

Es por ello que el MCC se utilizó como métrica en este modelo, al ser una métrica óptima y sólida para evaluar el rendimiento de modelos de clasificación binaria que se enfrentan a desequilibrios en los datos.

Métrica en test

Para evaluar el rendimiento del modelo entrenado en el conjunto de datos de test, principalmente se utilizó el coeficiente DICE, también conocido como índice de similitud de Sørensen-Dice, ya que es ampliamente conocida su aplicabilidad y buena utilidad en la evaluación de algoritmos de segmentación, especialmente en tareas de segmentación binaria, donde se trata de separar píxeles de fondo y objetos de interés. El coeficiente DICE mide la similitud entre el conjunto de píxeles predichos por el modelo y el conjunto de píxeles del *ground truth*. Al comparar la superposición de estas dos segmentaciones, el coeficiente DICE proporciona una medida cuantitativa de cuán bien el modelo de segmentación se ajusta a la máscara de referencia.

Pese a ser una de las métricas más recomendadas y empleadas en el área, calculamos también algunas de las métricas introducidas en la sección 1.3.2.2.6. con el fin de obtener una visión más completa de cómo se comporta el modelo en diferentes aspectos y entender mejor sus fortalezas y debilidades.

3.7 Biomarcadores de imagen: cálculo de métricas espaciales de la médula cervical

Con el fin de proporcionar la posibilidad de llevar en los pacientes de EM un seguimiento y estudiar la progresión de la atrofia de la médula espinal cervical, el método propuesto facilita, como biomarcador de imagen, el cálculo de la volumetría de la segmentación obtenida de manera automática.

Para realizar este cálculo es necesario consultar la cabecera de la segmentación en formato Nifti de la que se quiere calcular la volumetría. A esta cabecera se le asigna la misma información que su correspondiente imagen de datos reales original, tras el proceso de segmentación.

Capítulo 3. Materiales y métodos

En esa cabecera se consulta el campo `.get_data_shape()` para conocer las dimensiones de la imagen y el campo `.get_zooms()` para conocer las dimensiones del vóxel utilizando el siguiente código:

```
dimensiones_imagen = header.get_data_shape()
ancho_imagen = dimensiones_imagen[0]; alto_imagen = dimensiones_imagen[1]
dimensiones_voxel = header.get_zooms()
voxel_x = dimensiones_voxel[0]; voxel_y = dimensiones_voxel[1]; voxel_z
        = dimensiones_voxel[2]
```

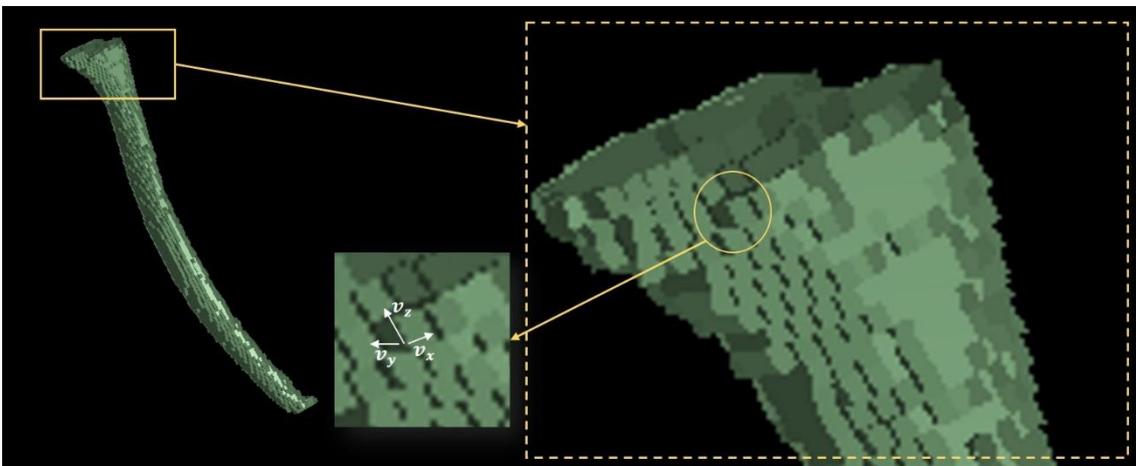


Figura 37. Representación de vóxel perteneciente a la máscara de segmentación y sus 3 componentes dimensionales.

Dichas dimensiones se pueden visualizar explicativamente en la Figura 37.

Conocer las dimensiones de la imagen es necesario para saber si ésta ha sufrido redimensionado durante la etapa de preprocesado de datos. Si ha sufrido redimensionado (sus dimensiones actuales no coinciden con la de la cabecera), su segmentación se devuelve a su tamaño original mediante interpolación con *nearest-neighbour* antes de calcular la volumetría. Si no ha sufrido redimensionado (las dimensiones actuales coinciden con las de la cabecera), no hay que volver a redimensionar a su tamaño original. Esto se realiza así para poder utilizar la información de tamaño de vóxel de la cabecera, ya que esta información es específica para el tamaño de imagen que se indica en la misma cabecera y no es válida tras aplicar redimensionado, pues se estaría considerando un tamaño de vóxel demasiado grande (en el caso de que la imagen se haya redimensionado a un tamaño mayor) o demasiado pequeño (en el caso de que la imagen se haya redimensionado a un tamaño menor).

Capítulo 3. Materiales y métodos

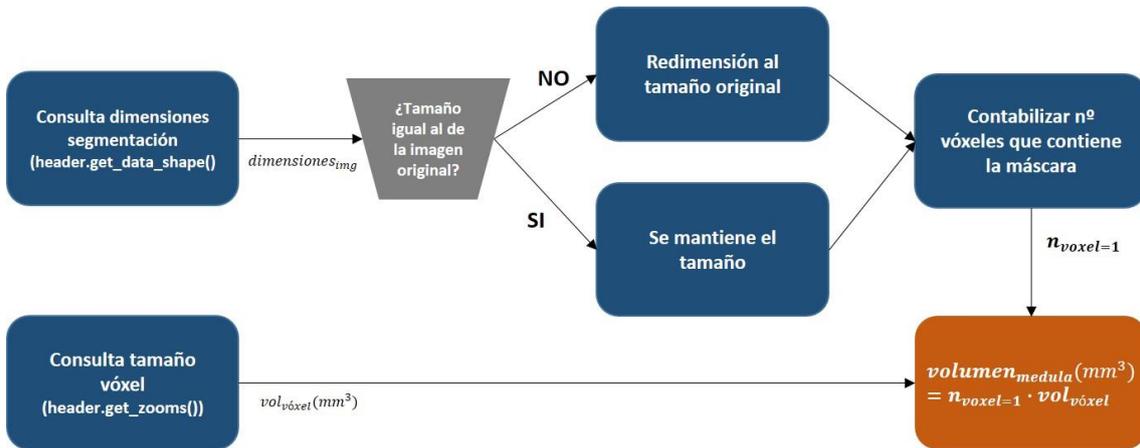


Figura 38. Diagrama de flujo para calcular la volumetría de la segmentación de médula espinal cervical.

Una vez que se tiene la segmentación con las dimensiones de su imagen *real world* original, se puede realizar el cálculo de la volumetría de la médula espinal cervical. Para ello, se contabilizan los vóxeles con valor distinto a cero ($n_{vóxeles=1}$) y este valor se multiplica por el volumen de un vóxel individual, siguiendo la siguiente expresión:

$$volumen_{medula}(mm^3) = n_{vóxeles=1} \cdot voxel_x \cdot voxel_y \cdot voxel_z \quad (29)$$

Siendo $voxel_x$, $voxel_y$ y $voxel_z$, respectivamente, el tamaño en mm de la componente horizontal, vertical y de profundidad de las dimensiones de vóxel de la secuencia en la que nos encontramos calculando el volumen, tal como podemos observar en la Figura 37.

Con el objetivo de obtener una métrica más homogénea entre las mediciones de las distintas médulas segmentadas automáticamente en la población componente de nuestro estudio, con nuestro método calculamos también la sección media transversal en mm^2 de cada segmentación, para salvar las divergencias de longitud de la médula cervical segmentada, pues en cada paciente que conforma nuestro estudio, esta longitud es distinta. Este cálculo se realiza a partir del volumen en mm^3 y teniendo en cuenta el número de cortes (n_{slices}) que compone cada segmentación y su espesor ($slice_{thickness}$), es decir, dividiendo el volumen entre la longitud total de la segmentación de manera que la sección media de cada médula espinal cervical segmentada se calcula mediante la siguiente expresión:

$$sección_{media}(mm^2) = \frac{volumen_{médula}}{longitud_{médula}} = \frac{volumen_{medula}}{(n_{slices} \cdot slice_{thickness})} \quad (30)$$

Capítulo 4. Resultados y discusión

Capítulo 4. Resultados y discusión

En este capítulo, analizamos los resultados obtenidos mediante el método de segmentación automática de médula espinal desarrollado, tanto en entrenamiento, en test, como en validación, de forma cuantitativa y cualitativa. Comparándolos con los resultados de las arquitecturas del estado del arte. Para finalizar mostrando la relación entre las dimensiones de segmentación, formas clínicas y la EDSS en EM.

4.1 Resultados en entrenamiento

Como se ha explicado en capítulos anteriores, el método desarrollado consiste en una arquitectura *Residual Attention-Aware U-Net* en la que empleamos Focal Tversky como función de pérdidas y el ajuste de tasa de aprendizaje se lleva a cabo mediante la técnica de *Cyclic learning rate*. De esta forma, conseguimos un método automático y preciso.

En la Figura 39, vemos la evolución del MCC y el valor calculado por la función de pérdidas en cada época del entrenamiento.

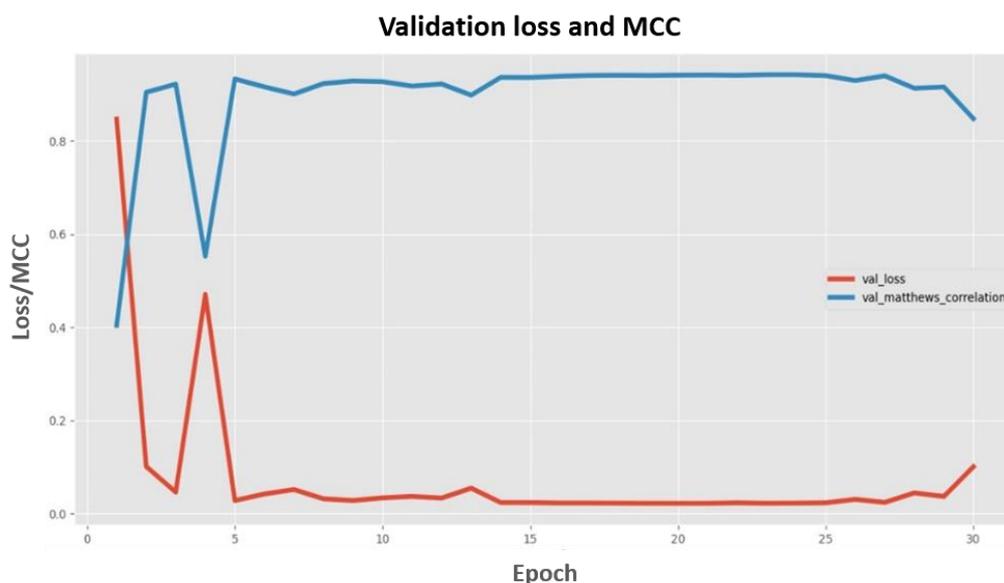


Figura 39. Pérdidas y MCC del set de validación durante el proceso de entrenamiento.

La Figura 40 muestra la comparación de los valores DICE que obtenemos utilizando una tasa de aprendizaje fija con un valor de $1e^{-4}$ y los que obtenemos empleando el método de tasa de aprendizaje cíclica (CLR). Podemos observar, cómo utilizando la CLR alcanzamos en menos épocas el máximo valor DICE, unas diez épocas antes que fijando la tasa de aprendizaje.

Capítulo 4. Resultados y discusión

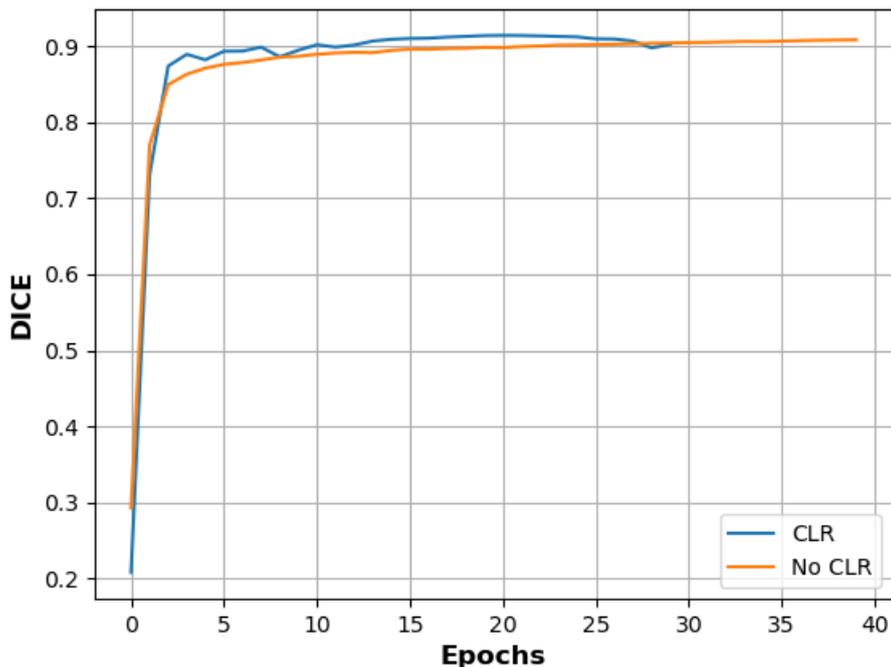


Figura 40. Evolución del valor DICE durante las épocas del entrenamiento utilizando CLR y una tasa de aprendizaje fija de $1e^{-4}$. Con CLR se alcanza en menos épocas el valor mayor DICE.

Conseguimos, en entrenamiento, un MCC de 0.95 y un coeficiente medio DICE de 0.9 como métrica de validación estadística. En la Figura 41 presentamos la distribución de coeficientes DICE generados durante la fase de entrenamiento. En ella, podemos observar un histograma estrecho y centrado en valores altos de coeficiente DICE, lo que indica que, la mayoría de las máscaras predichas por el modelo entrenado, presentan una gran similitud con su *ground-truth*. Los valores DICE cercanos a 0, pueden ser ocasionados por factores humanos, como el hecho de que el radiólogo haya realizado una segmentación manual de menor longitud, incluyendo menos *slices* que la generada por el modelo propuesto, o viceversa, de manera que, en este primer caso, el *ground-truth* contendría una máscara sin valores, mientras que nuestro modelo todavía seguiría encontrando médula espinal y generando su máscara.

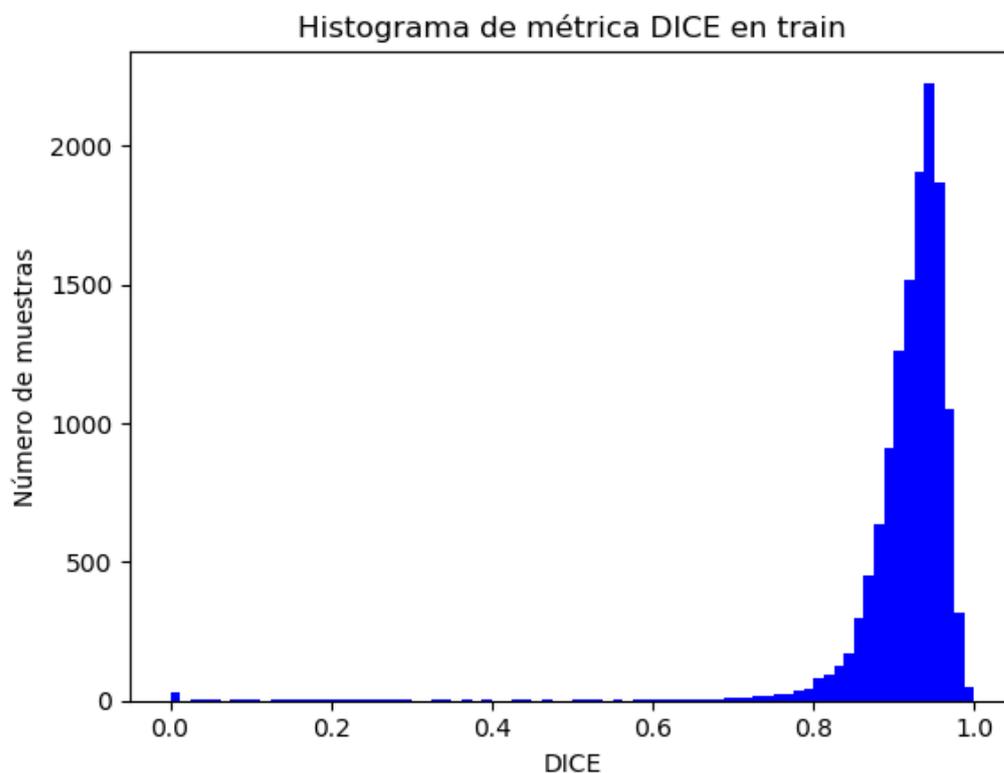


Figura 41. Histograma de coeficientes DICE generados durante el entrenamiento.

4.2 Resultados en test

El modelo entrenado fue validado con el 20% del conjunto de datos, denominándolo como conjunto de datos de test. En la Tabla 7 se muestran los resultados obtenidos con diferentes métricas.

Tabla 7: Resultados de métricas en el conjunto de datos de test para validación. Media \pm STD: media y desviación estándar. P25: percentil 25. P50 (mediana): percentil 50. P75: percentil 75.

	Media \pm STD	P25	P50 (mediana)	P75
DICE	0.904 \pm 0.101	0.90	0.92	0.94
IoU	0.77 \pm 0.09	0.74	0.78	0.82
Sensibilidad	0.777 \pm 0.099	0.75	0.79	0.83
Especificidad	0.98 \pm 0.07	0.99	1	1
Precisión	0.9999 \pm 0.0004	1	1	1

Capítulo 4. Resultados y discusión

El coeficiente DICE es una métrica que logra un equilibrio entre la precisión (proporción de píxeles correctamente predichos) y la sensibilidad (proporción de píxeles correctamente detectados). Esto es importante en tareas de segmentación de imagen, ya que es fundamental obtener tanto una buena delimitación de los objetos como evitar errores de predicción.

En cuanto a las métricas IoU y sensibilidad, observamos que, cuantitativamente, obtenemos valores muy similares, lo que se debe principalmente a la expresión que define el cálculo de estas métricas, lo que podemos recordar en la 1.3.2.2.6. Pese a ello, dichas métricas miden aspectos diferentes del rendimiento del modelo de segmentación. Mientras que el IoU se centra en la superposición entre la predicción y la máscara de referencia, la sensibilidad mide la capacidad del modelo para detectar correctamente los píxeles de la máscara de referencia que han sido correctamente detectados. En ciertas situaciones de segmentación de imágenes, estas dos métricas pueden ser semejantes debido a las características de las máscaras de referencia y las predicciones del modelo. Si el modelo tiene un bajo número de falsos positivos (es decir, pocos píxeles erróneamente clasificados como positivos), existirá una alta similitud entre IoU y sensibilidad.

En cuanto a la especificidad, en tareas de segmentación de imagen médica, indica la capacidad del modelo para detectar la clase perteneciente al fondo de una imagen. Debido a la gran cantidad de píxeles etiquetados como fondo a comparación de los pertenecientes a médula espinal, los rangos de especificidad cercanos a 1 como el obtenido (0.98) son el estándar y los esperados. Es por ello que la especificidad, puede ser una buena métrica para asegurar la funcionalidad del modelo, pero no para su rendimiento.

Capítulo 4. Resultados y discusión

Debido a la importancia que se les da a los verdaderos negativos en su definición, la precisión nos proporciona una puntuación alta errónea (0.9999) que no se corresponde con la realidad.

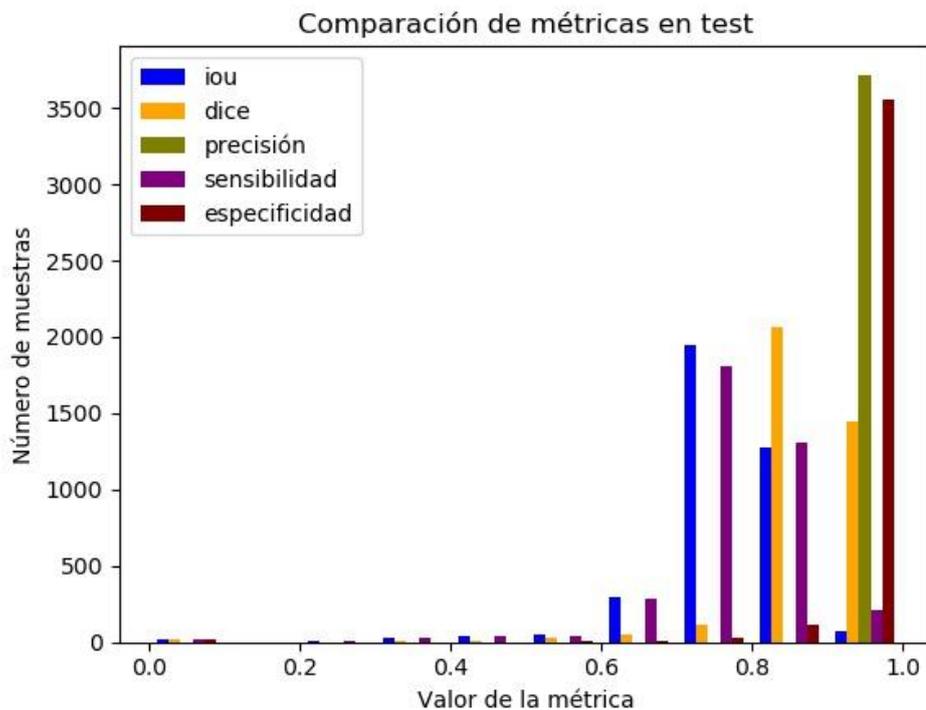


Figura 42. Representación comparativa de distribución de valores de las distintas métricas de evaluación utilizadas.

Es importante tener en cuenta que cada métrica tiene su propio propósito y contexto de aplicación y pueden ser más relevantes o útiles dependiendo de los objetivos específicos del problema de segmentación. Para nuestro objetivo, el coeficiente DICE fue el elegido como apoyo primario ya que es capaz de reflejar de manera precisa la calidad de la superposición entre las segmentaciones y nos proporciona un equilibrio entre precisión y sensibilidad, fundamental para obtener tanto una buena delimitación de los objetos como evitar errores de predicción. Esto se puede observar en los valores de P25, P50 y P75 de la Tabla 7, especialmente en los del parámetro DICE, los cuales son cercanos a 1, indicando que los valores de dicho parámetro tienen una dispersión baja y se encuentran en un rango relativamente estrecho y preciso.

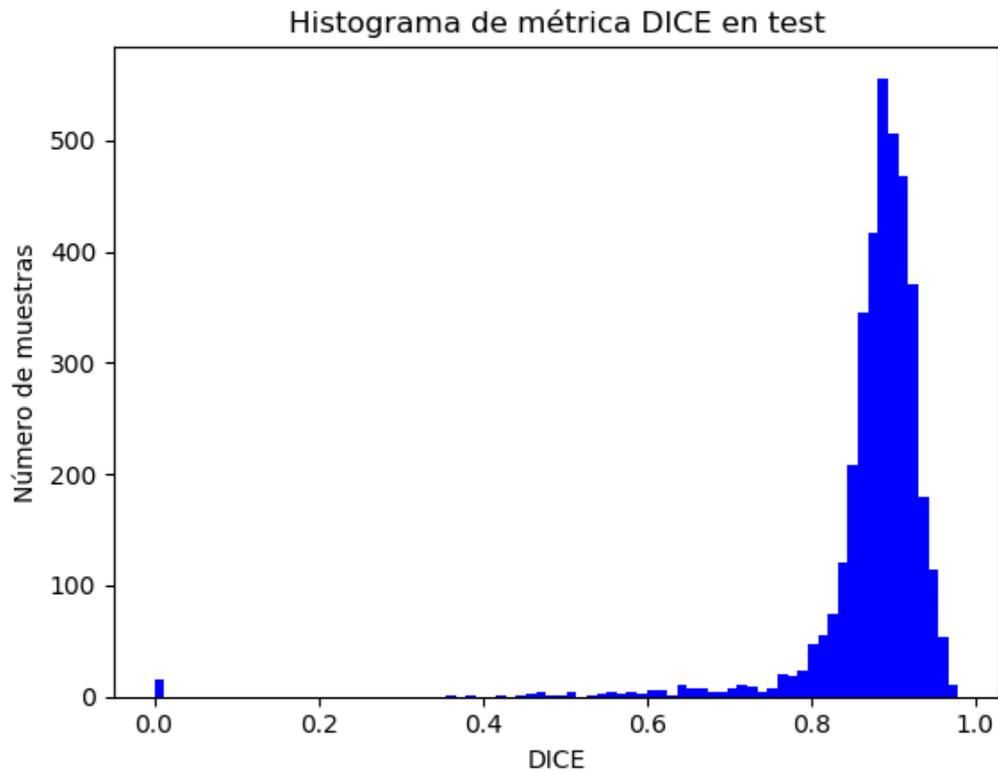


Figura 43. Histograma de coeficientes DICE en test.

4.3 Resultados en validación

La validación en un modelo de IA es fundamental para garantizar su robustez, capacidad de generalización y rendimiento en el mundo real. En nuestro estudio realizamos una validación interna con casos de pacientes de EM del Hospital Universitario y Politécnico de la Fe cuyas adquisiciones fueron adquiridas 5 años después de los utilizados para generar el modelo, tal como se describió en la sección 3.1.

Como podemos observar tanto en la Tabla 8 como en la Figura 44, obtenemos unos valores similares a los resultantes en las métricas de test, valores notablemente altos en precisión y especificidad y valores más fidedignos en DICE, IoU y sensibilidad.

Capítulo 4. Resultados y discusión

Tabla 8: Resultados de métricas en el conjunto de datos de validación. Media \pm STD: media y desviación estándar. P25: percentil 25. P50 (mediana): percentil 50. P75: percentil 75.

	Media \pm STD	P25	P50 (mediana)	P75
DICE	0.93 \pm 0.04	0.90	0.93	0.95
IoU	0.847 \pm 0.07	0.81	0.85	0.89
Sensibilidad	0.86 \pm 0.08	0.82	0.87	0.91
Especificidad	0.98 \pm 0.02	0.97	0.99	1
Precisión	0.9999 \pm 0.0002	1	1	1

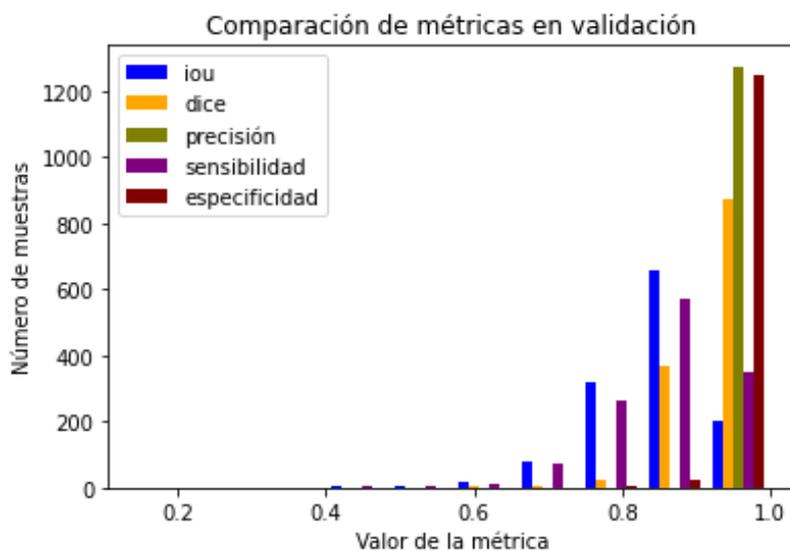


Figura 44: Representación comparativa de distribución de valores de las distintas métricas de evaluación utilizadas en validación.

En la Figura 45 se muestra el histograma de la distribución de coeficientes DICE obtenidos en el conjunto de datos de validación, el cual contiene un total de 1274 imágenes, de las cuales, observamos que la mayoría presentan un coeficiente DICE mayor a 0.9 y por lo tanto, las segmentaciones predichas son muy fidedignas a las delineadas manualmente por los expertos.

Capítulo 4. Resultados y discusión

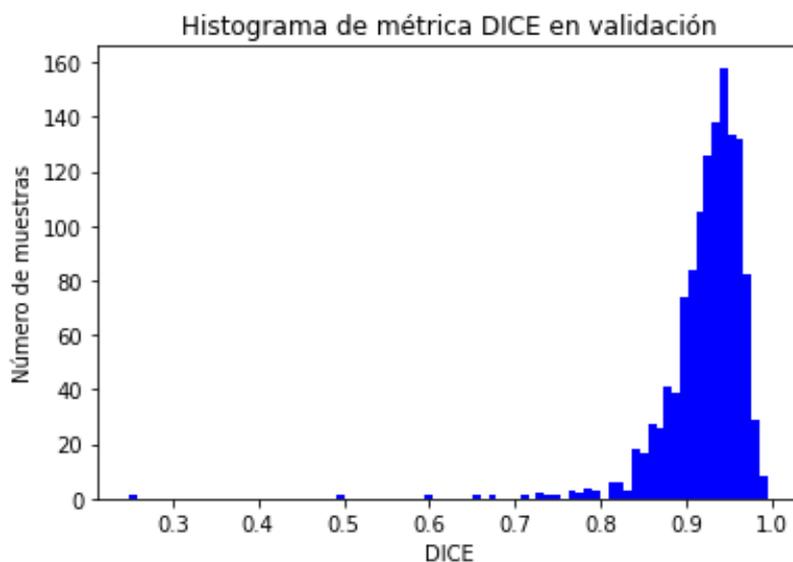


Figura 45. Histograma de distribución de coeficientes DICE obtenidos en validación.

4.4 Ejemplos de resultados cualitativos

A continuación, mostramos algunos ejemplos de resultados cualitativos de las segmentaciones de médula espinal cervical obtenidas, con el fin de mostrar una mejor visualización del rendimiento real del modelo propuesto.

En relación a ello, previamente podemos extraer algunas conclusiones, en concreto, que nuestro método de segmentación automática proporciona un resultado preciso y satisfactorio que apenas difiere de la experimentada delineación manual realizada por el radiólogo, tal como se puede deducir al observar el histograma DICE ya sea de entrenamiento, test o validación (Figura 41, Figura 43, Figura 45), donde la mayoría de las imágenes que componen tales conjuntos de datos presentan un coeficiente DICE mayor a 0.85, muestra de gran similitud entre la segmentación generada y la elaborada por los radiólogos.

En la Figura 46, se muestra en varias perspectivas la misma médula 3D. En verde aparecen los vóxeles coincidentes entre herramienta automática y radiólogo y en rojo los que no. En este caso observamos que la mayoría de vóxeles no coincidentes es debido a la consideración de la longitud de la segmentación, pues la herramienta detecta médula espinal y segmenta unos pocos cortes de imagen más por arriba y por abajo, pero manteniendo la homogeneidad y pertenencia a la médula cervical.

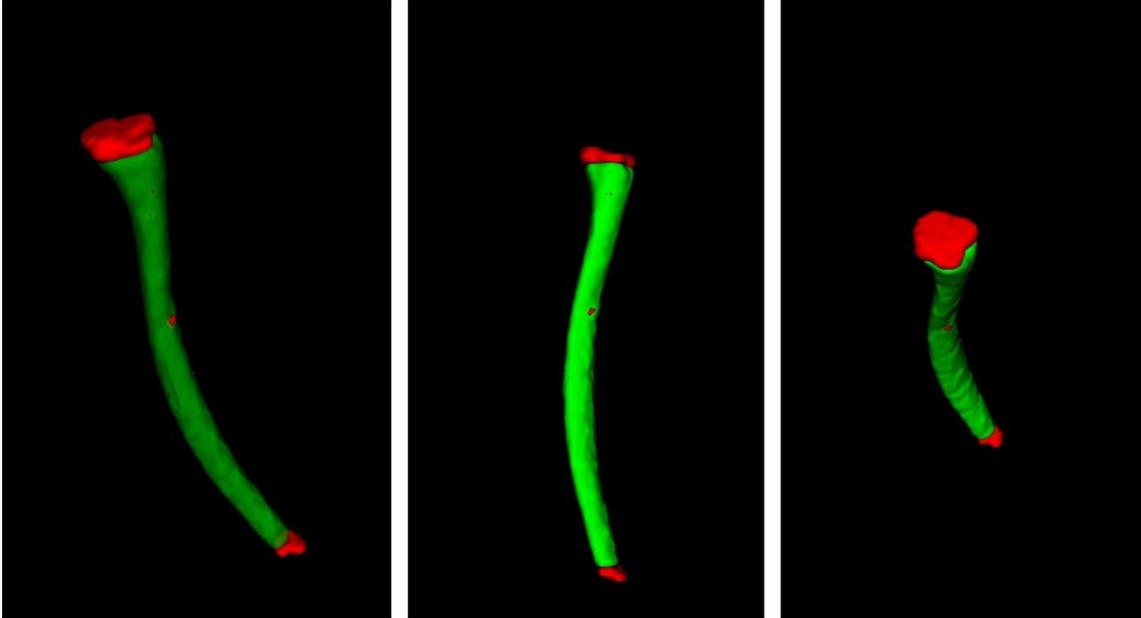


Figura 46. Médula cervical en 3D en varias perspectivas. En verde los vóxeles coincidentes entre herramienta propuesta y radiólogo. En rojo los no coincidentes, la mayoría por mayor longitud de segmentación de la porción medular.

En la Figura 47 podemos ver en un ejemplo del rendimiento del modelo en una imagen axial aleatoria de las numerosas que componen una segmentación de médula cervical completa. El cuadrado rojo indica la región que es aumentada en el resto de las imágenes para observar más claramente los detalles. En la Figura 47a observamos la superposición del *ground truth* (máscara etiquetada manualmente) sobre la imagen real, mientras que en la Figura 47b observamos la máscara predicha automáticamente por el modelo. Con el fin de comparar mejor estas dos máscaras, observamos la superposición de ambas en la Figura 47d. Los píxeles en rojo indican área no coincidente entre ambas máscaras (píxeles que solo están presentes en una de las dos) y los representados en verde indican áreas coincidentes (píxeles que están presentes tanto en la máscara generada manualmente como por el método automático).

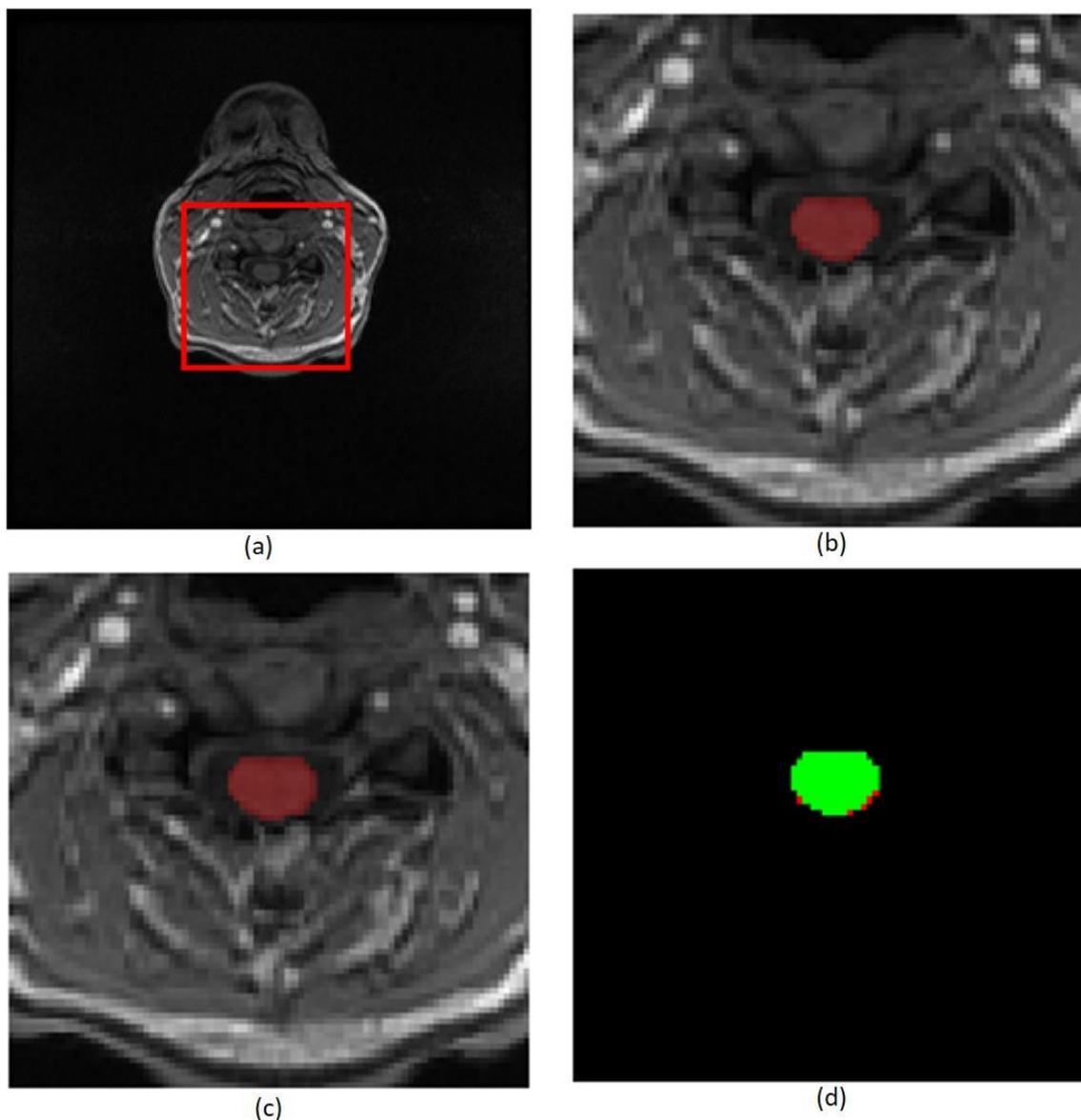


Figura 47. Ejemplo de segmentación en plano transversal con DICE alto. a) Real world data, b) ground truth, c) máscara predicha por el modelo propuesto, d) superposición entre ground truth y máscara predicha (en verde zonas coincidentes, en rojo zonas no coincidentes).

Como hemos mencionado, la imagen mostrada en la Figura 47 es sólo un ejemplo aleatorio de entre más de 2000 imágenes axiales que componen el conjunto de datos de validación.

En la Figura 48 mostramos una visualización explicativa similar a la anterior, pero esta vez con una imagen con un menor DICE (0.55). Como observamos, en esta imagen de RM se observa el artefacto de ruido bias-field explicado en la sección 1.2.1.1.3, lo que dificulta la detección de píxeles pertenecientes a la médula espinal debido a la mala diferenciación de intensidad entre las distintas regiones.

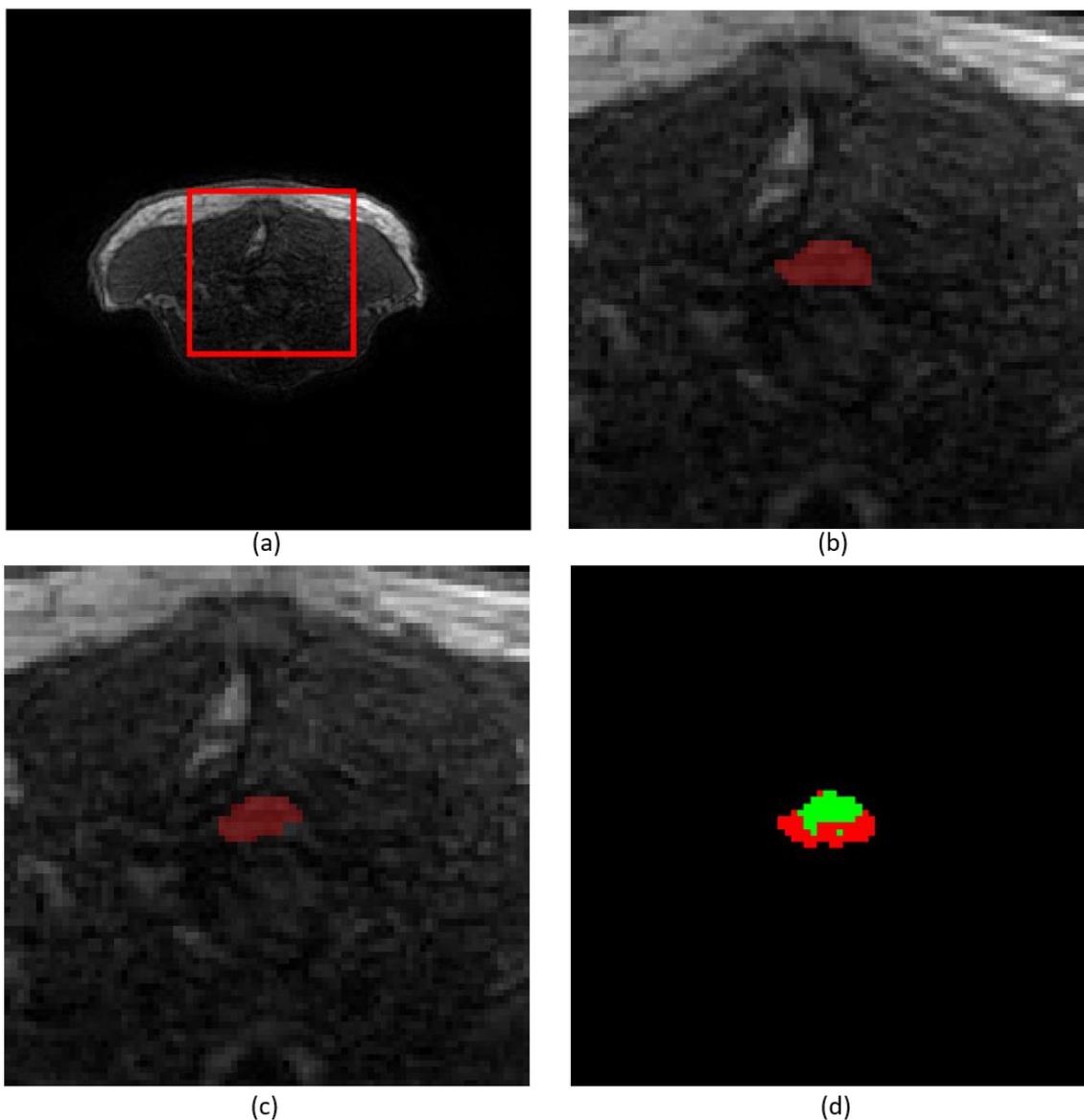


Figura 48. Ejemplo de segmentación en plano transversal con DICE bajo. Ejemplo de segmentación con DICE bajo. a) Real world data, b) ground truth, c) máscara predicha por el modelo propuesto, d) superposición entre ground truth y máscara predicha (en verde zonas coincidentes, en rojo zonas no coincidentes).

Continuando con la muestra de resultados, en la vista en 3D incluida en la Figura 49, observamos dos inhomogeneidades rodeadas en amarillo y en morado. Estas inhomogeneidades aparecen como saliente y un entrante en el volumen, respectivamente, y están formadas por las imágenes que están contenidas en las cajas de los mismos colores. En amarillo, observamos como para las imágenes axiales 2D que componen el saliente, nuestro modelo predijo una máscara que no se corresponde del todo con la médula, siendo un poco más grande el *ground truth* debido a la extrema similitud de intensidades entre los píxeles de la médula y de la zona circundante, causado posiblemente por efectos de volumen parcial.

Capítulo 4. Resultados y discusión

En morado se recoge la problemática contraria, este entrante que se observa en la vista 3D fue debido a que el modelo predijo una máscara menor a la que le corresponde en realidad. Si nos fijamos en la imagen real, la médula apenas se puede distinguir de entre la escala de grises de las demás estructuras que componen la imagen, por lo que la problemática para nuestro modelo era de una gran complejidad. En ambas heterogeneidades, el fallo es debido a la presencia de estructuras cercanas de una similitud de intensidad tan semejante a la médula que parecen estar conectadas.

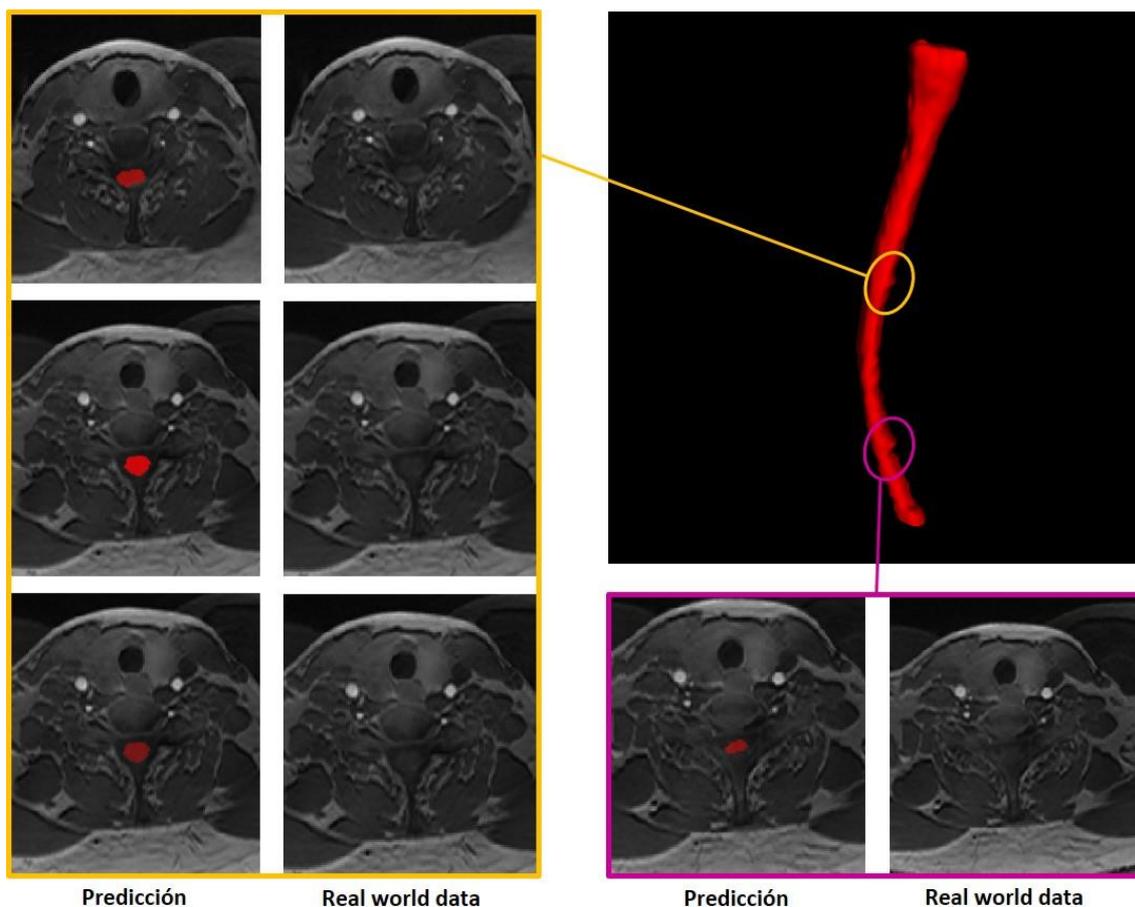


Figura 49. Ejemplo de menor precisión en la segmentación. En amarillo y en morado se indican las heterogeneidades provocadas por la mala diferenciación en la imagen entre la médula y las estructuras colindantes.

En la Figura 50 observamos un análisis visual de uno de los casos contenidos en el conjunto de datos de validación. En este caso y de manera similar al ejemplo de la Figura 46, la herramienta segmenta unos pocos cortes de imagen más por arriba y por abajo (indicado en rojo) con respecto a la segmentación manual (indicado en verde), aunque manteniendo la homogeneidad y pertenencia a la médula cervical. En esta misma figura e indicado por una circunferencia azul, observamos una diferencia entre segmentación manual y automática, en la que esta última incluye unos píxeles más a la máscara resultante, tal como podemos ver en la Figura 50b y en su

Capítulo 4. Resultados y discusión

ampliación en la Figura 50b1. Indicado por una circunferencia, esta vez en amarillo, observamos una imagen perteneciente a la extensión longitudinal que realiza la segmentación automática frente a la manual. En la Figura 50c observamos un corte de los que componen esta extensión con la máscara superpuesta a la imagen RM y en la Figura 50c1 el mismo corte transversal, pero sin la máscara superpuesta, con el fin de mejorar la visualización de la región incluida en la segmentación por la herramienta automática.

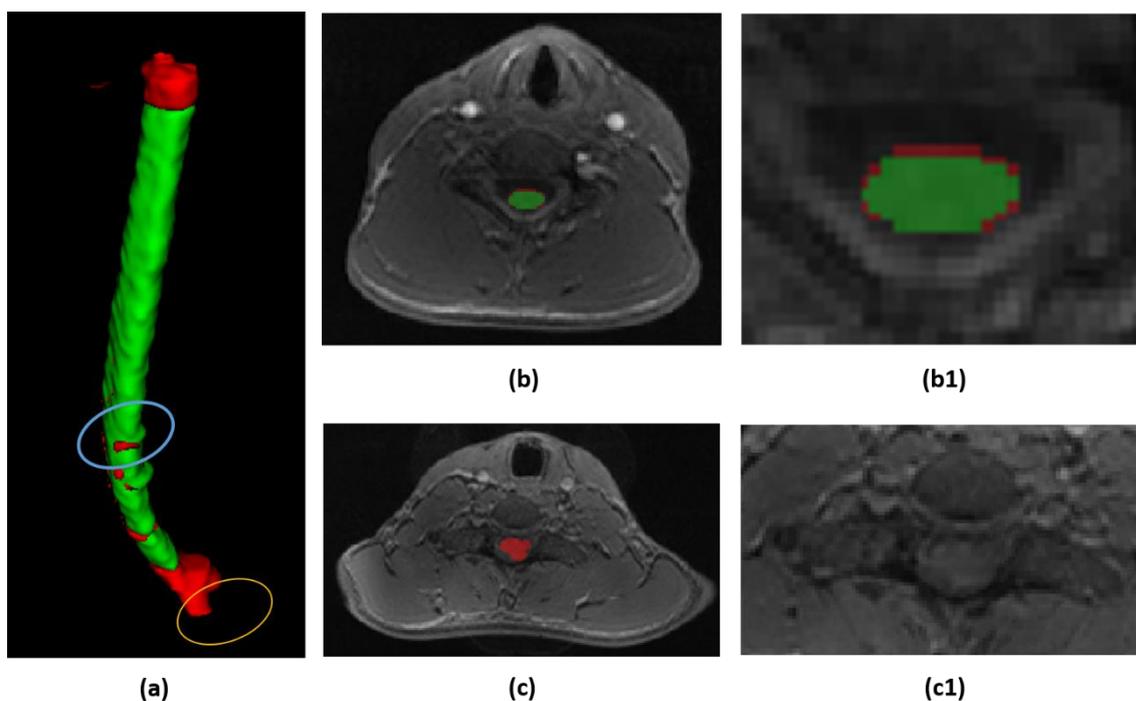


Figura 50. Análisis de caso de validación. a) Segmentación en 3D, b) Corte transversal en el que difieren segmentación manual y automática, b1) Ampliación de b), en rojo los píxeles que añade la segmentación automática frente a la manual, c) Corte transversal perteneciente a mayor longitud de segmentación, c1) Ampliación de c).

Siguiendo con el mismo caso de validación, en la Figura 51 analizamos la pequeña región que la herramienta automática segmenta por error en la zona superior, indicado en amarillo. En este ejemplo, vemos como la segmentación manual y automática coinciden hasta el corte visualizado en la Figura 51b, a partir de ahí, la herramienta sigue identificando píxeles como pertenecientes a región de médula espinal en una serie de imágenes (Figura 51c, Figura 51d, Figura 51e), hasta que llega a un corte (Figura 51f) en el que detecta dos zonas diferenciadas, una de ellas que sigue la referencia espacial de las máscaras de los cortes anteriores (indicada en azul) y otra, señalada en amarillo, donde detecta la estructura del diente del axis como si también fuese médula espinal.

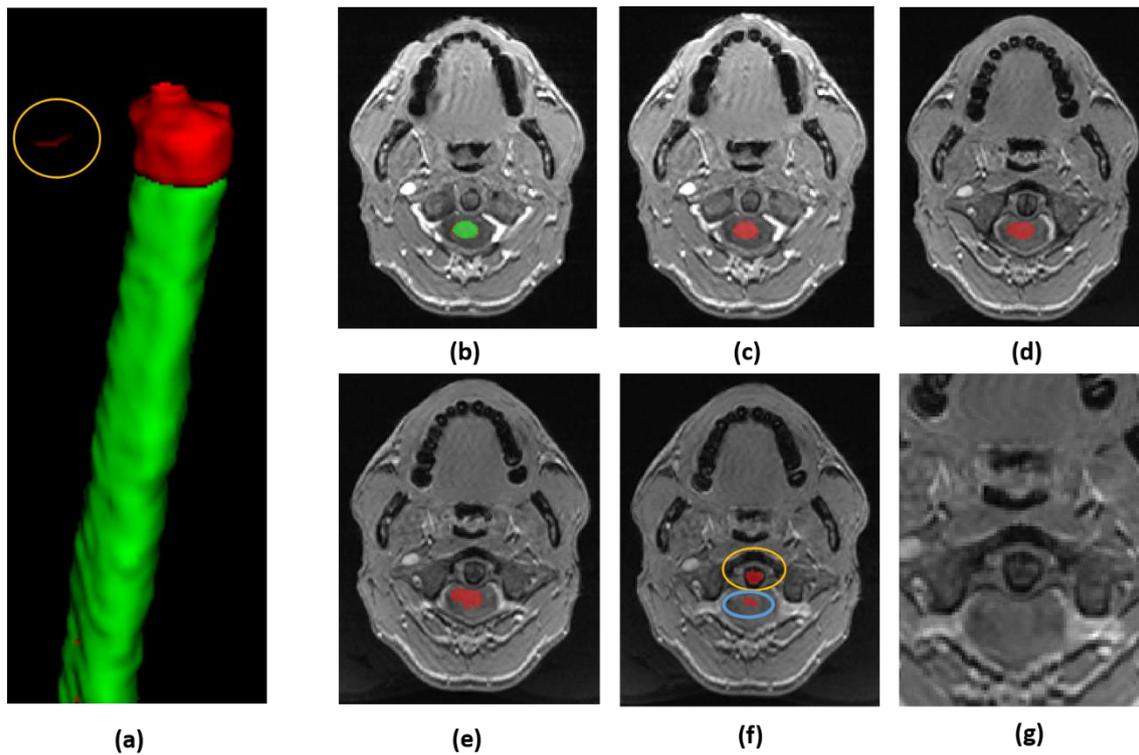


Figura 51. Análisis de caso de validación. a) Segmentación 3D, error de la herramienta indicado en amarillo, b) Último corte superior en el que coinciden segmentación manual y herramienta, c)-e) extensión longitudinal de cortes pertenecientes a la segmentación manual, f) corte con zona errónea, g) mismo corte sin superposición de máscara.

En la Figura 52 podemos analizar otro caso perteneciente al conjunto de validación, en el cual, al igual que en los ejemplos anteriores, encontramos dos patrones de diferencias entre segmentación manual y automática: indicado en azul, diferencia por prolongación de cortes (Figura 52b y Figura 52b1) e, indicado en amarillo, diferencia en unos pocos píxeles integrados como máscara (Figura 52c y Figura 52c1).

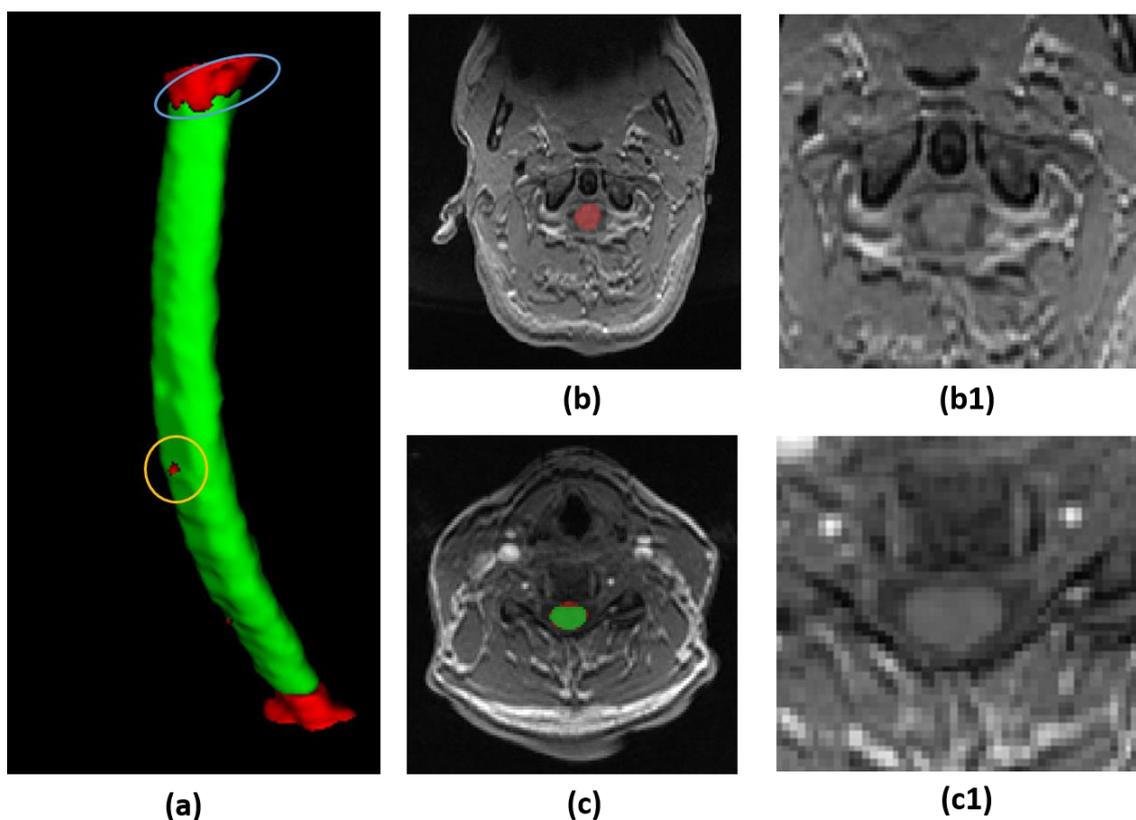


Figura 52. Análisis de caso de validación. a) Segmentación en 3D, b) Corte transversal perteneciente a extensión de longitud de segmentación, b1) Ampliación de b), c) Corte transversal en el que difieren segmentación manual y automática, en rojo los píxeles que añade la segmentación automática frente a la manual, c1) Ampliación de c.

Siguiendo con la visualización de resultados obtenidos dentro del conjunto de datos de validación, en la Figura 53 se muestra una de las segmentaciones que más difieren entre método automático y *ground truth*. Destacamos dos errores como ejemplo. En la Figura 53a, señalamos rodeado en amarillo cómo la herramienta automática genera en un corte una máscara en la que, además de la médula espinal, detecta la vértebra perteneciente a ese nivel. En la Figura 53b, mostramos cómo la segmentación automática añade más píxeles a la máscara que la manual y cómo se observa este efecto en la representación 3D, rodeado en color azul. En la imagen ampliada de la ROI, se observa como esta confusión está dentro de unas condiciones controladas en las que, incluso la subjetividad humana de los expertos podría llevar a un efecto similar.

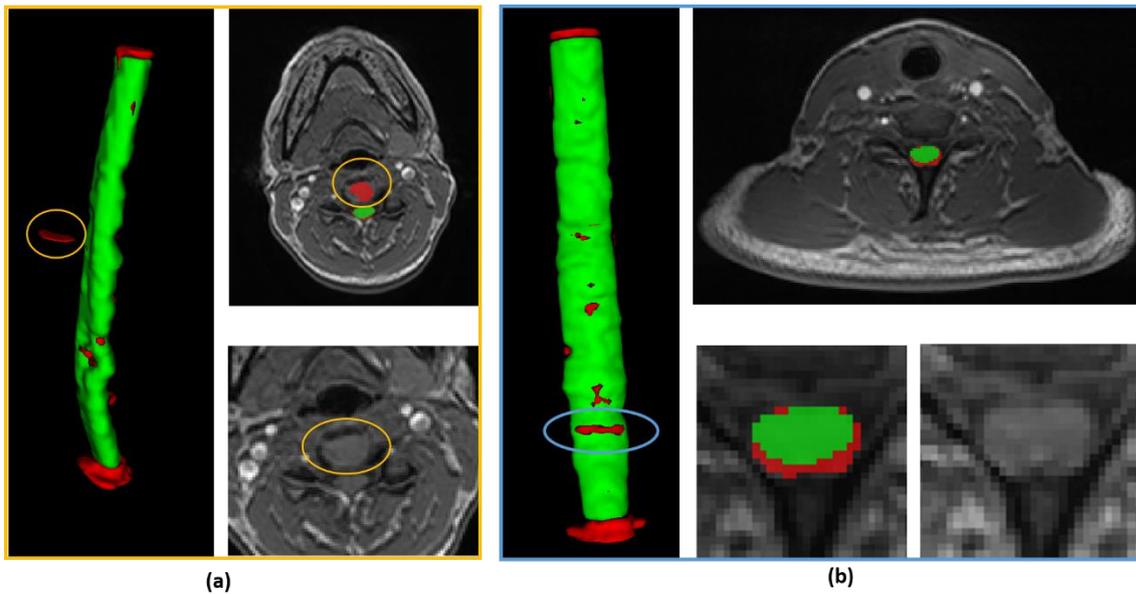


Figura 53. Análisis de caso de validación. a) Confusión de médula con vértebra por la herramienta automática, b) mayor adición de píxeles a la máscara por la segmentación automática.

Pese a casos como el descrito anteriormente, en el que la precisión de la herramienta puede disminuir debido a problemáticas de artefactos de imagen comunes en las imágenes de RM o a confusiones con otras estructuras (en contadas ocasiones), el modelo propuesto cumple con su objetivo y presenta un buen rendimiento.

La segmentación manual de la médula espinal cervical es una tarea laboriosa para los radiólogos, por lo que el poder contar con una herramienta que les facilite esta tarea será de gran utilidad y una importante base para aspectos del análisis médico como la cuantificación de la atrofia en la médula espinal cervical mediante el cálculo de la volumetría, ya que se ha comprobado que la reducción del volumen tiene una relación directa con la atrofia, tal como veremos en la próxima sección 4.6.

4.4.1 Discusión del método de segmentación propuesto

Esta herramienta fue desarrollada con el objetivo de solventar algunas deficiencias que presentaban otros modelos existentes en el estado del arte, como la herramienta JIM, un método basado en superficies (*Surface-based method*) que delinea la médula tras la necesidad de que el usuario marque cuál es el centro de la médula espinal (*Jim 9 Software | Xinapse Systems*, n.d.). Esta herramienta se basa en *Active Surfaced Model (ASM)* y el usuario necesita introducir manualmente marcas en los extremos de la región de estudio y señalar cual es la línea central de la médula. Este método proporciona imágenes sagitales que después se reformatean al plano transversal para obtener cinco *slices* contiguas de 3mm de espesor; la herramienta

Capítulo 4. Resultados y discusión

calcula automáticamente el radio y el centro de la médula en cada imagen axial y el área transversal se obtiene mediante el promediado de estas *slices* contiguas. Otro método semiautomático es *NeuroQLab* (Lukas et al., 2013a; Weiler et al., 2015), un método basado en imagen que segmenta la parte superior de la médula espinal cervical a partir del tejido circundante usando un método de modelos Gaussianos basado en intensidad. Este método requiere de la interacción del usuario, el cual, debe definir la sección de médula que se quiere analizar mediante la inclusión de un plano oblicuo que la debe atravesar. Además, al utilizar un modelo basado en intensidad para la segmentación, este método es propenso a incluir píxeles similares en esta característica, pero pertenecientes a regiones fuera de la de interés.

También se han dedicado esfuerzos al intento de desarrollar métodos totalmente automáticos como *Spinal Cord Toolbox* (SCT) (De Leener et al., 2017). Este método se basa en la secuencia de dos CNN y consigue un coeficiente DICE medio de 0.95. La primera CNN con convoluciones 2D dilatadas, detecta la línea central de la espina dorsal siendo necesaria la interacción humana para inicializar la localización de la estructura (señalar tres puntos en la espina dorsal) para un buen funcionamiento y utilizando la siguiente CNN para la segmentación en una metodología que depende de la calidad de este módulo previo de detección. Existe otro método que a la vez depende del descrito previamente, orientado para la segmentación automática de médula espinal cervical tras el daño por contusión traumática a partir de imágenes de RM axiales potenciadas T2 y cuyo coeficiente DICE es 0.93. Este método (McCoy et al., 2019), utiliza primeramente el SCT para detectar la línea central de la médula espinal y crear una máscara cuadrada a su alrededor. Seguidamente, se aplican tres arquitecturas diferentes de CNN 2D para conseguir la segmentación. Alejándose de nuestro propósito de método generalizable, la cohorte de este estudio proviene de una institución en la que se utiliza siempre el mismo equipo de adquisición de RM con los mismos parámetros, sesgando potencialmente los resultados. Además, el paso de pre-procesado que aplica de recortar la imagen alrededor de la línea central de la médula espinal depende de una tercera herramienta.

En nuestro estudio (Bueno et al., 2022), solventamos lo que consideramos que son limitaciones de otros métodos y con el objetivo de evitar cualquier tipo de técnica de pre-procesado y reducir la necesidad de interacción humana, se implementó una arquitectura de aprendizaje de atención residual 2D basada en U-Net para obtener la segmentación de médula espinal cervical deseada. Elegimos una arquitectura profunda ya que, como es conocido, las redes más profundas tienen mejor capacidad de aprendizaje, aunque también presentan más inconvenientes como el desvanecimiento de gradiente (cuando se lleva a cabo la propagación hacia atrás) y el problema de degradación, que da lugar a peores resultados de entrenamiento.

Capítulo 4. Resultados y discusión

Se utilizan bloques residuales en las capas de la red U-Net, excepto en la primera y la última, para solucionar este problema de desvanecimiento de gradiente a la vez que se mantiene una arquitectura profunda. El mecanismo de atención aprende a centrarse en las localizaciones de la imagen que son relevantes permitiendo la retención de la información procedentes de características originales de la imagen mediante la “rama fuerte” mientras se pone atención a estas localizaciones relevantes pertenecientes a la médula espinal mediante la “rama débil”. El entrenamiento fue diseñado con una función de pérdidas focal basada en el índice Tversky para solventar el problema de desbalanceo de clases, común en segmentación de imagen médica, y un buscador automático de tasa de aprendizaje óptima. Finalmente, se consiguió un coeficiente DICE de 0.904.

Este algoritmo de segmentación automática permitirá la cuantificación del volumen de la médula espinal cervical y será de utilidad para la evaluación y seguimiento de la cantidad de atrofia del paciente. Además, este modelo será la base para el desarrollo de un modelo agregado de detección de lesiones intramedulares a partir de secuencias de imagen STIR que se centrará en la detección de lesiones en la región cervical, ya que se utilizará como primer paso para obtener la localización de la médula espinal cervical en el espacio STIR correspondiente tras aplicar técnicas de registro de imágenes entre ambas secuencias. Esta propuesta en investigación, la podemos encontrar descrita y desarrollada con detalle en el Anexo I: Detección automática de lesiones en médula cervical.

4.5 Comparación de resultados entre nuestro método y otros modelos de segmentación

En esta sección presentamos una comparación, tanto a nivel cuantitativo como cualitativo, del rendimiento de los modelos introducidos en la sección 3.5.1 al procesar nuestro conjunto de datos junto a la configuración de distintas combinaciones de hiperparámetros. De esta manera podemos evaluar nuestro modelo frente a otros existentes en tareas de segmentación y ser capaces de verificar su solidez y eficacia a la vez que entender las limitaciones y áreas de mejora de los modelos estudiados.

En la Tabla 9 se muestra la comparación de nuestro modelo con otras configuraciones entrenadas con nuestros datos y reflejando una evaluación de la segmentación mediante el coeficiente DICE medio y sus correspondientes parámetros estadísticos.

Capítulo 4. Resultados y discusión

Tabla 9. Resultados y comparación entre configuraciones de modelos y funciones de pérdidas en los datos de test.

Modelos: U-Net, RA-U-Net, Deep U-Net. Funciones de pérdidas, TL: Tversky Loss, FTL: Focal Tversky Loss. MCC: Mathews Correlation Coefficient; P25: percentil 25. Median (P50): DICE medio y percentil 50. P75: percentil 75.

Modelo	Función de pérdidas	Parámetros	Learning rate	DICE \pm STD	MCC	P25	Median (P50)	P75
U-Net	TL	$\alpha=0.4, \beta=0.6$	$1e^{-5}$	0.66 ± 0.24	0.68	0.57	0.75	0.84
U-Net	TL	$\alpha=0.3, \beta=0.7$	$1e^{-5}$	0.68 ± 0.24	0.71	0.61	0.77	0.85
U-Net	FTL	$\gamma=1.33, \alpha=0.7, \beta=0.3$	$1e^{-5}$	0.62 ± 0.25	0.64	0.53	0.71	0.81
U-Net	FTL	$\gamma=1.33, \alpha=0.6, \beta=0.4$	$1e^{-5}$	0.68 ± 0.27	0.70	0.62	0.80	0.87
U-Net	FTL	$\gamma=1.6, \alpha=0.6, \beta=0.4$	$1e^{-5}$	0.74 ± 0.23	0.77	0.70	0.82	0.88
RA-U-Net	FTL	$\gamma=1.33, \alpha=0.6, \beta=0.4$	$1e^{-5}$	0.86 ± 0.11	0.89	0.85	0.89	0.92
RA-U-Net	FTL	$\gamma=1.6, \alpha=0.6, \beta=0.4$	$1e^{-4}$	0.80 ± 0.22	0.84	0.82	0.87	0.90
Deep U-Net	FTL	$\gamma=1.33, \alpha=0.6, \beta=0.4$	CLR ($max_{lr} = 1e^{-2}, base_{lr} = 1e^{-4}$)	0.79 ± 0.25	0.83	0.82	0.89	0.92
RA-U-Net	FTL	$\gamma=1.33, \alpha=0.6, \beta=0.4$	CLR ($max_{lr} = 1e^{-3}, base_{lr} = 1e^{-5}$)	0.904 ± 0.101	0.95	0.90	0.92	0.94

Como podemos observar en la tabla anterior, los modelos entrenados con el algoritmo U-Net tienen un rendimiento menor independientemente de que se use la función de pérdidas Tversky o Focal Tversky. El hecho de utilizar la arquitectura RA-U-Net nos ofrece una clara mejora del rendimiento. Cuando empleamos el algoritmo Deep U-Net, podemos observar una mejora con respecto a aquellos resultados procedentes de los modelos U-Net, ya que como sabemos, las redes profundas tienen una mejor capacidad de aprendizaje. Aun así, la arquitectura RA-U-Net

Capítulo 4. Resultados y discusión

mediante la técnica de CLR sigue mostrando el mejor comportamiento en forma de un alto coeficiente DICE, una desviación estándar baja y valores de percentiles altos.

En la Figura 54, se muestran todos los histogramas de distribución de valores DICE de cada una de las configuraciones recogidas en la Tabla 9, en el mismo orden de aparición que en esta. Observamos cómo, en cada ejemplo, los histogramas son progresivamente más estrechos y centrados en valores alto de coeficiente DICE, lo que indica que los modelos tienen un mejor resultado de predicción en el conjunto de datos, tanto en la calidad como en la cantidad de las predicciones. Si observamos los casos extremos en la Figura 54.a y Figura 54.i, vemos como con una arquitectura U-Net tradicional, se obtiene un histograma de valores DICE ancho y disperso, observando también un alto número de imágenes en los que se han obtenido valores DICE a 0 (segmentaciones incorrectas). Estos valores DICE en 0 son provocados, en la mayoría de los casos, por la extensión longitudinal de la segmentación, es decir, que en algunos casos la segmentación manual tiene un mayor o menor número de slices incluidas como médula que la segmentación automática, o viceversa. Por otra parte, en la Figura 54.i, los valores DICE obtenidos con la arquitectura y configuración propuesta en la presente tesis, presentan una distribución estrecha, poco dispersa y con la mayoría de valores centrados en un valor DICE mayor a 0.85, indicando que nuestro modelo propuesto proporciona una segmentación precisa en la mayoría de las imágenes del conjunto de datos.

Capítulo 4. Resultados y discusión

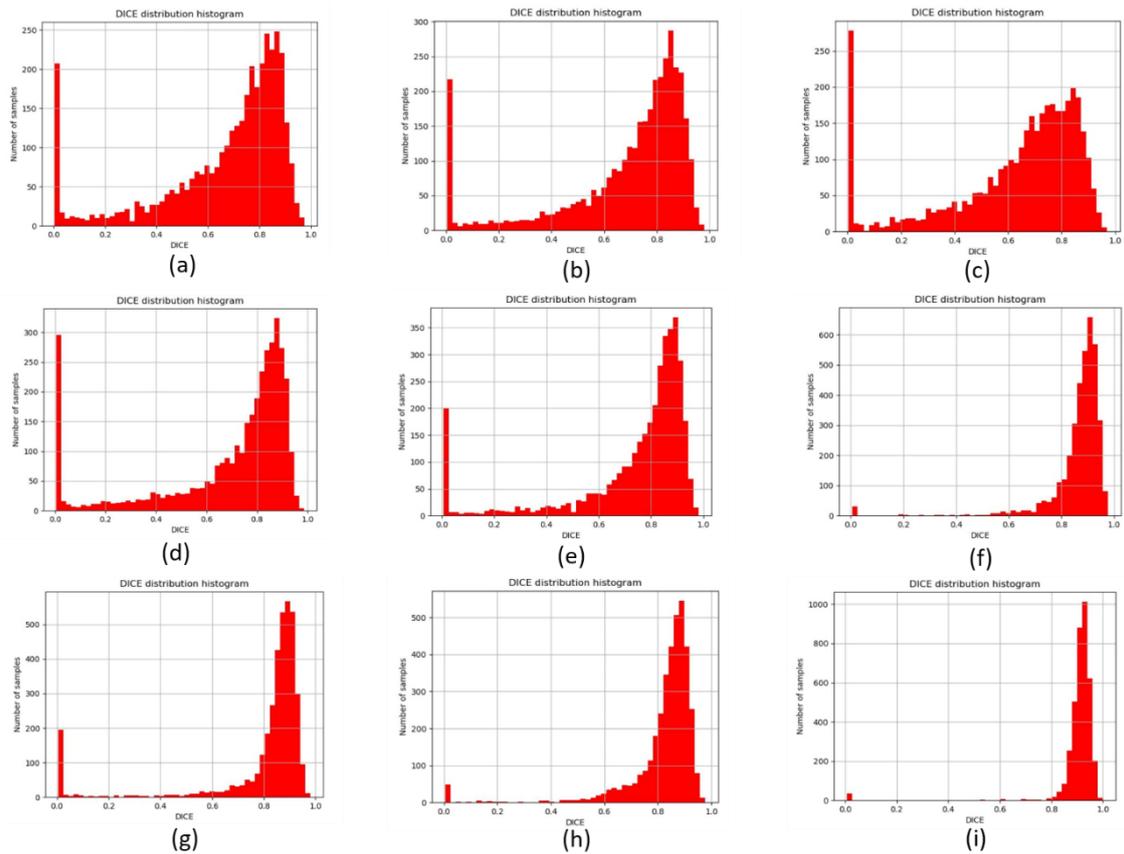


Figura 54. Histogramas de distribución DICE de las configuraciones mostradas en la Tabla 6, desde la imagen (a) hasta la (i), en el mismo orden de aparición. Evolución progresiva de los modelos, distribuciones menos dispersas y más centradas en valores altos en cada ensayo.

Esta evolución de mejora también puede ser observada cualitativamente. A continuación, observamos una muestra visual de la evolución de los resultados. Como ejemplo, observamos la tercera, quinta, octava y última configuración presentes en la Tabla 9 (Figura 55, Figura 56, Figura 57 e Figura 58, respectivamente), con el fin de plasmar ilustrativamente la respuesta de cada una de estas configuraciones al ser aplicadas a cada modelo y este, a nuestro conjunto de datos.

Capítulo 4. Resultados y discusión

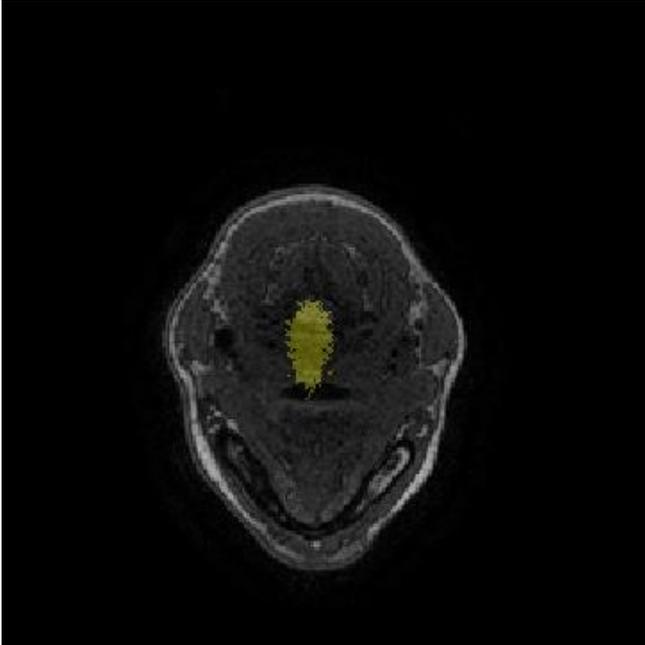
Modelo	U-Net	
Optimizador	<i>Adam(lr=1e-5)</i>	
Función de pérdidas	<i>Focal Tversky Loss (Focal loss (gamma=1.33, beta=0.3, alpha=0.7)</i>	
Métrica	<i>Coefficiente de Correlación Matthews</i>	
DICE medio	0.62	

Figura 55. Muestra visual de prueba con arquitectura U-Net. Tercera configuración de la Tabla 9.

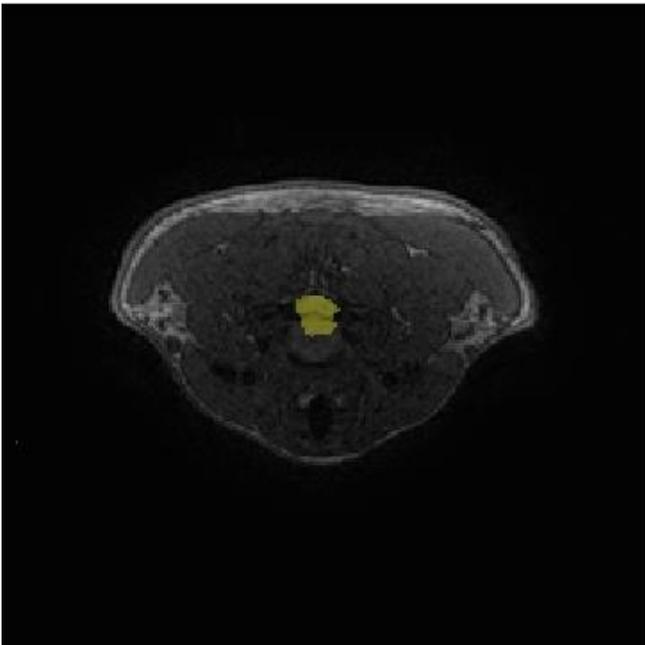
Modelo	U-Net	
Optimizador	<i>Adam(lr=1e-5)</i>	
Función de pérdidas	<i>Focal loss (gamma=1.6, alpha=0.6)</i>	
Métrica	<i>Coefficiente de Correlación Matthews</i>	
DICE medio	0.74	

Figura 56. Muestra visual de prueba con arquitectura U-Net. Quinta configuración de la Tabla 9.

Capítulo 4. Resultados y discusión

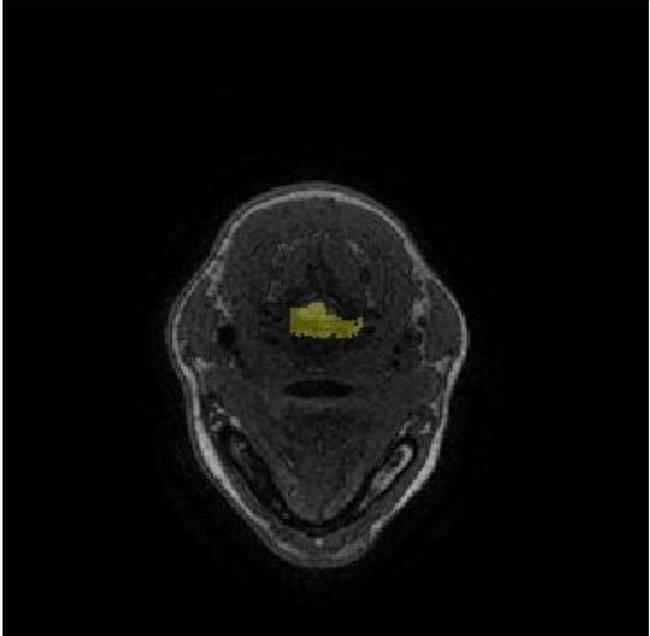
Modelo	Deep U-Net	
Optimizador	<i>CLR</i>	
Función de pérdidas	<i>Focal loss (gamma=1.33, alpha=0.6)</i>	
Métrica	<i>Coefficiente de Correlación Matthews</i>	
DICE medio	0.79	

Figura 57. Muestra visual de prueba con arquitectura U-Net. Octava configuración de la Tabla 9.

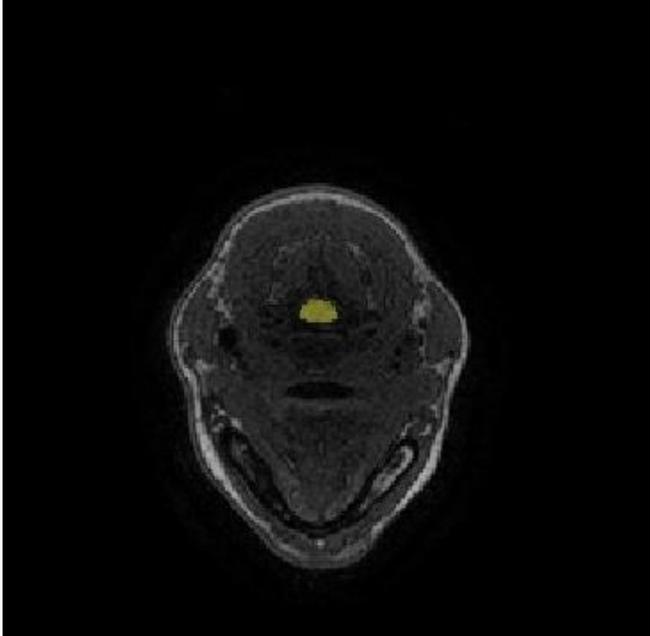
Modelo	RA-UNet	
Optimizador	<i>CLR</i>	
Función de pérdidas	<i>Focal loss (gamma=1.33, alpha=0.6)</i>	
Métrica	<i>Coefficiente de Correlación Matthews</i>	
DICE medio	0.98	

Figura 58. Muestra visual con nuestro método.

Capítulo 4. Resultados y discusión

Podemos observar cómo en la Figura 55, Figura 56 y Figura 57, el modelo entrenado no es lo suficientemente preciso, pues la segmentación resultante, representada sobre la imagen de resonancia en color amarillo, no se ajusta en detalle a la región de médula espinal, de manera que incluye como región de interés no sólo los píxeles pertenecientes a esta estructura objetivo, sino también los pertenecientes a zonas circundantes cercanas, generando una máscara predicha poco cercana al *ground-truth*, viéndose reflejado también en el bajo valor del coeficiente DICE.

En cambio, en la Figura 58 observamos como la máscara predicha por el modelo entrenado en nuestro método se ajusta más fidedignamente a la región de píxeles pertenecientes a médula espinal, excluyendo aquellos cercanos, pero no pertenecientes, viéndose también reflejado este satisfactorio resultado en el alto coeficiente DICE de la muestra.

4.6 Relación entre biomarcadores extraídos, formas clínicas y EDSS en EM

La atrofia de la médula espinal cervical está relacionada a la discapacidad en EM ya que es una parte crucial del SNC y está involucrada en la transmisión de señales entre el cerebro y el resto del cuerpo. La pérdida de tejido nervioso en esta área puede dar lugar a síntomas motores, sensoriales y otros síntomas neurológicos, además de afectar la función de los tractos nerviosos que pasan por esta región, contribuyendo así a la discapacidad en términos de movilidad, coordinación y otras funciones motoras. La EDSS, tal como se explica en la sección 1.1.4, tiene en cuenta estos aspectos al evaluar la discapacidad en personas con EM, demostrándose su relación con la atrofia en la médula espinal y reflejando que es un biomarcador fiable para caracterizar la progresión de la enfermedad (Mina et al., 2021b; Losseff et al., 1996b; Valsasina et al., 2013). En relación a las formas clínicas, y como mencionamos en el apartado 1.2.4.3.1, también se ha demostrado que la atrofia de la médula espinal es más pronunciada en los fenotipos de la EM que se caracterizan por una progresión continuada de la discapacidad (Bernitsas et al., 2015; Mina et al., 2021).

En la sección 3.7, hemos visto cómo nuestro método, además de obtener la segmentación automática de la médula espinal cervical, también la caracteriza proporcionando, como biomarcadores de imagen, dos métricas espaciales, el volumen en mm^3 y la sección media en mm^2 . Empleando durante el estudio de los resultados el valor de sección media en mm^2 , pues debido a la variabilidad de longitudes entre las distintas segmentaciones que componen nuestro estudio, ésta resulta una medida más homogénea y normalizada, además de que, de esta

Capítulo 4. Resultados y discusión

manera, será más intuitiva la comparación de nuestros resultados con los de otros métodos presentes en el estado del arte, los cuales emplean dicha métrica y unidad de medida.

En la Figura 59 podemos observar la distribución de valores de secciones medias de los pacientes de EM que componen nuestro grupo de estudio en función de la forma clínica. Esta gráfica se complementa con la información que observamos en la Tabla 10, de manera que, podemos advertir cómo la sección transversal media de la médula cervical es menor en pacientes con fenotipos progresivos de EM (EMSP y EMPP) que en los recurrentes (EMRR), es decir, que la pérdida de tejido medular es más pronunciada en estos fenotipos progresivos, coincidiendo con resultados obtenidos en estudios previos (Bischof et al., 2022; Papinutto et al., 2018).

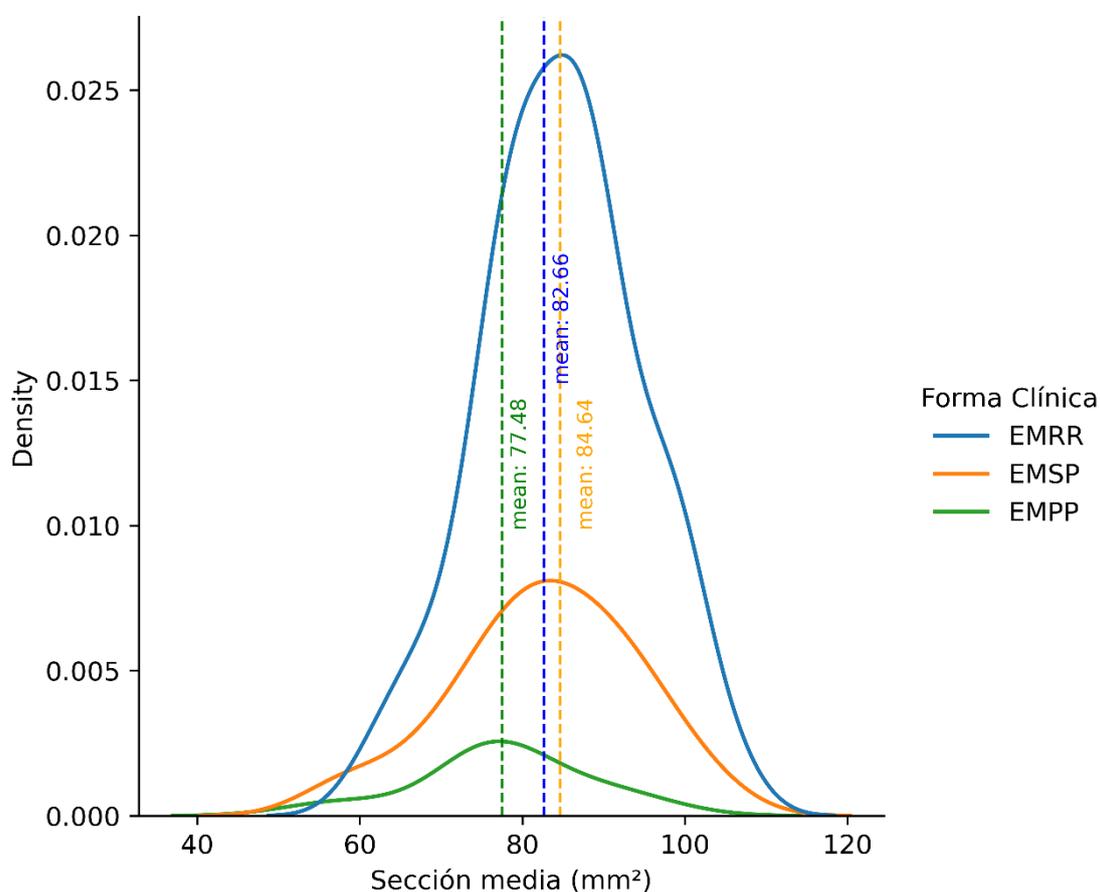


Figura 59. Distribución de valores de sección media entre las distintas formas clínicas que componen nuestra base de datos.

En cuanto a la distribución de valores de la EDSS en nuestro conjunto de datos, observamos en la Figura 60, una discapacidad clínica media mayor en los fenotipos progresivos que en los

Capítulo 4. Resultados y discusión

recurrentes, siendo la $EDSS_{media}=1.98$ en EMRR, $EDSS_{media}=4.41$ en EMSP y $EDSS_{media}=6.18$ en EMPP.

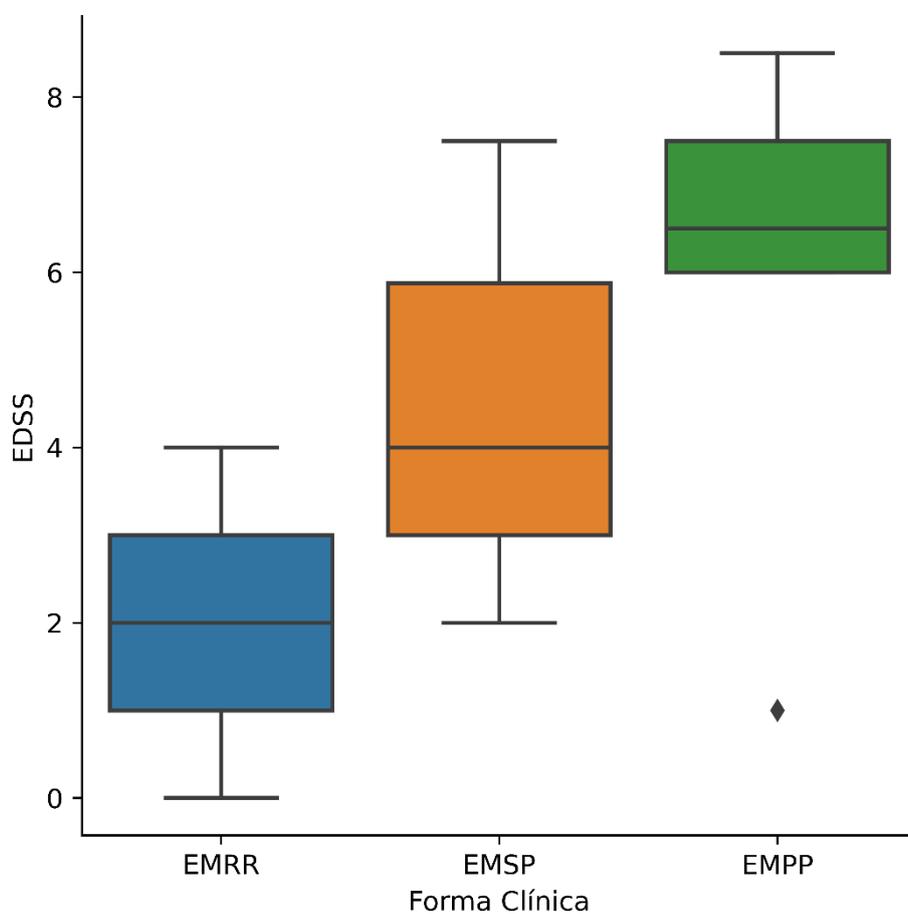


Figura 60. Valores de EDSS por forma clínica.

Con respecto a la relación entre la EDSS y la sección media de cada segmentación, podemos extraer algunas conclusiones observando la representación de la Figura 61. En términos globales del conjunto de datos, visualizamos una leve correlación negativa entre las secciones calculadas y los valores de la EDSS, confirmando esta tendencia mediante el cálculo del coeficiente de correlación de Pearson ($r=-0.3$ y $p=0.001$ ($p<0.05$)).

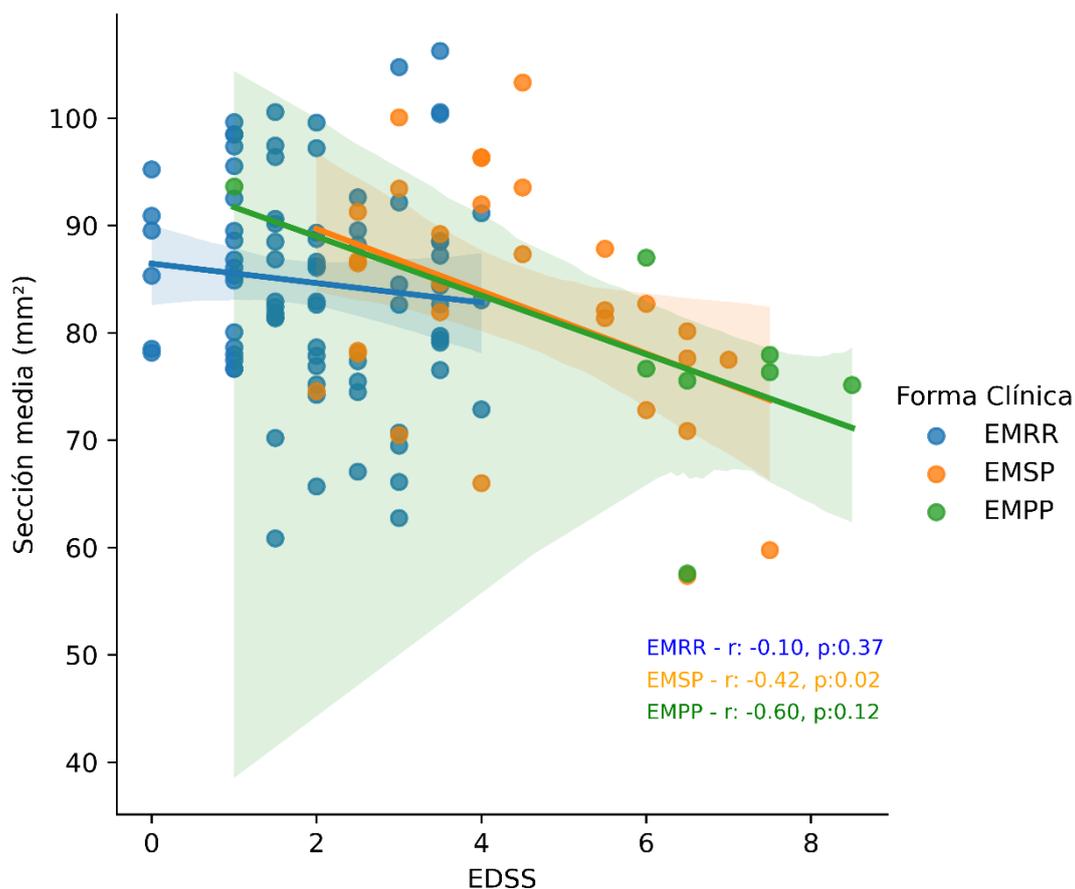


Figura 61. Relación entre sección media y EDSS en las diferentes formas clínicas. Las líneas muestran la tendencia lineal de los datos y se muestran los distintos valores de correlación de Pearson.

Cuando realizamos este mismo estudio entre subgrupos de población distinguidos por las diferentes formas clínicas, obtenemos en EMSP ($r=-0.42$ y $p=0.02$ ($p<0.05$)) también una ligera correlación negativa, aunque más marcada que la obtenida mediante el conjunto de datos en su totalidad ($r=-0.3$ y $p=0.001$ ($p<0.05$)). En EMRR ($r=-0.10$ y $p=0.37$ ($p>0.05$)) la correlación negativa, además de ser muy pequeña, muestra que no hay suficiente evidencia en los datos observados para afirmar que hay una diferencia significativa ($p>0.05$). Esto puede ser debido a la falta de homogeneidad en los valores de secciones medias presentes en la forma clínica de EMRR, pues como contamos en la sección 1.2.4.3.1, las formas remitentes de la enfermedad apenas presentan atrofia en médula espinal, teniendo valores significativamente semejantes a los de los pacientes de control sanos (Bernitsas et al., 2015; Rocca et al., 2011), por lo que en los pacientes con EMRR resulta lógico que sea complejo encontrar una relación entre dimensiones de médula y discapacidad o EDSS. En EMPP, es la forma clínica en la que observamos mayor correlación negativa entre tamaño de la sección media y EDSS ($r=-0.60$ y $p=0.12$ ($p>0.05$)). Pese a que los resultados, en teoría, no son estadísticamente significativos ($p>0.05$), esto puede ser

Capítulo 4. Resultados y discusión

debido al pequeño tamaño de muestra, pues para tener una validez estadística, se recomienda que cada subgrupo tenga al menos una muestra de $n=30$ y, en cambio, sí tener relevancia práctica, pues en los fenotipos progresivos como EMSP y EMPP, la ocurrencia del evento de atrofia medular en los pacientes es más alta que en la forma recurrente y hay una relación significativa con la discapacidad clínica cuantificada con la EDSS (Valsasina et al., 2013), tal como podemos observar en la Figura 62.

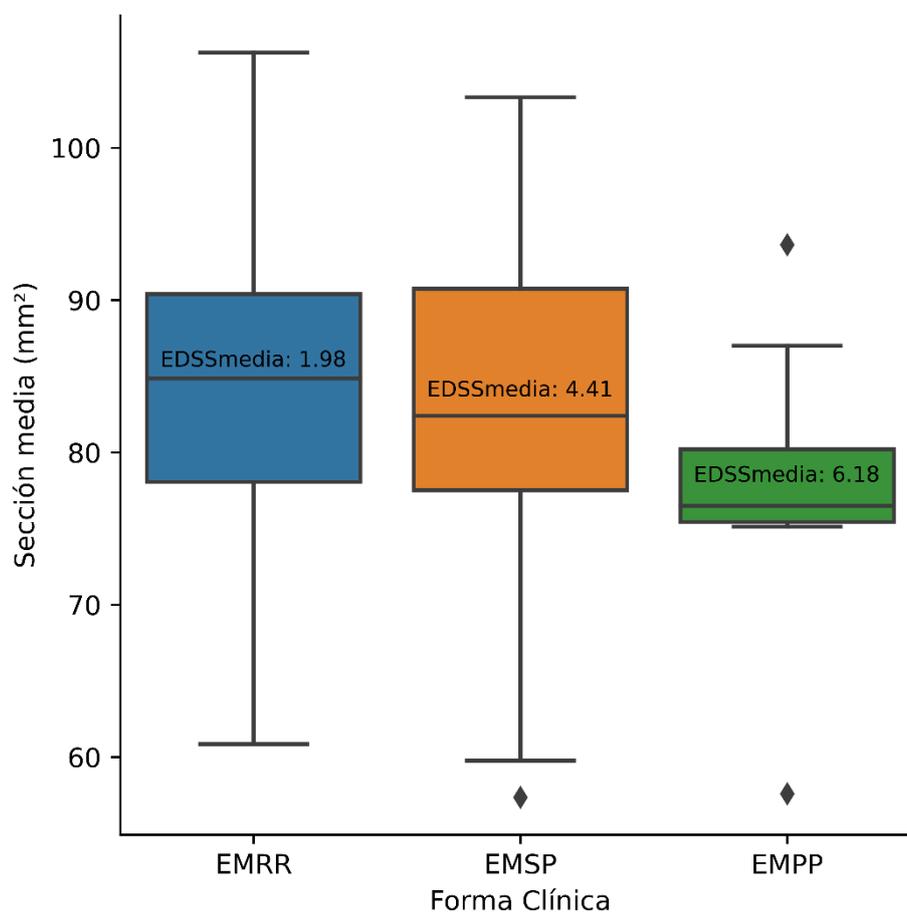


Figura 62. Distribución de valores de sección media por forma clínica y relación con EDSS media.

4.6.1 Discusión de estudio de relación entre biomarcadores de imagen extraídos, formas clínicas y EDSS

Mostramos, en la presente sección, una comparación de nuestros resultados junto a otros presentes en el estado del arte que también estudian la relación del tamaño de la sección media de la médula espinal cervical con la EDSS y las formas clínicas. En la Tabla 10 se puede observar una comparación entre los resultados de nuestro estudio y otros presentes en el estado del arte que abordan la misma problemática (Losseff et al., 1996a; Lukas et al., 2013b) encontrando

Capítulo 4. Resultados y discusión

ciertas similitudes y pequeñas diferencias entre las dimensiones de sección media y aspectos clínicos como la escala de discapacidad EDSS y las formas clínicas.

En nuestro método, obtenemos una sección transversal media en la médula cervical de 84,64 en EMRR y de 82,6 y 77,48 mm² en los fenotipos de EMSP y EMPP, en ese orden. En el método de C.Lukas (Lukas et al., 2013b), como medida de sección media se calcula *la UCCA (Upper Cervical Cord Area)*, esto es, la sección de médula entre C2 y C5, obteniendo 84.05 mm² en EMRR, 79 mm² en EMSP y 77.3 en EMPP (Horsfield et al., 2010). Mientras que los valores en EMRR y EMPP son semejantes a los nuestros (84.64 mm² y 77,48 mm²), difieren en EMSP (82.66 mm²). Esto puede ser debido a que en este método (Lukas et al., 2013b) el cálculo de la sección media se realiza en una menor porción de médula espinal cervical (C2-C5) frente a nuestro método, en el cual, esta porción no es fija y el área media se obtiene normalizando el volumen a la longitud de la segmentación. En todos los métodos reflejados en la Tabla 10, las formas progresivas de la EM tienen una menor sección que la recurrente. En cuanto a las formas progresivas, tanto en nuestro método como en (Lukas et al., 2013b) la sección media se ve más reducida en EMPP que en EMSP, mientras que en (Losseff et al., 1996a) es al contrario, en dicho estudio, los autores tienen como resultado una mayor sección media en la forma progresiva primaria que en la secundaria, hecho potencialmente debido al pequeño tamaño de muestra de cada una de las formas clínicas componentes de su estudio, no siendo lo suficientemente grande como para tener una validez representativa y comprometiendo así la precisión y validez de las conclusiones. Esto se puede deducir también al observar los valores de *Rango (mm²)* en la tabla para este método, pues los intervalos de confianza son más amplios que en el resto y es más probable que no se capture el verdadero valor que representa a la muestra poblacional.

Capítulo 4. Resultados y discusión

Tabla 10. Información de valores de pacientes, secciones y EDSS en nuestro método y otros del estado del arte. Correlación de nuestro método calculada mediante coeficiente de Pearson. UCCA: Upper Cervical Cord Area, C2: segunda vértebra cervical, FC: Forma Clínica.

	Nuestro método			C.Lukas et al(Lukas et al., 2013b)			N.A.Losseff et al(Losseff et al., 1996a)		
	EMRR	EMSP	EMPP	EMRR	EMSP	EMPP	EMRR	EMSP	EMPP
Zona estudiada	Toda la imagen			UCCA			C2		
Número pacientes	83	30	8	311	92	37	15	15	15
Sección media (mm ²)	84.64	82.66	77.47	84	79	77.3	85.6	61.2	73.1
Rango (mm ²)	60.85-106.24	57.34-103.31	57.58-93.63	78.7-89.3	72.4-84.9	69-81.5	74.1-95	44.8-72.2	48.3-88.8
EDSS media	1.98	4.41	6.18	2.5	5.75	4.5	2.5	7	4
Rango EDSS	0-4	2-7.5	1-8.5	2-3.5	4-6.5	4-6	1-6	4.5-8.5	2-7.5
Correlación EDSS-sección media	Spearman rank ($\rho \leq -0.22$, $P < 0.01$) Sin distinción de FC			Spearman rank ($\rho \leq -0.29$, $P < 0.001$) Sin distinción de FC			Spearman Rank ($\rho -0.7$, $P < 0.01$) Sin distinción de FC		

La relación entre las formas clínicas de la EM y la puntuación de EDSS puede ser variable y depende de factores individuales. En general, como hemos visto anteriormente, se espera que los pacientes con formas progresivas de la enfermedad tengan puntuaciones de EDSS más altas debido a la acumulación de discapacidad a lo largo del tiempo. Aquellos con EMRR pueden tener puntuaciones de EDSS más bajas durante las fases de remisión. En nuestro método, el valor medio de EDSS más bajo lo encontramos en el grupo con EMRR (1.98), coincidiendo con lo anteriormente indicado y con los resultados mostrados por los restantes métodos (Losseff et al., 1996a; Lukas et al., 2013a). Obtenemos mayores valores de EDSS (6.18) en la forma primaria progresiva de la enfermedad (EMPP), mientras que en los demás métodos presentes en la Tabla 10, el valor medio de EDSS es mayor en la forma secundaria progresiva. Esto puede ser debido al amplio rango de valores EDSS que tenemos en nuestros grupos de población. Si observamos la fila "Rango EDSS", vemos que, en nuestro conjunto de datos, los rangos son diferenciadamente más amplios que los de los restantes métodos, teniendo una mayor

Capítulo 4. Resultados y discusión

variabilidad en los valores de EDSS y, por tanto, perdiendo poder de generalización a la hora de reflejar la tendencia en términos de EDSS media de cada forma clínica.

En cuanto a las secciones medias de médula espinal cervical reflejadas, algo que coincide entre nuestros resultados y los presentes en otros estudios, es la analogía de que en la forma clínica de EMRR las secciones de médula espinal cervical son mayores que en las formas progresivas (EMPP y EMSP). Basándonos en nuestros resultados, obtenemos una sección transversal media en la médula cervical de 77,48 y 82,6 mm² en los fenotipos de EMPP y EMSP, respectivamente, y de 84,64 en EMRR. Estos valores coinciden con algunas conclusiones extraídas de estudios existentes (Lin & Blumhardt, 2001; C. Liu et al., 1999; Losseff et al., 1996a; Stevenson et al., 1998), como por ejemplo con el hecho de que en los fenotipos progresivos la reducción de la sección media de la médula espinal cervical es mayor que en las formas recurrentes. Sin embargo, dentro de las formas progresivas de la EM, la reducción del área o volumen suele ser mayor en la EMSP que en la EMPP (Lukas et al., 2013b) debido a diferencias en la evolución de la enfermedad y la progresión de la discapacidad en estos dos fenotipos, como por ejemplo, que la EMPP tiene una progresión más lenta y constante de la discapacidad sin la presencia de recaídas y remisiones evidentes mientras que en la EMSP la atrofia de la médula espinal cervical suele ser más pronunciada, al tener una progresión continua de la enfermedad asociada con una mayor pérdida de tejido tanto en la médula espinal como en el cerebro. Esto no es lo que refleja nuestro estudio, pues en nuestro conjunto de datos, los pacientes con EMPP muestran una menor área media de médula espinal cervical. Esto puede ser debido a diversos factores, uno de los principales, el bajo número de muestra con EMPP (8 pacientes) de entre nuestro conjunto de datos (121 pacientes) ya que con un tamaño de muestra insuficientemente grande, los resultados pueden no ser representativos de la población que se está estudiando, pues estas estimaciones son más propensas a variaciones aleatorias y, por lo tanto, son menos confiables, reduciendo así la potencia estadística y, con ella, la probabilidad de que el estudio identifique relaciones reales. En cambio, como podemos observar en la síntesis de resultados entre diferentes estudios que recoge la Tabla 10, existe otra investigación que, a diferencia de lo mencionado anteriormente, obtiene una relación entre sección media y formas clínicas semejante a la nuestra, siendo la sección media de la EMPP menor que la de la EMSP (Lukas et al., 2013b). Este estudio cuenta con un conjunto de datos mayor que el que hemos mencionado anteriormente como ejemplo de disimilitud con nuestros resultados, por lo que, al menos en términos estadísticos, podríamos considerar que tenga mayor valor. Encontramos algunas pequeñas divergencias entre los distintos estudios que hemos encontrado en el estado del arte, lo que resulta natural en el análisis de este tipo de investigaciones en las que factores como el

Capítulo 4. Resultados y discusión

tamaño de la muestra, el método de recopilación de datos, la duración del estudio, la variabilidad biológica o el tipo de imágenes empleadas pueden variar y afectar a los resultados.

Capítulo 5. Conclusiones y líneas futuras

Capítulo 5. Conclusiones y líneas futuras

En la presente tesis se ha implementado un método automático para la segmentación de médula espinal a partir de imágenes de RM adquiridas de datos observacionales de pacientes de EM y la subsecuente extracción de biomarcadores de imagen. Las principales contribuciones del método propuesto son: la falta de necesidad de interacción humana, el uso de datos del mundo real para acercar la implementación de soluciones a la práctica clínica habitual, la minimización del problema de desbalanceo de datos sin la necesidad de técnicas de preprocesamiento y la optimización en términos temporal debido al uso del método de ajuste de tasa de aprendizaje cíclica.

Debido a la gran variedad de equipos y la falta de un consenso común de adquisición estándar de imágenes de RM en la rutina clínica (*real world data*), las imágenes de RM difieren entre ellas, cubriendo diferentes regiones de la médula espinal cervical con diferentes intensidades de tejidos, diferentes ángulos de visión, diferente tamaño de vóxeles y diferentes intensidades de señal. En relación a esto, se ha conseguido la segmentación de médula espinal cervical mediante el diseño de CNN y su parametrización, evitando técnicas de preprocesado de imágenes de RM que requieran interacción humana, de manera que se pueda obtener la segmentación automática de datos normales adquiridos en cualquier equipo de adquisición de RM de un hospital común. Para entrenar y validar el algoritmo se extrajo y utilizó un conjunto de datos compuesto por 121 imágenes axiales 3D potenciadas T1 de pacientes de EM, secuencia común en los protocolos de adquisición de médula cervical en EM. Estos casos fueron segmentados manualmente con la ayuda de radiólogos expertos dan lugar a la generación del *ground-truth*. Nuestro método puede abordar el problema de desbalanceo de clases sin la necesidad de técnicas de preprocesado, de manera que, pudiendo utilizar directamente los datos adquiridos del mundo real se puede conseguir una mejor integración clínica. Hemos cumplido satisfactoriamente el objeto de segmentación automática de médula espinal cervical mediante una CNN optimizada con un mecanismo de atención residual híbrida, con la función de pérdidas Focal Tversky y un método de tasa de aprendizaje cíclica para determinar, cambiar y ajustar la tasa de aprendizaje durante el entrenamiento. Este favorable modelo ha sido validado en un conjunto de datos compuesto por imágenes adquiridas 5 años más tarde a las empleadas en la fase de diseño, demostrando así la correcta capacidad de generalización en nuevos datos. Por último, se presenta una comparación de nuestra arquitectura y otras del estado del arte en combinación con varias funciones de pérdidas y métodos de aprendizaje. El modelo de segmentación automática propuesto tendrá implicaciones directas para la aceleración del fenotipado, seguimiento y extracción de biomarcadores de imagen en la EM, pues conseguimos una gran optimización temporal aliviando esfuerzo en la tarea de segmentación y, además,

Capítulo 5. Conclusiones y líneas futuras

extrayendo biomarcadores de imagen como el volumen en mm^3 y el área cervical media en mm^2 de las segmentaciones obtenidas, incrementando así los medios para la monitorización de la evolución de la enfermedad con el fin de definir los factores que hay detrás del aumento de la discapacidad irreversible, síntoma principal de la EM.

Mediante este estudio también se ha realizado un análisis de nuestro conjunto de datos en términos de relación entre las secciones medias calculadas de las segmentaciones de médula espinal cervical y los factores clínicos de formas de la enfermedad y discapacidad definida en términos de la EDSS. Encontramos una mayor presencia de atrofia en los fenotipos progresivos de la enfermedad con respecto a los recurrentes, así como una mayor discapacidad en términos de EDSS coincidiendo con estudios que han abordado la misma problemática.

Las limitaciones detectadas durante el estudio también serán investigadas en el futuro. Algunas de ellas principalmente relacionadas con el tamaño relativamente pequeño de muestra de pacientes y contar con un grupo de control sin EM, pues es común la dificultad en aplicaciones de IA especializadas en medicina el conseguir un amplio conjunto de datos debido a problemas como la dificultad de realizar el etiquetado de estos. Llevar a cabo una comparación más exhaustiva que la mostrada en la sección 4.5, del rendimiento entre las herramientas ya existentes y la nuestra, también hubiese sido adecuado para clarificar mejor la calidad de nuestro modelo, pero en este caso, esta práctica era difícil de llevar a cabo ya que no todos manipulábamos ni el mismo tipo de imágenes RM ni las mismas poblaciones de estudio, lo que añaden variabilidad a los resultados. Otra de las principales limitaciones es la falta de validación externa, ya que, pese a la satisfactoria validación interna mostrada en la sección 4.3, nuestro método debería ser validado externamente en otros centros, empleando adquisiciones de RM de otros escáneres de diferentes proveedores, campos magnéticos y parámetros de adquisición. De esta manera seremos capaces de evaluar la interoperabilidad de nuestro método.

Con respecto a la arquitectura empleada en este proyecto, señalar que, aunque U-Net y sus variaciones siguen siendo populares y eficientes en el campo de la segmentación de imagen, son varios los modelos y técnicas que se han desarrollado desde entonces, pues la comunidad de investigación en IA está continuamente en busca avances significativos, generando una velocidad de desarrollo notable que provoca una continua evolución. Es por ello que, en la actualidad, podemos encontrar modelos relevantes en la segmentación de imagen como: *DeepLab* (L. C. Chen et al., 2016), una familia de modelos de segmentación semántica desarrollados por Google que utilizan una red neuronal convolucional profunda e incorporan *atrous convolutions* (convoluciones dilatadas) para expandir el campo receptivo sin aumentar el

Capítulo 5. Conclusiones y líneas futuras

número de parámetros; *High-Resolution Network (HRNet)* (J. Wang et al., 2019), la cual preserva los detalles finos en la segmentación centrándose en mantener una alta resolución espacial durante toda la red a través de conexiones entre diferentes escalas de características; *Pyramid Scene Parsing Network (PSPNet)* (Zhao et al., 2016), que captura información contextual a diferentes escalas mediante la introducción del módulo de *pooling piramidal (PSP)*, permitiendo así una comprensión más global de la escena o *Hybrid Dilated Networks (HD-Net)* (Z. Wang et al., 2021), la cual mejora la segmentación, capturando detalles finos y manteniendo un campo receptivo amplio, combinando módulos dilatados con una red convencional. La arquitectura de *Transformer* (Vaswani et al., 2017) también ha tenido una irrupción relevante en el campo de la IA, pues, pese a ser inicialmente popular en tareas de procesamiento de lenguaje natural, ha demostrado ser eficaz también en tareas de visión, incluida la segmentación de imágenes. En este ámbito, podemos destacar algunos modelos de arquitectura *Transformer*: *Vision Transformer (ViT)* (Dosovitskiy et al., 2020), el cual captura las relaciones de largo alcance introduciendo la idea de tratar las imágenes como secuencias de parches y utilizando bloques de atención de transformer; *Swin Transformer* (Z. Liu et al., 2021), que permite manejar información a diferentes escalas mediante una estructura jerárquica que organiza los bloques de atención de manera escalonada; *Class-Attention in Image Transformer (CaiT)* (Touvron et al., 2021), el cual mejora la capacidad de discriminación introduciendo bloques de atención de transformer con atención de clase en lugar de atención estándar con el fin de aumentar la utilidad en tareas de segmentación que requieren identificación detallada entre clases.

Finalmente, a partir de las segmentaciones realizadas con nuestro método, se ha iniciado el desarrollo de un modelo agregado de detección de lesiones intramedulares a partir de secuencias de imagen STIR que se centrará en la detección de lesiones en la región cervical. El método de segmentación automática presentado en esta tesis se utilizará como primer paso para obtener la localización de la médula espinal cervical en el espacio STIR correspondiente tras aplicar técnicas de registro de imágenes entre ambas secuencias. El estado de esta propuesta, aparece más detalladamente descrito y desarrollado en Anexo I: Detección automática de lesiones en médula cervical.

Referencias

- Acharjya, P. P., Das, R., Ghoshal, D., & Pratim Acharjya, P. (2012). *Study and Comparison of Different Edge Detectors for Image Segmentation Study and Comparison of Different Edge Detectors for Image Segmentation Study and Comparison of Different Edge Detectors for Image Segmentation*.
- Activation Functions: With Real-life analogy and Python Code - Shiksha Online*. (n.d.). Retrieved January 12, 2024, from <https://www.shiksha.com/online-courses/articles/activation-functions-with-real-life-analogy-and-python-code/>
- Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*. *Neural Networks and Deep Learning*. <https://doi.org/10.1007/978-3-319-94463-0>
- Andelova, M., Uher, T., Krasensky, J., Sobisek, L., Kusova, E., Srpova, B., Vodehnalova, K., Friedova, L., Motyl, J., Preiningerova, J. L., Havrdova, E. K., Horakova, D., & Vaneckova, M. (2019). Additive effect of spinal cord volume, diffuse and focal cord pathology on disability in multiple sclerosis. *Frontiers in Neurology*, *10*(JUL). <https://doi.org/10.3389/FNEUR.2019.00820>
- Aquib Ansari, M., Kurchaniya, D., & Dixit, M. (2017). A Comprehensive Analysis of Image Edge Detection Techniques. *International Journal of Multimedia and Ubiquitous Engineering*, *12*(11), 1–12. <https://doi.org/10.14257/ijmue.2017.12.11.01>
- Arnold, D. L., & Matthews, P. M. (2002). MRI in the diagnosis and management of multiple sclerosis. *Neurology*, *58*(8 SUPPL. 5). https://doi.org/10.1212/WNL.58.8_SUPPL_4.S23
- Arrambide, G., Rovira, A., Sastre-Garriga, J., Tur, C., Castelló, J., Ríó, J., Vidal-Jordana, A., Galán, I., Rodríguez-Acevedo, B., Midaglia, L., Nos, C., Mulero, P., Arévalo, M. J., Comabella, M., Huerga, E., Auger, C., Montalban, X., & Tintore, M. (2017). Spinal cord lesions: A modest contributor to diagnosis in clinically isolated syndromes but a relevant prognostic factor. <https://doi.org/10.1177/1352458517697830>, *24*(3), 301–312. <https://doi.org/10.1177/1352458517697830>
- Asgari Taghanaki, S., Abhishek, K., Cohen, J. P., Cohen-Adad, J., & Hamarneh, G. (2019). Deep Semantic Segmentation of Natural and Medical Images: A Review. *Artificial Intelligence Review*, *54*(1), 137–178. <https://doi.org/10.1007/s10462-020-09854-1>
- Azodi, S., Nair, G., Enose-Akahata, Y., Charlip, E., Vellucci, A., Cortese, I., Dwyer, J., Billioux, B. J., Thomas, C., Ohayon, J., Reich, D. S., & Jacobson, S. (2017). Imaging spinal cord atrophy in progressive myelopathies: HTLV-I-associated neurological disease (HAM/TSP) and multiple sclerosis (MS). *Annals of Neurology*, *82*(5), 719–728. <https://doi.org/10.1002/ANA.25072>
- Baldi, P. (2012). *Autoencoders, Unsupervised Learning, and Deep Architectures* (Vol. 27, pp. 37–49). JMLR Workshop and Conference Proceedings.

Referencias

- <https://proceedings.mlr.press/v27/baldi12a.html>
- Benito-León, J., & Laurence, M. (2017). The role of fungi in the etiology of multiple sclerosis. *Frontiers in Neurology*, 8(OCT), 298966. <https://doi.org/10.3389/FNEUR.2017.00535/BIBTEX>
- Bernitsas, E., Bao, F., Seraji-Bozorgzad, N., Chorostecki, J., Santiago, C., Tselis, A., Caon, C., Zak, I., Millis, S., & Khan, O. (2015). Spinal cord atrophy in multiple sclerosis and relationship with disability across clinical phenotypes. *Multiple Sclerosis and Related Disorders*, 4(1), 47–51. <https://doi.org/10.1016/J.MSARD.2014.11.002>
- Bernstein, M. A., King, K. F., & Zhou, X. J. (2004). Handbook of MRI Pulse Sequences. *Handbook of MRI Pulse Sequences*, 1–1017. <https://doi.org/10.1016/B978-0-12-092861-3.X5000-6>
- Biberacher, V., Boucard, C. C., Schmidt, P., Engl, C., Buck, D., Berthele, A., Hoshi, M. M., Zimmer, C., Hemmer, B., & Mühlau, M. (2015). Atrophy and structural variability of the upper cervical cord in early multiple sclerosis. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 21(7), 875–884. <https://doi.org/10.1177/1352458514546514>
- Bischof, A., Papinutto, N., Keshavan, A., Rajesh, A., Kirkish, G., Zhang, X., Mallott, J. M., Asteggiano, C., Sacco, S., Gundel, T. J., Zhao, C., Stern, W. A., Caverzasi, E., Zhou, Y., Gomez, R., Ragan, N. R., Santaniello, A., Zhu, A. H., Juwono, J., ... Henry, R. G. (2022). Spinal Cord Atrophy Predicts Progressive Disease in Relapsing Multiple Sclerosis. *Annals of Neurology*, 91(2), 268–281. <https://doi.org/10.1002/ANA.26281>
- Bishop, C. M. (2006). Pattern Recognition and Machine Learning. *Information Science and Statistics*, 738. <https://www.springer.com/gp/book/9780387310732>
- Bjartmar, C., Kidd, G., Mörk, S., Rudick, R., & Trapp, B. D. (2000). Neurological disability correlates with spinal cord axonal loss and reduced N-acetyl aspartate in chronic multiple sclerosis patients. *Annals of Neurology*, 48(6), 893–901. [https://doi.org/10.1002/1531-8249\(200012\)48:6<893::AID-ANA10>3.0.CO;2-B](https://doi.org/10.1002/1531-8249(200012)48:6<893::AID-ANA10>3.0.CO;2-B)
- Boiko, A., Vorobeychik, G., Paty, D., Devonshire, V., Sadovnick, D., Hashimoto, S., Hooge, J., Kastrukoff, L., Oger, J., & Traboulsee, T. (2002). Early onset multiple sclerosis: a longitudinal study. *Neurology*, 59(7), 1006–1010. <https://doi.org/10.1212/WNL.59.7.1006>
- Bonati, U., Fisniku, L. K., Altmann, D. R., Yiannakas, M. C., Furby, J., Thompson, A. J., Miller, D. H., & Chard, D. T. (2011). Cervical cord and brain grey matter atrophy independently associate with long-term MS disability. *Journal of Neurology, Neurosurgery, and Psychiatry*, 82(4), 471–472. <https://doi.org/10.1136/JNNP.2010.205021>
- Bot, J. C., & Barkhof, F. (2009). Spinal-cord MRI in multiple sclerosis: conventional and nonconventional MR techniques. *Neuroimaging Clinics of North America*, 19(1), 81–99. <https://doi.org/10.1016/J.NIC.2008.09.005>
- Bot, J. C. J., Barkhof, F., Polman, C. H., Lycklama À Nijeholt, G. J., De Groot, V., Bergers,

Referencias

- E., Ader, H. J., & Castelijns, J. A. (2004). Spinal cord abnormalities in recently diagnosed MS patients: added value of spinal MRI examination. *Neurology*, *62*(2), 226–233. <https://doi.org/10.1212/WNL.62.2.226>
- Brex, P. A., Ciccarelli, O., O’Riordan, J. I., Sailer, M., Thompson, A. J., & Miller, D. H. (2002). A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. *The New England Journal of Medicine*, *346*(3), 158–164. <https://doi.org/10.1056/NEJMOA011341>
- Brodsky, M., Nazarian, S., Orengo-Nania, S., Hutton, G. J., Buckley, E. G., Massey, E. W., Bhatti, M. T., Greer, M., Goodwin, J., Wall, M., Savino, P. J., Leist, T., Miller, N. R., Irani, D., Trobe, J. D., Cornblath, W., Kaufman, D. I., Eggenberger, E., Kupersmith, M. J., ... Smith, C. H. (2008). Multiple sclerosis risk after optic neuritis: final optic neuritis treatment trial follow-up. *Archives of Neurology*, *65*(6), 727–732. <https://doi.org/10.1001/ARCHNEUR.65.6.727>
- Brownlee, W. J., Altmann, D. R., Alves Da Mota, P., Swanton, J. K., Miszkiel, K. A., Wheeler-Kingshott, C. A. M. G., Ciccarelli, O., & Miller, D. H. (2017). Association of asymptomatic spinal cord lesions and atrophy with disability 5 years after a clinically isolated syndrome. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, *23*(5), 665–674. <https://doi.org/10.1177/1352458516663034>
- Buda, M., Maki, A., & Mazurowski, M. A. (2017). *A systematic study of the class imbalance problem in convolutional neural networks*. <https://doi.org/10.1016/j.neunet.2018.07.011>
- Bueno, A., Bosch, I., Rodríguez, A., Jiménez, A., Carreres, J., Fernández, M., Martí-Bonmati, L., & Alberich-Bayarri, A. (2022). Automated Cervical Spinal Cord Segmentation in Real-World MRI of Multiple Sclerosis Patients by Optimized Hybrid Residual Attention-Aware Convolutional Neural Networks. *Journal of Digital Imaging*, *35*(5), 1131–1142. <https://doi.org/10.1007/S10278-022-00637-4>
- Carbonell-Caballero, J., Manjón, J. V., Martí-Bonmatí, L., Olalla, J. R., Casanova, B., De La Iglesia-Vayá, M., Coret, F., & Robles, M. (2006). Accurate quantification methods to evaluate cervical cord atrophy in multiple sclerosis patients. *Magma (New York, N.Y.)*, *19*(5), 237–246. <https://doi.org/10.1007/S10334-006-0052-0>
- Casserly, C., Seyman, E. E., Alcaide-Leon, P., Guenette, M., Lyons, C., Sankar, S., Svendrovski, A., Baral, S., & Oh, J. (2018). Spinal Cord Atrophy in Multiple Sclerosis: A Systematic Review and Meta-Analysis. *Journal of Neuroimaging : Official Journal of the American Society of Neuroimaging*, *28*(6), 556–586. <https://doi.org/10.1111/JON.12553>
- Cawley, N., Tur, C., Prados, F., Plantone, D., Kearney, H., Abdel-Aziz, K., Ourselin, S., Wheeler-Kingshott, C. A. M. G., Miller, D. H., Thompson, A. J., & Ciccarelli, O. (2018). Spinal cord atrophy as a primary outcome measure in phase II trials of progressive multiple sclerosis. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, *24*(7), 932–941. <https://doi.org/10.1177/1352458517709954>
- Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2016). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution,

Referencias

- and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834–848. <https://doi.org/10.1109/TPAMI.2017.2699184>
- Chen, M., Carass, A., Oh, J., Nair, G., Pham, D. L., Reich, D. S., & Prince, J. L. (2013). Automatic magnetic resonance spinal cord segmentation with topology constraints for variable fields of view. *NeuroImage*, 83, 1051–1062. <https://doi.org/10.1016/j.neuroimage.2013.07.060>
- Chicco, D. (2017). Ten quick tips for machine learning in computational biology. *BioData Mining*, 10(1), 1–17. <https://doi.org/10.1186/S13040-017-0155-3/FIGURES/3>
- Chicco, D., & Jurman, G. (n.d.). *The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation*. <https://doi.org/10.1186/s12864-019-6413-7>
- Cielen, D., Meysman, A. D. B., & Ali, M. (2016). The Data Science Process. *Introducing Data Science*, 22–56. <https://www.manning.com/books/introducing-data-science>
- Cohen, A. B., Neema, M., Arora, A., Dell’Oglio, E., Benedict, R. H. B., Tauhid, S., Goldberg-Zimring, D., Chavarro-Nieto, C., Ceccarelli, A., Klein, J. P., Stankiewicz, J. M., Houtchens, M. K., Buckle, G. J., Alsop, D. C., Guttmann, C. R. G., & Bakshi, R. (2012a). The relationships among MRI-defined spinal cord involvement, brain involvement, and disability in multiple sclerosis. *Journal of Neuroimaging*, 22(2), 122. <https://doi.org/10.1111/J.1552-6569.2011.00589.X>
- Cohen, A. B., Neema, M., Arora, A., Dell’Oglio, E., Benedict, R. H. B., Tauhid, S., Goldberg-Zimring, D., Chavarro-Nieto, C., Ceccarelli, A., Klein, J. P., Stankiewicz, J. M., Houtchens, M. K., Buckle, G. J., Alsop, D. C., Guttmann, C. R. G., & Bakshi, R. (2012b). The Relationships among MRI-Defined Spinal Cord Involvement, Brain Involvement, and Disability in Multiple Sclerosis. *Journal of Neuroimaging*, 22(2), 122–128. <https://doi.org/10.1111/j.1552-6569.2011.00589.x>
- Compston, A., & Coles, A. (2008). Multiple sclerosis. *Lancet (London, England)*, 372(9648), 1502–1517. [https://doi.org/10.1016/S0140-6736\(08\)61620-7](https://doi.org/10.1016/S0140-6736(08)61620-7)
- Cope, B., Kalantzis, M., & Searsmith, D. (2021). Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. *Educational Philosophy and Theory*, 53(12), 1229–1245. <https://doi.org/10.1080/00131857.2020.1728732>
- Daams, M., Steenwijk, M. D., Wattjes, M. P., Geurts, J. J. G., Uitdehaag, B. M. J., Tewarie, P. K., Balk, L. J., Pouwels, P. J. W., Killestein, J., & Barkhof, F. (2015). Unraveling the neuroimaging predictors for motor dysfunction in long-standing multiple sclerosis. *Neurology*, 85(3), 248–255. <https://doi.org/10.1212/WNL.0000000000001756>
- Daams, M., Weiler, F., Steenwijk, M. D., Hahn, H. K., Geurts, J. J. G., Vrenken, H., Van Schijndel, R. A., Balk, L. J., Tewarie, P. K., Tillema, J. M., Killestein, J., Uitdehaag, B. M. J., & Barkhof, F. (2014). Mean upper cervical cord area (MUCCA) measurement in long-standing multiple sclerosis: relation to brain findings and clinical disability. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 20(14), 1860–1865.

Referencias

- <https://doi.org/10.1177/1352458514533399>
- Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94. <https://doi.org/10.7861/FUTUREHOSP.6-2-94>
- De Leener, B., Cohen-Adad, J., & Kadoury, S. (2015). Automatic Segmentation of the Spinal Cord and Spinal Canal Coupled With Vertebral Labeling. *IEEE Transactions on Medical Imaging*, 34(8), 1705–1718. <https://doi.org/10.1109/TMI.2015.2437192>
- De Leener, B., Lévy, S., Dupont, S. M., Fonov, V. S., Stikov, N., Louis Collins, D., Callot, V., & Cohen-Adad, J. (2017). SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. *NeuroImage*, 145, 24–43. <https://doi.org/10.1016/j.neuroimage.2016.10.009>
- De Leener, B., Taso, M., Cohen-Adad, J., & Callot, V. (2016). Segmentation of the human spinal cord. In *Magnetic Resonance Materials in Physics, Biology and Medicine* (Vol. 29, Issue 2, pp. 125–153). Springer Verlag. <https://doi.org/10.1007/s10334-015-0507-2>
- Dey, D., & Polley, D. (2016). Edge Detection by Using Canny and Prewitt. *International Journal of Scientific & Engineering Research*, 7(4). <http://www.ijser.org>
- Dhruv, P., & Naskar, S. (2020). Image classification using convolutional neural network (CNN) and Recurrent Neural Network (RNN): A Review. *Advances in Intelligent Systems and Computing*, 1101, 367–381. https://doi.org/10.1007/978-981-15-1884-3_34/COVER
- dicom2nifti — dicom2nifti documentation*. (n.d.). Retrieved February 1, 2024, from <https://icometrix.github.io/dicom2nifti/readme.html>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR 2021 - 9th International Conference on Learning Representations*. <https://arxiv.org/abs/2010.11929v2>
- Dupuy, S. L., Khalid, F., Healy, B. C., Bakshi, S., Neema, M., Tauhid, S., & Bakshi, R. (2016). The effect of intramuscular interferon beta-1a on spinal cord volume in relapsing-remitting multiple sclerosis. *BMC Medical Imaging*, 16(1). <https://doi.org/10.1186/S12880-016-0158-4>
- Durand-Dubief, F., Belaroussi, B., Armspach, J. P., Dufour, M., Roggerone, S., Vukusic, S., Hannoun, S., Sappey-Marinier, D., Confavreux, C., & Cotton, F. (2012). Reliability of longitudinal brain volume loss measurements between 2 sites in patients with multiple sclerosis: comparison of 7 quantification techniques. *AJNR. American Journal of Neuroradiology*, 33(10), 1918–1924. <https://doi.org/10.3174/AJNR.A3107>
- Ecabert, O., Peters, J., Schramm, H., Lorenz, C., Von Berg, J., Walker, M. J., Vembar, M., Olszewski, M. E., Subramanyan, K., Lavi, G., & Weese, J. (2008). Automatic model-

Referencias

- based segmentation of the heart in CT images. *IEEE Transactions on Medical Imaging*, 27(9), 1189–1202. <https://doi.org/10.1109/TMI.2008.918330>
- El-Baz, A., Farag, A. A., Yuksel, S. E., El-Ghar, M. E. A., Eldiasty, T. A., & Ghoneim, M. A. (2007). Application Of Deformable Models For The Detection Of Acute Renal Rejection. *Deformable Models*, 293–333. https://doi.org/10.1007/978-0-387-68413-0_10
- Elnakib, A., Gimel'farb, G., Suri, J. S., & El-Baz, A. (2011). Medical Image Segmentation: A Brief Survey. *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*, 1–39. https://doi.org/10.1007/978-1-4419-8204-9_1
- Eshaghi, A., Prados, F., Brownlee, W. J., Altmann, D. R., Tur, C., Cardoso, M. J., De Angelis, F., van de Pavert, S. H., Cawley, N., De Stefano, N., Stromillo, M. L., Battaglini, M., Ruggieri, S., Gasperini, C., Filippi, M., Rocca, M. A., Rovira, A., Sastre-Garriga, J., Vrenken, H., ... Ciccarelli, O. (2018). Deep gray matter volume loss drives disability worsening in multiple sclerosis. *Annals of Neurology*, 83(2), 210–222. <https://doi.org/10.1002/ANA.25145>
- Filippi, M., Bar-Or, A., Piehl, F., Preziosa, P., Solari, A., Vukusic, S., & Rocca, M. A. (2018). Multiple sclerosis. *Nature Reviews Disease Primers* 2018 4:1, 4(1), 1–27. <https://doi.org/10.1038/s41572-018-0041-4>
- Filippi, M., Preziosa, P., Banwell, B. L., Barkhof, F., Ciccarelli, O., De Stefano, N., Geurts, J. J. G., Paul, F., Reich, D. S., Toosy, A. T., Traboulsee, A., Wattjes, M. P., Yousry, T. A., Gass, A., Lubetzki, C., Weinshenker, B. G., & Rocca, M. A. (2019). Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines. *Brain*, 142(7), 1858–1875. <https://doi.org/10.1093/BRAIN/AWZ144>
- Filippi, M., & Rocca, M. A. (2007). Conventional MRI in multiple sclerosis. *Journal of Neuroimaging : Official Journal of the American Society of Neuroimaging*, 17 Suppl 1(SUPPL. 1). <https://doi.org/10.1111/J.1552-6569.2007.00129.X>
- Fisniku, L. K., Brex, P. A., Altmann, D. R., Miszkiel, K. A., Benton, C. E., Lanyon, R., Thompson, A. J., & Miller, D. H. (2008). Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain : A Journal of Neurology*, 131(Pt 3), 808–817. <https://doi.org/10.1093/BRAIN/AWM329>
- Francois-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018). An Introduction to Deep Reinforcement Learning. *Foundations and Trends in Machine Learning*, 11(3–4), 219–354. <https://doi.org/10.1561/22000000071>
- Furby, J., Hayton, T., Anderson, V., Altmann, D., Brenner, R., Chataway, J., Hughes, R. A. C., Smith, K. J., Miller, D. H., & Kapoor, R. (2008). Magnetic resonance imaging measures of brain and spinal cord atrophy correlate with clinical impairment in secondary progressive multiple sclerosis. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 14(8), 1068–1075. <https://doi.org/10.1177/1352458508093617>
- García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L., & Collins, D. L. (2013). Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. In *Medical Image Analysis*

Referencias

- (Vol. 17, Issue 1, pp. 1–18). <https://doi.org/10.1016/j.media.2012.09.004>
- Giardino, A., Gupta, S., Olson, E., Sepulveda, K., Lenchik, L., Ivanidze, J., Rakow-Penner, R., Patel, M. J., Subramaniam, R. M., & Ganesan, D. (2017). Role of Imaging in the Era of Precision Medicine. *Academic Radiology*, *24*(5), 639–649. <https://doi.org/10.1016/J.ACRA.2016.11.021>
- Gonem, S., Janssens, W., Das, N., & Topalovic, M. (2020). Applications of artificial intelligence and machine learning in respiratory medicine. *Thorax*, *75*(8), 695–701. <https://doi.org/10.1136/THORAXJNL-2020-214556>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks. *Communications of the ACM*, *63*(11), 139–144. <https://doi.org/10.1145/3422622>
- Grau, V., Mewes, A. U. J., Alcañiz, M., Kikinis, R., & Warfield, S. K. (2004). Improved watershed transform for medical image segmentation using prior information. *IEEE Transactions on Medical Imaging*, *23*(4), 447–458. <https://doi.org/10.1109/TMI.2004.824224>
- Greenspan, H., Ruf, A., & Goldberger, J. (2006). Constrained Gaussian mixture model framework for automatic segmentation of MR brain images. *IEEE Transactions on Medical Imaging*, *25*(9), 1233–1245. <https://doi.org/10.1109/TMI.2006.880668>
- Gros, C., De Leener, B., Badji, A., Maranzano, J., Eden, D., Dupont, S. M., Talbott, J., Zhuoquiong, R., Liu, Y., Granberg, T., Ouellette, R., Tachibana, Y., Hori, M., Kamiya, K., Chougar, L., Stawiarz, L., Hillert, J., Bannier, E., Kerbrat, A., ... Cohen-Adad, J. (2019). Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *NeuroImage*, *184*, 901–915. <https://doi.org/10.1016/j.neuroimage.2018.09.081>
- Hagström, I. T., Schneider, R., Bellenberg, B., Salmen, A., Weiler, F., Köster, O., Gold, R., & Lukas, C. (2017). Relevance of early cervical cord volume loss in the disease evolution of clinically isolated syndrome and early multiple sclerosis: a 2-year follow-up study. *Journal of Neurology*, *264*(7), 1402–1412. <https://doi.org/10.1007/S00415-017-8537-5>
- Harris, C., Costello, K., Halper, J., Hinrichs, J., Hutchinson, B., Naunton, A., & Yeung, M. (2003). Consortium of Multiple Sclerosis Centers Recommendations for Care of Those Affected by Multiple Sclerosis. *International Journal of MS Care*, *5*(3), 67–78. <https://doi.org/10.7224/1537-2073-5.3.67>
- Harris, J. O., Frank, J. A., Patronas, N., McFarlin, D. E., & McFarland, H. F. (1991). Serial gadolinium-enhanced magnetic resonance imaging scans in patients with early, relapsing-remitting multiple sclerosis: implications for clinical trials and natural history. *Annals of Neurology*, *29*(5), 548–555. <https://doi.org/10.1002/ANA.410290515>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>

Referencias

- Heaton, J. (2018). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp, ISBN: 0262035618 Article. *Genetic Programming and Evolvable Machines*, 19(1–2), 305–307.
https://books.google.com/books/about/Deep_Learning.html?hl=es&id=Np9SDQA AQBAJ
- Hittmair, K., Mallek, R., Prayer, D., Schindler, E. G., & Kollegger, H. (1555). Spinal Cord Lesions in Patients with Multiple Sclerosis: Comparison of MR Pulse Sequences. *AJNR Am J Neuroradiol*, 17(8), 1555–1565. <http://www.ajnr.org/content/17/8/>
- Hohol, M. J., Orav, E. J., & Weiner, H. L. (1999). Disease steps in multiple sclerosis: a longitudinal study comparing Disease Steps and EDSS to evaluate disease progression. <Http://Dx.Doi.Org/10.1177/135245859900500508>, 5(5), 349–354.
<https://doi.org/10.1177/135245859900500508>
- Horsfield, M. A., Sala, S., Neema, M., Absinta, M., Bakshi, A., Sormani, M. P., Rocca, M. A., Bakshi, R., & Filippi, M. (2010). Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: application in multiple sclerosis. *NeuroImage*, 50(2), 446–455.
<https://doi.org/10.1016/J.NEUROIMAGE.2009.12.121>
- Hua, Y., Guo, J., & Zhao, H. (2015). Deep Belief Networks and deep learning. *Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things, ICIT 2015*, 1–4. <https://doi.org/10.1109/ICAOT.2015.7111524>
- Ioffe, S., & Szegedy, C. (2015). *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift* (pp. 448–456). PMLR.
<https://proceedings.mlr.press/v37/ioffe15.html>
- ITK-SNAP Home*. (n.d.). Retrieved November 10, 2023, from <http://www.itksnap.org/pmwiki/pmwiki.php>
- Jamshidi, M., & Zilouchian, A. (2001). *Fundamentals of Neural Networks*. 17–38.
<https://doi.org/10.1201/9781420058147-2>
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586.
<https://doi.org/10.1016/J.BUSHOR.2018.03.007>
- Jim 9 Software | Xinapse Systems*. (n.d.). Retrieved January 17, 2024, from <https://www.xinapse.com/j-im-9-software/>
- Joyner, M. J., & Paneth, N. (2019). Promises, promises, and precision medicine. *The Journal of Clinical Investigation*, 129(3), 946–948.
<https://doi.org/10.1172/JCI126119>
- Jurman, G., Riccadonna, S., & Furlanello, C. (2012). A Comparison of MCC and CEN Error Measures in Multi-Class Prediction. *PLoS ONE*, 7(8), 41882.
<https://doi.org/10.1371/JOURNAL.PONE.0041882>
- Kallenberg, M., Petersen, K., Nielsen, M., Ng, A. Y., Diao, P., Igel, C., Vachon, C. M., Holland, K., Winkel, R. R., Karssemeijer, N., & Lillholm, M. (2016). Unsupervised Deep Learning Applied to Breast Density Segmentation and Mammographic Risk

Referencias

- Scoring. *IEEE Transactions on Medical Imaging*, 35(5), 1322–1331.
<https://doi.org/10.1109/TMI.2016.2532122>
- Kang, H., Metz, L. M., Traboulsee, A. L., Eliasziw, M., Zhao, G. J., Cheng, Y., Zhao, Y., & Li, D. K. B. (2014). Application and a proposed modification of the 2010 McDonald criteria for the diagnosis of multiple sclerosis in a Canadian cohort of patients with clinically isolated syndromes. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 20(4), 458–463. <https://doi.org/10.1177/1352458513501230>
- Kapoor, R., Furby, J., Hayton, T., Smith, K. J., Altmann, D. R., Brenner, R., Chataway, J., Hughes, R. A., & Miller, D. H. (2010). Lamotrigine for neuroprotection in secondary progressive multiple sclerosis: a randomised, double-blind, placebo-controlled, parallel-group trial. *The Lancet. Neurology*, 9(7), 681–688.
[https://doi.org/10.1016/S1474-4422\(10\)70131-9](https://doi.org/10.1016/S1474-4422(10)70131-9)
- Karimi, D., Dou, H., Warfield, S. K., & Gholipour, A. (2019). Deep learning with noisy labels: exploring techniques and remedies in medical image analysis. *Medical Image Analysis*, 65. <https://doi.org/10.1016/j.media.2020.101759>
- Kearney, H., Miller, D. H., & Ciccarelli, O. (2015). Spinal cord MRI in multiple sclerosis--diagnostic, prognostic and clinical value. *Nature Reviews. Neurology*, 11(6), 327–338. <https://doi.org/10.1038/NRNEUROL.2015.80>
- Kearney, H., Rocca, M. A., Valsasina, P., Balk, L., Sastre-Garriga, J., Reinhardt, J., Ruggieri, S., Rovira, A., Stippich, C., Kappos, L., Sprenger, T., Tortorella, P., Rovaris, M., Gasperini, C., Montalban, X., Geurts, J. J. G., Polman, C. H., Barkhof, F., Filippi, M., ... Chard, D. T. (2014). Magnetic resonance imaging correlates of physical disability in relapse onset multiple sclerosis of long disease duration. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 20(1), 72–80.
<https://doi.org/10.1177/1352458513492245>
- Klein, K., & Martinez, L. F. (2022). The impact of anthropomorphism on customer satisfaction in chatbot commerce: an experimental study in the food sector. *Electronic Commerce Research*, 23(4), 2789–2825.
<https://doi.org/10.1007/S10660-022-09562-8/FIGURES/5>
- Koh, J., Kim, T., Chaudhary, V., & Dhillon, G. (2010). Automatic segmentation of the spinal cord and the dural sac in lumbar MR images using gradient vector flow field. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2010*, 3117–3120.
<https://doi.org/10.1109/IEMBS.2010.5626097>
- Kremenutzky, M., Rice, G. P. A., Baskerville, J., Wingerchuk, D. M., & Ebers, G. C. (2006). The natural history of multiple sclerosis: a geographically based study 9: observations on the progressive phase of the disease. *Brain : A Journal of Neurology*, 129(Pt 3), 584–594. <https://doi.org/10.1093/BRAIN/AWH721>
- KumarSingh, R., Shekhar, S., Bhawan Singh, R., & Chauhan, V. (2014). A Comparative Study of Edge Detection Techniques. *International Journal of Computer Applications*, 100(19), 5–8. <https://doi.org/10.5120/17631-5949>

Referencias

- Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, *33*(11), 1444–1452. <https://doi.org/10.1212/WNL.33.11.1444>
- Lee, M. A., Smith, S., Palace, J., Narayanan, S., Silver, N., Minicucci, L., Filippi, M., Miller, D. H., Arnold, D. L., & Matthews, P. M. (1999). Spatial mapping of T2 and gadolinium-enhancing T1 lesion volumes in multiple sclerosis: evidence for distinct mechanisms of lesion genesis? *Brain : A Journal of Neurology*, *122* (Pt 7)(7), 1261–1270. <https://doi.org/10.1093/BRAIN/122.7.1261>
- Li, R., Liu, W., Yang, L., Sun, S., Hu, W., Zhang, F., & Li, W. (2017). DeepUNet: A Deep Fully Convolutional Network for Pixel-level Sea-Land Segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(11), 3954–3962. <http://arxiv.org/abs/1709.00201>
- Li, R., Liu, W., Yang, L., Sun, S., Hu, W., Zhang, F., Member, S., & Li, W. (n.d.). *DeepUNet: A Deep Fully Convolutional Network for Pixel-level Sea-Land Segmentation*.
- Lin, X., & Blumhardt, L. D. (2001). Inflammation and atrophy in multiple sclerosis: MRI associations with disease course. *Journal of the Neurological Sciences*, *189*(1–2), 99–104. [https://doi.org/10.1016/S0022-510X\(01\)00576-7](https://doi.org/10.1016/S0022-510X(01)00576-7)
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, *42*, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>
- Liu-Thompkins, Y., Okazaki, S., & Li, H. (2022). Artificial empathy in marketing interactions: Bridging the human-AI gap in affective and social customer experience. *Journal of the Academy of Marketing Science* *2022* *50*:6, *50*(6), 1198–1218. <https://doi.org/10.1007/S11747-022-00892-5>
- Liu, C., Edwards, S., Gong, Q., Roberts, N., & Blumhardt, L. D. (1999). Three dimensional MRI estimates of brain and spinal cord atrophy in multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, *66*(3), 323–330. <https://doi.org/10.1136/JNRP.66.3.323>
- Liu, X., Song, L., Liu, S., & Zhang, Y. (2021). A Review of Deep-Learning-Based Medical Image Segmentation Methods. *Sustainability* *2021*, Vol. *13*, Page 1224, *13*(3), 1224. <https://doi.org/10.3390/SU13031224>
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *Proceedings of the IEEE International Conference on Computer Vision*, 9992–10002. <https://doi.org/10.1109/ICCV48922.2021.00986>
- Lladó, X., Oliver, A., Cabezas, M., Freixenet, J., Vilanova, J. C., Quiles, A., Valls, L., Ramió-Torrent, L., & Rovira, L. (2012). Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches. *Information Sciences*, *186*(1), 164–185. <https://doi.org/10.1016/j.ins.2011.10.011>

Referencias

- Local Adaptive Thresholding. (2009). *Encyclopedia of Biometrics*, 939–939.
https://doi.org/10.1007/978-0-387-73003-5_506
- Losseff, N. A., Webb, S. L., O’Riordan, J. I., Page, R., Wang, L., Barker, G. J., Tofts, P. S., McDonald, W. I., Miller, D. H., & Thompson, A. J. (1996a). Spinal cord atrophy and disability in multiple sclerosis. A new reproducible and sensitive MRI method with potential to monitor disease progression. *Brain*, *119*(3), 701–708.
<https://doi.org/10.1093/brain/119.3.701>
- Losseff, N. A., Webb, S. L., O’Riordan, J. I., Page, R., Wang, L., Barker, G. J., Tofts, P. S., McDonald, W. I., Miller, D. H., & Thompson, A. J. (1996b). Spinal cord atrophy and disability in multiple sclerosis. A new reproducible and sensitive MRI method with potential to monitor disease progression. *Brain : A Journal of Neurology*, *119* (Pt 3)(3), 701–708. <https://doi.org/10.1093/BRAIN/119.3.701>
- Lublin, F. D., Baier, M., & Cutter, G. (2003). Effect of relapses on development of residual deficit in multiple sclerosis. *Neurology*, *61*(11), 1528–1532.
<https://doi.org/10.1212/01.WNL.0000096175.39831.21>
- Lublin, F. D., & Reingold, S. C. (1996). Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology*, *46*(4), 907–911. <https://doi.org/10.1212/WNL.46.4.907>
- Lublin, F. D., Reingold, S. C., Cohen, J. A., Cutter, G. R., Sørensen, P. S., Thompson, A. J., Wolinsky, J. S., Balcer, L. J., Banwell, B., Barkhof, F., Bebo, B., Calabresi, P. A., Clanet, M., Comi, G., Fox, R. J., Freedman, M. S., Goodman, A. D., Inglese, M., Kappos, L., ... Polman, C. H. (2014). Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology*, *83*(3), 278–286.
<https://doi.org/10.1212/WNL.0000000000000560>
- Lukas, C., Knol, D. L., Sombekke, M. H., Bellenberg, B., Hahn, H. K., Popescu, V., Weier, K., Radue, E. W., Gass, A., Kappos, L., Naegelin, Y., Uitdehaag, B. M. J., Geurts, J. J. G., Barkhof, F., & Vrenken, H. (2015). Cervical spinal cord volume loss is related to clinical disability progression in multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, *86*(4), 410–418. <https://doi.org/10.1136/JNNP-2014-308021>
- Lukas, C., Sombekke, M. H., Bellenberg, B., Hahn, H. K., Popescu, V., Bendfeldt, K., Radue, E. W., Gass, A., Borgwardt, S. J., Kappos, L., Naegelin, Y., Knol, D. L., Polman, C. H., Geurts, J. J. G., Barkhof, F., & Vrenken, H. (2013a). Relevance of spinal cord abnormalities to clinical disability in multiple sclerosis: MR imaging findings in a large cohort of patients. *Radiology*, *269*(2), 542–552.
<https://doi.org/10.1148/RADIOLOGY.13122566>
- Lukas, C., Sombekke, M. H., Bellenberg, B., Hahn, H. K., Popescu, V., Bendfeldt, K., Radue, E. W., Gass, A., Borgwardt, S. J., Kappos, L., Naegelin, Y., Knol, D. L., Polman, C. H., Geurts, J. J. G., Barkhof, F., & Vrenken, H. (2013b). Relevance of spinal cord abnormalities to clinical disability in multiple sclerosis: MR imaging findings in a large cohort of patients. *Radiology*, *269*(2), 542–552.
<https://doi.org/10.1148/RADIOLOGY.13122566>

Referencias

- Lycklama À Nijeholt, G., & Barkhof, F. (2003). Differences between subgroups of MS: MRI findings and correlation with histopathology. *Journal of the Neurological Sciences*, 206(2), 173–174. [https://doi.org/10.1016/S0022-510X\(02\)00336-2](https://doi.org/10.1016/S0022-510X(02)00336-2)
- Martí-Bonmatí, L., & Alberich-Bayarri, A. (2016). Imaging biomarkers: Development and clinical integration. *Imaging Biomarkers: Development and Clinical Integration*, 1–376. <https://doi.org/10.1007/978-3-319-43504-6/COVER>
- McCoy, D. B., Dupont, S. M., Gros, C., Cohen-Adad, J., Huie, R. J., Ferguson, A., Duong-Fernandez, X., Thomas, L. H., Singh, V., Narvid, J., Pascual, L., Kyritsis, N., Beattie, M. S., Bresnahan, J. C., Dhall, S., Whetstone, W., & Talbott, J. F. (2019). Convolutional neural network–based automated segmentation of the spinal cord and contusion injury: Deep learning biomarker correlates of motor impairment in acute spinal cord injury. *American Journal of Neuroradiology*, 40(4), 737–744. <https://doi.org/10.3174/ajnr.A6020>
- McGowan, J. C. (2000). Technical issues for MRI examination of the spinal cord. *Journal of the Neurological Sciences*, 172(SUPPL. 1). [https://doi.org/10.1016/S0022-510X\(99\)00274-9](https://doi.org/10.1016/S0022-510X(99)00274-9)
- Meier, D. S., Guttmann, C. R. G., Tummala, S., Moscufo, N., Cavallari, M., Tauhid, S., Bakshi, R., & Weiner, H. L. (2018). Dual-Sensitivity Multiple Sclerosis Lesion and CSF Segmentation for Multichannel 3T Brain MRI. *Journal of Neuroimaging : Official Journal of the American Society of Neuroimaging*, 28(1), 36–47. <https://doi.org/10.1111/JON.12491>
- Miller, D. H., Chard, D. T., & Ciccarelli, O. (2012). Clinically isolated syndromes. *The Lancet Neurology*, 11(2), 157–169. [https://doi.org/10.1016/S1474-4422\(11\)70274-5](https://doi.org/10.1016/S1474-4422(11)70274-5)
- Miller, R. H., Fyffe-Maricich, S., & Caprariello, A. C. (2017). Animal Models for the Study of Multiple Sclerosis. *Animal Models for the Study of Human Disease: Second Edition*, 967–988. <https://doi.org/10.1016/B978-0-12-809468-6.00037-1>
- Mina, Y., Azodi, S., Dubuche, T., Andrada, F., Osuorah, I., Ohayon, J., Cortese, I., Wu, T., Johnson, K. R., Reich, D. S., Nair, G., & Jacobson, S. (2021). Cervical and thoracic cord atrophy in multiple sclerosis phenotypes: Quantification and correlation with clinical disability. *NeuroImage: Clinical*, 30, 102680. <https://doi.org/10.1016/J.NICL.2021.102680>
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature* 2015 518:7540, 518(7540), 529–533. <https://doi.org/10.1038/nature14236>
- Neuroimaging in Python - Pipelines and Interfaces — nipy pipeline and interfaces package*. (n.d.). Retrieved February 1, 2024, from <https://nipy.readthedocs.io/en/latest/api/generated/nipy.interfaces.dcm2nii.html>
- NITRC: MRicroGL: Tool/Resource Info. (n.d.). Retrieved February 1, 2024, from

Referencias

- <https://www.nitrc.org/projects/microgl>
- Norouzi, A., Rahim, M. S. M., Altameem, A., Saba, T., Rad, A. E., Rehman, A., & Uddin, M. (2014). Medical image segmentation methods, algorithms, and applications. *IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India)*, 31(3), 199–213. <https://doi.org/10.1080/02564602.2014.906861>
- NVIDIA and IBM Cloud Support ImageNet Large Scale Visual Recognition Challenge | NVIDIA Technical Blog*. (n.d.). Retrieved November 10, 2023, from <https://developer.nvidia.com/blog/nvidia-ibm-cloud-support-imagenet-large-scale-visual-recognition-challenge/>
- Oh, J., Sotirchos, E. S., Saidha, S., Whetstone, A., Chen, M., Newsome, S. D., Zackowski, K., Balcer, L. J., Frohman, E., Prince, J., Diener-West, M., Reich, D. S., & Calabresi, P. A. (2015). Relationships between quantitative spinal cord MRI and retinal layers in multiple sclerosis. *Neurology*, 84(7), 720. <https://doi.org/10.1212/WNL.0000000000001257>
- Oktay, O., Schlemper, J., Folgoc, L. Le, Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). *Attention U-Net: Learning Where to Look for the Pancreas*. <https://arxiv.org/abs/1804.03999v3>
- Okuda, D. T., Siva, A., Kantarci, O., Inglese, M., Katz, I., Tutuncu, M., Keegan, B. M., Donlon, S., Hua, L. H., Vidal-Jordana, A., Montalban, X., Rovira, A., Tintoré, M., Amato, M. P., Brochet, B., De Seze, J., Brassat, D., Vermersch, P., De Stefano, N., ... Lebrun, C. (2014). Radiologically isolated syndrome: 5-year risk for an initial clinical event. *PLoS One*, 9(3). <https://doi.org/10.1371/JOURNAL.PONE.0090509>
- Oldenhuis, C. N. A. M., Oosting, S. F., Gietema, J. A., & de Vries, E. G. E. (2008). Prognostic versus predictive value of biomarkers in oncology. *European Journal of Cancer*, 44(7), 946–953. <https://doi.org/10.1016/J.EJCA.2008.03.006>
- Overview of the epidemiology, diagnosis, and disease progression associated with multiple sclerosis - PubMed*. (n.d.). Retrieved July 31, 2023, from <https://pubmed.ncbi.nlm.nih.gov/23544716/>
- Papinutto, N., Bakshi, R., Bischof, A., Calabresi, P. A., Caverzasi, E., Constable, R. T., Datta, E., Kirkish, G., Nair, G., Oh, J., Pelletier, D., Pham, D. L., Reich, D. S., Rooney, W., Roy, S., Schwartz, D., Shinohara, R. T., Sicotte, N. L., Stern, W. A., ... Henry, R. G. (2018). Gradient nonlinearity effects on upper cervical spinal cord area measurement from 3D T1-weighted brain MRI acquisitions. *Magnetic Resonance in Medicine*, 79(3), 1595–1601. <https://doi.org/10.1002/MRM.26776>
- Pezold, S., Fundana, K., Amann, M., Andelova, M., Pfister, A., Sprenger, T., & Cattin, P. C. (2015). Automatic segmentation of the spinal cord using continuous max flow with cross-sectional similarity prior and tubularity features. *Lecture Notes in Computational Vision and Biomechanics*, 20, 107–118. https://doi.org/10.1007/978-3-319-14148-0_10
- Popovic, A., de la Fuente, M., Engelhardt, M., & Radermacher, K. (2007). Statistical

Referencias

- validation metric for accuracy assessment in medical image segmentation. *International Journal of Computer Assisted Radiology and Surgery*, 2(3–4), 169–181. <https://doi.org/10.1007/S11548-007-0125-1/METRICS>
- Poser, C. M., Paty, D. W., Scheinberg, L., McDonald, W. I., Davis, F. A., Ebers, G. C., Johnson, K. P., Sibley, W. A., Silberberg, D. H., & Tourtellotte, W. W. (1983). New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Annals of Neurology*, 13(3), 227–231. <https://doi.org/10.1002/ANA.410130302>
- Powers, D. M. W., & Ailab. (2020). *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation*. <https://arxiv.org/abs/2010.16061v1>
- Prastawa, M., Bullitt, E., & Gerig, G. (2009). Simulation of brain tumors in MR images for evaluation of segmentation efficacy. *Medical Image Analysis*, 13(2), 297–311. <https://doi.org/10.1016/J.MEDIA.2008.11.002>
- Ransohoff, R. M., Hafler, D. A., & Lucchinetti, C. F. (2015). Multiple sclerosis—a quiet revolution. *Nature Reviews. Neurology*, 11(3), 134–142. <https://doi.org/10.1038/NRNEUROL.2015.14>
- Rocca, M. A., Horsfield, M. A., Sala, S., Copetti, M., Valsasina, P., Mesaros, S., Martinelli, V., Caputo, D., Stosic-Opincal, T., Drulovic, J., Comi, G., & Filippi, M. (2011). A multicenter assessment of cervical cord atrophy among MS clinical phenotypes. *Neurology*, 76(24), 2096–2102. <https://doi.org/10.1212/WNL.0B013E31821F46B8>
- Rocca, M. A., Valsasina, P., Meani, A., Gobbi, C., Zecca, C., Rovira, À., Montalban, X., Kearney, H., Ciccarelli, O., Matthews, L., Palace, J., Gallo, A., Biseco, A., Gass, A., Eisele, P., Lukas, C., Bellenberg, B., Barkhof, F., Vrenken, H., ... Filippi, M. (2019). Clinically relevant cranio-caudal patterns of cervical cord atrophy evolution in MS. *Neurology*, 93(20), E1852–E1866. <https://doi.org/10.1212/WNL.0000000000008466>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- Rovaris, M., Confavreux, C., Furlan, R., Kappos, L., Comi, G., & Filippi, M. (2006). Secondary progressive multiple sclerosis: current knowledge and future challenges. *The Lancet. Neurology*, 5(4), 343–354. [https://doi.org/10.1016/S1474-4422\(06\)70410-0](https://doi.org/10.1016/S1474-4422(06)70410-0)
- Rovira, À., Wattjes, M. P., Tintoré, M., Tur, C., Yousry, T. A., Sormani, M. P., De Stefano, N., Filippi, M., Auger, C., Rocca, M. A., Barkhof, F., Fazekas, F., Kappos, L., Polman, C., Miller, D., & Montalban, X. (2015). MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis—clinical implementation in the diagnostic process. *Nature Reviews Neurology* 2015 11:8, 11(8), 471–482. <https://doi.org/10.1038/nrneuro.2015.106>
- Roy, S., Butman, J. A., Reich, D. S., Calabresi, P. A., & Pham, D. L. (2018). *Multiple*

Referencias

- Sclerosis Lesion Segmentation from Brain MRI via Fully Convolutional Neural Networks*. <https://arxiv.org/abs/1803.09172v1>
- Sand, I. K., Krieger, S., Farrell, C., & Miller, A. E. (2014). Diagnostic uncertainty during the transition to secondary progressive multiple sclerosis. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 20(12), 1654–1657. <https://doi.org/10.1177/1352458514521517>
- Singhal, T., Tauhid, S., Hurwitz, S., Neema, M., & Bakshi, R. (2017). The Effect of Glatiramer Acetate on Spinal Cord Volume in Relapsing-Remitting Multiple Sclerosis. *Journal of Neuroimaging : Official Journal of the American Society of Neuroimaging*, 27(1), 33–36. <https://doi.org/10.1111/JON.12378>
- Smith, M. E., Stone, L. A., Albert, P. S., Frank, J. A., Martin, R., Armstrong, M., Maloni, H., McFarlin, D. E., & McFarland, H. F. (1993). Clinical worsening in multiple sclerosis is associated with increased frequency and area of gadopentetate dimeglumine-enhancing magnetic resonance imaging lesions. *Annals of Neurology*, 33(5), 480–489. <https://doi.org/10.1002/ANA.410330511>
- Sombekke, M. H., Wattjes, M. P., Balk, L. J., Nielsen, J. M., Vrenken, H., Uitdehaag, B. M. J., Polman, C. H., & Barkhof, F. (2013). Spinal cord lesions in patients with clinically isolated syndrome: a powerful tool in diagnosis and prognosis. *Neurology*, 80(1), 69–75. <https://doi.org/10.1212/WNL.0B013E31827B1A67>
- Sonawane, M. S., & Dhawale, C. A. (2015). A Brief Survey on Image Segmentation Methods. *IJCA Proceedings on National Conference on Digital Image and Signal Processing, DISP 2015(1)*, 1–5. <http://www.ijcaonline.org/proceedings/disp2015/number1/20475-3002>
- Song, X., Li, D., Qiu, Z., Su, S., Wu, Y., Wang, J., Liu, Z., & Dong, H. (2020). Correlation between EDSS scores and cervical spinal cord atrophy at 3T MRI in multiple sclerosis: A systematic review and meta-analysis. *Multiple Sclerosis and Related Disorders*, 37. <https://doi.org/10.1016/J.MSARD.2019.101426>
- Standardized MR imaging protocol for multiple sclerosis: Consortium of MS Centers consensus guidelines - PubMed*. (n.d.). Retrieved July 31, 2023, from <https://pubmed.ncbi.nlm.nih.gov/16484429/>
- Stevenson, V. L., Leary, S. M., Losseff, N. A., Parker, G. J. M., Barker, G. J., Husmani, Y., Miller, D. H., & Thompson, A. J. (1998). Spinal cord atrophy and disability in MS: a longitudinal study. *Neurology*, 51(1), 234–238. <https://doi.org/10.1212/WNL.51.1.234>
- Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Jorge Cardoso, M. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10553 LNCS, 240–248. https://doi.org/10.1007/978-3-319-67558-9_28/COVER
- Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging*, 15(1), 1–28. <https://doi.org/10.1186/S12880-015-0068-X/TABLES/5>

Referencias

- Tang, L., Wen, Y., Zhou, Z., von Deneen, K. M., Huang, D., & Ma, L. (2013). Reduced field-of-view DTI segmentation of cervical spine tissue. *Magnetic Resonance Imaging, 31*(9), 1507–1514. <https://doi.org/10.1016/J.MRI.2013.07.003>
- The Pareto Principle | Dunford | The Plymouth Student Scientist.* (n.d.). Retrieved December 9, 2019, from <https://bcurl.org/journals/index.php/TPSS/article/view/408>
- Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., ... Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet. Neurology, 17*(2), 162–173. [https://doi.org/10.1016/S1474-4422\(17\)30470-2](https://doi.org/10.1016/S1474-4422(17)30470-2)
- Tintore, M., Rovira, A., Arrambide, G., Mitjana, R., Río, J., Auger, C., Nos, C., Edo, M. C., Castelló, J., Horga, A., Perez-Miralles, F., Huerga, E., Comabella, M., Sastre-Garriga, J., & Montalban, X. (2010). Brainstem lesions in clinically isolated syndromes. *Neurology, 75*(21), 1933–1938. <https://doi.org/10.1212/WNL.0B013E3181FEB26F>
- Touvron, H., Cord, M., Sablayrolles, A., Synnaeve, G., & Jégou, H. (2021). Going deeper with Image Transformers. *Proceedings of the IEEE International Conference on Computer Vision, 32–42*. <https://doi.org/10.1109/ICCV48922.2021.00010>
- Traboulsee, A., Simon, J. H., Stone, L., Fisher, E., Jones, D. E., Malhotra, A., Newsome, S. D., Oh, J., Reich, D. S., Richert, N., Rammohan, K., Khan, O., Radue, E. W., Ford, C., Halper, J., & Li, D. (2016). Revised Recommendations of the Consortium of MS Centers Task Force for a Standardized MRI Protocol and Clinical Guidelines for the Diagnosis and Follow-Up of Multiple Sclerosis. *AJNR. American Journal of Neuroradiology, 37*(3), 394–401. <https://doi.org/10.3174/AJNR.A4539>
- Tsagkas, C., Magon, S., Gaetano, L., Pezold, S., Naegelin, Y., Amann, M., Stippich, C., Cattin, P., Wuerfel, J., Bieri, O., Sprenger, T., Kappos, L., & Parmar, K. (2018). Spinal cord volume loss: A marker of disease progression in multiple sclerosis. *Neurology, 91*(4), e349–e358. <https://doi.org/10.1212/WNL.0000000000005853>
- Tsagkas, C., Magon, S., Gaetano, L., Pezold, S., Naegelin, Y., Amann, M., Stippich, C., Cattin, P., Wuerfel, J., Bieri, O., Sprenger, T., Kappos, L., & Parmar, K. (2019). Preferential spinal cord volume loss in primary progressive multiple sclerosis. *Multiple Sclerosis (Houndmills, Basingstoke, England), 25*(7), 947–957. <https://doi.org/10.1177/1352458518775006>
- Tur, C., Moccia, M., Barkhof, F., Chataway, J., Sastre-Garriga, J., Thompson, A. J., & Ciccarelli, O. (2018). Assessing treatment outcomes in multiple sclerosis trials and in the clinical setting. *Nature Reviews. Neurology, 14*(2), 75–93. <https://doi.org/10.1038/NRNEUROL.2017.171>
- Ukkonen, M., Dastidar, P., Heinonen, T., Laasonen, E., & Elovaara, I. (2003). Volumetric quantitation by MRI in primary progressive multiple sclerosis: volumes of plaques and atrophy correlated with neurological disability. *European Journal of Neurology, 10*(6), 663–669. <https://doi.org/10.1046/J.1468-1331.2003.00617.X>

Referencias

- Valsasina, P., Rocca, M. A., Horsfield, M. A., Absinta, M., Messina, R., Caputo, D., Comi, G., & Filippi, M. (2013). Regional cervical cord atrophy and disability in multiple sclerosis: A voxel-based analysis. *Radiology*, *266*(3), 853–861. <https://doi.org/10.1148/radiol.12120813>
- Valsecchi, A., Bermejo, E., Damas, S., & Cordón, O. (2018). Metaheuristics for medical image registration. *Handbook of Heuristics*, *2–2*, 1079–1101. https://doi.org/10.1007/978-3-319-07124-4_56/COVER
- Valverde, S., Cabezas, M., Roura, E., González-Villà, S., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Rovira, À., Oliver, A., & Lladó, X. (2017). Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. *Neuroimage*, *155*, 159–168. <https://doi.org/10.1016/J.NEUROIMAGE.2017.04.034>
- Valverde, S., Oliver, A., Roura, E., González-Villà, S., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Rovira, À., & Lladó, X. (2017). Automated tissue segmentation of MR brain images in the presence of white matter lesions. *Medical Image Analysis*, *35*, 446–457. <https://doi.org/10.1016/J.MEDIA.2016.08.014>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems, 2017-December*, 5999–6009. <https://arxiv.org/abs/1706.03762v7>
- Vincent, O., & Folorunso, O. (2009). A Descriptive Algorithm for Sobel Image Edge Detection. *Ith SITE 2009: Informing Science + IT Education Conference*, *9*. <https://doi.org/10.28945/3351>
- Vollmer, T. (2007). The natural history of relapses in multiple sclerosis. *Journal of the Neurological Sciences*, *256 Suppl 1*(SUPPL. 1). <https://doi.org/10.1016/J.JNS.2007.01.065>
- Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., & Xiao, B. (2019). Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(10), 3349–3364. <https://doi.org/10.1109/TPAMI.2020.2983686>
- Wang, Z., Zou, Y., & Liu, P. X. (2021). Hybrid dilation and attention residual U-Net for medical image segmentation. *Computers in Biology and Medicine*, *134*. <https://doi.org/10.1016/j.compbiomed.2021.104449>
- Weiler, F., Daams, M., Lukas, C., Barkhof, F., & Hahn, H. K. (2015). Highly accurate volumetry of the spinal cord. *Medical Imaging 2015: Image Processing*, *9413*, 941302. <https://doi.org/10.1117/12.2080803>
- Welcome! — *Anaconda documentation*. (n.d.). Retrieved November 10, 2023, from <https://docs.anaconda.com/>
- Westbrook, C., & Talbot, J. (Writer on magnetic resonance imaging). (n.d.). *MRI in practice*. 395. Retrieved August 1, 2023, from <https://www.wiley.com/en-sg/MRI+in+Practice%2C+5th+Edition-p-9781119392002>

Referencias

- What DALL-E Reveals About Human Creativity*. (n.d.). Retrieved August 30, 2023, from <https://hai.stanford.edu/news/what-dall-e-reveals-about-human-creativity>
- Xu, Z., Zhou, Y., Wang, S., Wang, L., Li, F., Wang, S., & Wang, Z. (2020). A novel intelligent classification method for urban green space based on high-resolution remote sensing images. *Remote Sensing*, *12*(22), 1–18. <https://doi.org/10.3390/RS12223845>
- Yin, W., Kann, K., Yu, M., SchützeSch, H., & Munich, L. (2017). *Comparative Study of CNN and RNN for Natural Language Processing*. <https://arxiv.org/abs/1702.01923v1>
- Zahavy, T., Veeriah, V., Hou, S., Waugh, K., Lai, M., Leurent, E., Tomasev, N., Schut, L., Hassabis, D., & Singh, S. (2023). *Diversifying AI: Towards Creative Chess with AlphaZero*. 2023–2039. <https://arxiv.org/abs/2308.09175v1>
- Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2016). Pyramid Scene Parsing Network. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-January*, 6230–6239. <https://doi.org/10.1109/CVPR.2017.660>
- Zhou, Z. H. (2018). A brief introduction to weakly supervised learning. *National Science Review*, *5*(1), 44–53. <https://doi.org/10.1093/NSR/NWX106>
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11045 LNCS*, 3–11. https://doi.org/10.1007/978-3-030-00889-5_1
- Ziemssen, T. (2009). Multiple sclerosis beyond EDSS: depression and fatigue. *Journal of the Neurological Sciences*, *277*(SUPPL. 1), S37–S41. [https://doi.org/10.1016/S0022-510X\(09\)70011-5](https://doi.org/10.1016/S0022-510X(09)70011-5)

Anexo I: Detección automática de lesiones en médula cervical

Anexo I: Detección automática de lesiones en médula cervical

En este anexo, se explican los futuros objetivos, los cuales se encuentran en línea con el método de segmentación automática de médula espinal cervical desarrollado en esta tesis, y se recoge y describe el trabajo que se ha invertido en el desarrollo de estos propósitos hasta el momento. El modelo de segmentación, ya implementado, será la base para la elaboración de un método agregado de detección de lesiones en médula cervical a partir de secuencias de imagen de RM STIR y Sagital T2, siendo utilizado como paso previo para obtener la localización de la médula cervical en el espacio de la secuencia STIR tras el correcto registro de imágenes entre la secuencia axial 3D T1 donde se ha segmentado la médula cervical y la STIR donde se pretenden localizar las lesiones.

1 Introducción

Un síntoma común en los pacientes de EM, además de los mencionados en la sección 1.1 de esta tesis, es la presencia de daño en la médula espinal en forma de placas o lesiones que se forman durante el proceso inflamatorio y destrucción de la mielina, provocando diversas alteraciones neurológicas cuya gravedad y tipo de síntomas dependen de su ubicación y extensión. Sin embargo, es importante destacar que no todas las personas con EM experimentan dichas lesiones, y no todas las lesiones en la médula espinal son causadas por la EM. Estas lesiones pueden estar causadas por otras afecciones, como traumatismos, infecciones o tumores. Por lo tanto, es necesario realizar una evaluación médica completa para determinar la causa exacta y los síntomas asociados.

Las lesiones de la médula espinal están presentes en hasta el 90% de los pacientes con EM (J. C. J. Bot et al., 2004) y se visualizan en la RM como hiperintensidades en T2 en forma de cuña que se extienden menos de tres segmentos. Pueden presentarse en cualquier nivel de la médula espinal, pero se encuentran con mayor frecuencia en el nivel cervical (Kearney et al., 2015; Losseff et al., 1996a). Evaluar estas lesiones puede resultar complicado debido a la respiración y las pulsaciones de la sangre, pero tienen un papel fundamental en la estimación del pronóstico de la enfermedad (Hittmair et al., 1555; McGowan, 2000).

Su detección ayuda a confirmar el diagnóstico de la EM y a diferenciarla de otras afecciones que pueden presentar síntomas similares. Su tamaño y ubicación pueden proporcionar información sobre la gravedad de la enfermedad y su impacto en la función neurológica del paciente, lo que puede ayudar a los clínicos a determinar el pronóstico y a planificar el tratamiento más adecuado. Al igual que en el caso de la atrofia medular, la detección y el seguimiento regular de las lesiones en la médula espinal pueden proporcionar información sobre la evolución de la enfermedad a lo largo del tiempo, lo que permite ajustar el tratamiento, de ser necesario, y

Anexo I: Detección automática de lesiones en médula cervical

evaluar su efectividad. Estas lesiones, pueden llevar a complicaciones adicionales como problemas de movilidad, debilidad muscular, alteraciones del control de la vejiga y del intestino, entre otros, por lo que su detección temprana puede ayudar a prevenir o tratar estas complicaciones de manera oportuna.

2 Lesiones en médula espinal

La lesión en EM se define, a nivel de imagen radiológica, como un área de hiperintensidad focal. Normalmente estas lesiones presentan forma ovoide y un diámetro de un tamaño que puede variar entre unos pocos milímetros hasta más de 1 o 2 cm. Para cumplir el criterio diagnóstico, deben de tener al menos 3 mm de longitud, aunque la topografía también debe tomarse en cuenta. Por otra parte, en su identificación en imagen médica, las lesiones deberían ser visibles al menos en dos cortes de imagen consecutivos para eliminar la posibilidad de que sean artefactos o pequeñas hiperintensidades no relacionadas con la enfermedad, excepto en el caso de que las adquisiciones tengan un espesor de corte mayor (>3 mm), en el que sí que se puede considerar lesión, aunque aparezca en una única imagen (Filippi et al., 2019).

En cerebro, las lesiones en EM normalmente se desarrollan en ambos hemisferios cerebrales, aunque su distribución suele ser ligeramente asimétrica en las primeras fases de la enfermedad. Pueden darse en cualquier zona del sistema nervioso, pero si lo comparamos con otros desórdenes causantes de lesiones en la materia blanca, la EM tiende a afectar a zonas específicas, como la materia blanca periventricular y yuxtacortical, el cuerpo calloso, el área infratentorial (sobre todo en el puente de Varolio y cerebelo) y la médula espinal (principalmente en el segmento cervical).

Las lesiones en la médula espinal suelen ser múltiples y pequeñas en el diámetro cráneo-caudal. Son hiperintensas en secuencias de potenciación T2 y se pueden dar a lo largo de toda la médula espinal, tanto a nivel cervical, torácico o lumbar, aunque en la porción cervical la aparición suele ser mayor. Para asegurarnos de que las lesiones no son artefactos de la imagen, es una buena práctica identificarlas en al menos dos secuencias (normalmente T2 y STIR) o dos planos.

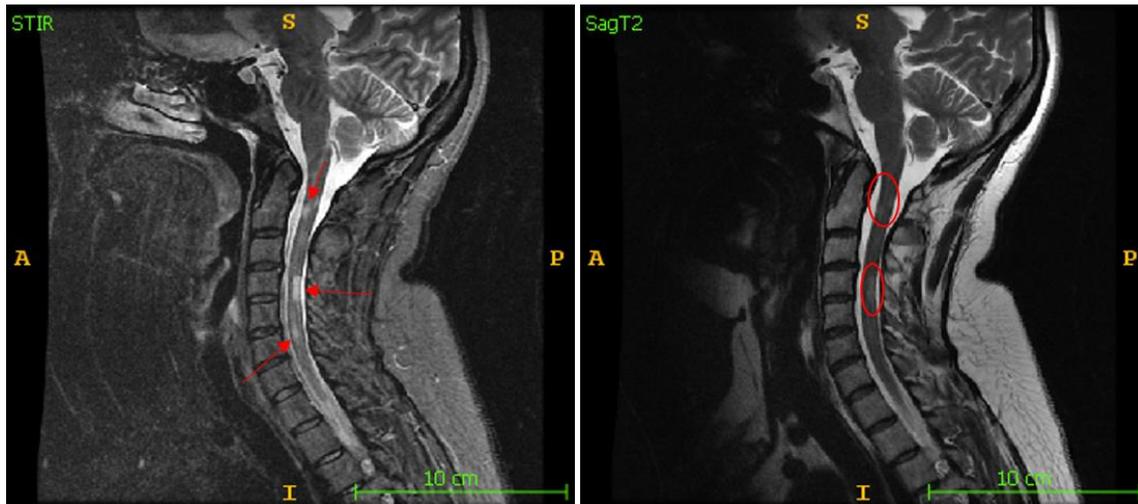


Figura 63. Secuencias procedentes de la misma adquisición. STIR (izquierda), mayor sensibilidad, presencia de claras hiperintensidades focales en la médula espinal, señaladas en rojo. Sagital T2 (derecha), mayor especificidad, falta de claridad en las supuestas lesiones, las que se intuyen, rodeadas en rojo.

A lo largo del tiempo se han observado dos tipos diferentes de lesiones en pacientes con EM: discretas (focales) y áreas difusas con una intensidad de señal intermedia que carecen de un borde bien delimitado. Estas matizaciones no están dentro del criterio actual (Thompson et al., 2018) de diagnóstico de esclerosis múltiple ya que todavía no cuentan con la suficiente especificidad y fiabilidad. Por lo tanto, actualmente, para que las lesiones en médula espinal se consideren válidas para el diagnóstico de la EM, deben de ser focales, con un borde claramente delimitado, forma lenticular en el plano sagital y forma de cuña en el axial.

Son frecuentemente lesiones pequeñas (aunque, como hemos mencionado anteriormente, de al menos 3 mm) y que ocupen menos de tres segmentos vertebrales y menos de la mitad del área de la médula. En las imágenes axiales, la mayoría de las lesiones se encuentran en la periferia de la médula espinal, principalmente en la parte de la columna lateral o dorsal, aunque también pueden afectar a la materia blanca anterior y la materia gris central. Las lesiones focales que están estrictamente limitadas a la materia gris no son usuales en EM.

Estas lesiones en la médula espinal no se suelen ver en lo que se considera un deterioro humano natural o en desórdenes neurológicos más corrientes como la migraña o algunas enfermedades cerebrovasculares. Sí que se pueden dar en algunas condiciones como en la mielopatía espondilótica o cuando existe presencia de fístulas durales o malformaciones arteriovenosas.

Podemos observar indicios de que lo que estamos observando no se debe considerar una lesión en lo que respecta a EM, lo más significantes son: la extensión a lo largo de tres o más segmentos vertebrales, la mayor implicación en la materia gris central, la presencia de inflamación en la

Anexo I: Detección automática de lesiones en médula cervical

médula espinal, la mayor afectación en la parte más caudal de la médula, la afectación leptomeníngea o de raíces nerviosas, la cavitación de la médula espinal o la evidencia de micro o macrohemorragias.

3 Secuencias empleadas

Para el objetivo de detección de lesiones en la médula espinal empleamos la secuencia STIR y la Sagital T2, siguiendo las orientaciones de nuestros radiólogos y coincidiendo éstas con lo recomendado en guías clínicas redactadas por comités oficiales (Rovira et al., 2015; *Standardized MR Imaging Protocol for Multiple Sclerosis: Consortium of MS Centers Consensus Guidelines - PubMed*, n.d.). De este modo, se pretende combinar la sensibilidad de la secuencia Sagital STIR para buscar todo lo que pueda ser potencialmente una lesión, con la especificidad de la secuencia Sagital T2 para confirmar si verdaderamente lo es. En la Figura 63 podemos ver un ejemplo visual de ello.

3.1 Secuencia Sagital T2

La secuencia sagital T2-FSE (*Fast-Spin-Echo*) es una adquisición estándar en RM neurológica y se caracteriza por la alta señal que muestran los fluidos corporales. El término T2 se refiere a la relajación transversal, una propiedad física de los tejidos que se mide en la RM. Los tejidos tienen diferentes tiempos de relajación T1 y T2, lo que significa que se recuperan a diferentes velocidades después de haber sido excitados por el campo magnético. La secuencia sagital T2 se configura para resaltar las características anatómicas y patológicas que tienen un tiempo de relajación T2 largo, como el líquido cefalorraquídeo (LCR), el cual aparece hiperintenso (brillante) debido a su alto contenido de agua y su largo tiempo de relajación T2. Las estructuras de la médula espinal, como la sustancia gris y blanca, también se visualizan, aunque con menor intensidad. Las lesiones patológicas, como tumores, quistes o hernias discales, pueden ser identificadas y evaluadas en función de su apariencia y ubicación en estas imágenes. El hecho de que esta secuencia muestre hiperintensidad en las posibles lesiones junto al hecho de que tenga un nivel relativamente bajo de artefactos y una buena especificidad, hace que sea una secuencia muy utilizada en neurología. Un ejemplo de secuencia sagital T2 se encuentra en la Figura 64.

Anexo I: Detección automática de lesiones en médula cervical



Figura 64. Imagen sagital potenciada T2. La médula cervical aparece como una estructura gris oscura rodeada por el LCR y grasa intraespinal.

Para la realización de una secuencia Sagital T2, el paciente se ha de colocar en posición supina (tumbado boca arriba) en el escáner de RM. Se utilizan técnicas de supresión de grasa para eliminar la señal de los tejidos grasos circundantes y mejorar la visibilidad de la médula espinal. Se adquieren múltiples cortes sagitales de la columna vertebral, cubriendo todo el rango de la médula espinal, desde la región cervical hasta la región lumbar, como en el ejemplo de la Figura 65.

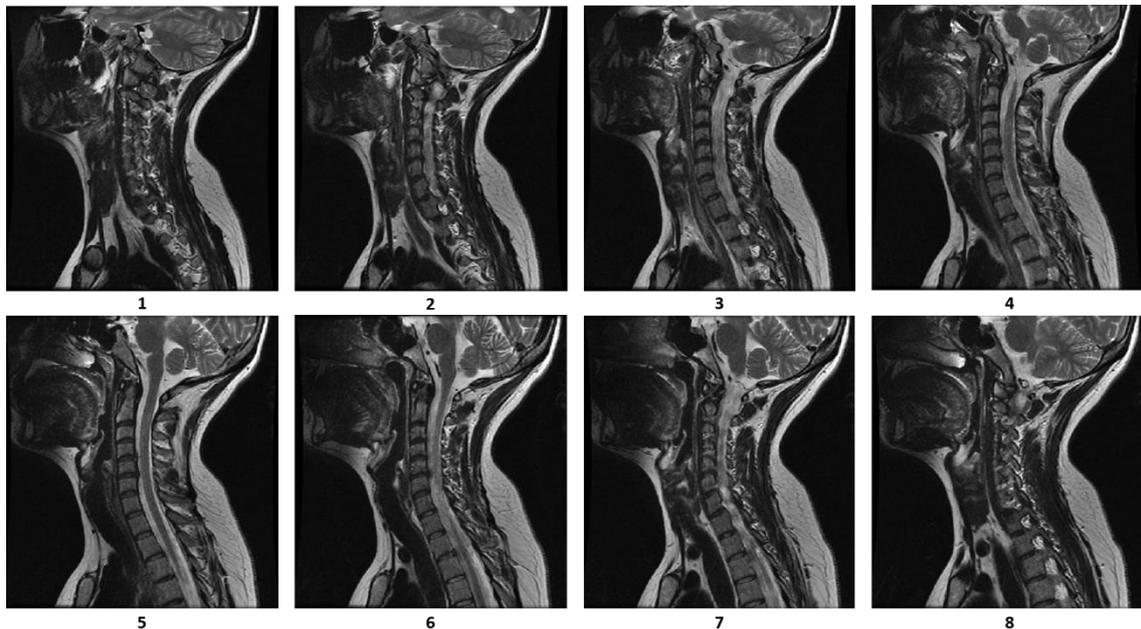


Figura 65. Distintos cortes sagitales componentes de una secuencia Sagital T2.

Anexo I: Detección automática de lesiones en médula cervical

Como se puede observar, la secuencia produce una buena señal, ya que tanto el agua como la grasa contribuyen a la imagen. La contribución de la grasa proviene del uso de muchos ecos en un tren de ecos, característico de la técnica FSE, por lo tanto, cuando muchos ecos son recolectados, algunos de ellos se medirán temporalmente muy cerca a la excitación y contendrán más señal de la grasa. El contraste T2 en general se produce midiendo grandes frecuencias espaciales en tiempos de eco correspondientes al contraste de imagen deseado. El tiempo de eco deseado es el tiempo de eco efectivo, lo cual suele estar controlado por el valor de entrada TE (tiempo de eco) del escáner de RM (Westbrook & Talbot, n.d.).

Como las secuencias T2 tienen una señal inherente relativamente alta, generalmente es posible incrementar la resolución en relación con otras frecuencias, como la STIR, sin la necesidad de aumentar el tiempo de adquisición. Una resolución mayor puede ser útil para la detección de las lesiones.

3.2 Secuencia STIR

La secuencia STIR (*Short Tau Inversion Recovery*) es una secuencia de RM que se utiliza para la detección de lesiones en la médula espinal. Es especialmente útil para visualizar estructuras con alto contenido de agua, como la médula espinal, y para resaltar las anomalías presentes en ellas.

En la secuencia STIR, se utiliza un pulso de inversión corto para suprimir la señal de tejidos con tiempos de relajación T1 cortos, como la grasa, mientras que se preserva la señal de los tejidos con tiempos de relajación T1 largos, como el líquido LCR y las lesiones de la médula espinal. Luego de aplicar el pulso de inversión, se aplica una secuencia de eco de gradiente o eco de espín para obtener las imágenes.

Al suprimir la señal de la grasa, la secuencia STIR permite resaltar las lesiones que afectan a la médula espinal, ya sean tumores, quistes, abscesos, hernias discales o EM, entre otros. Las imágenes resultantes de la secuencia STIR son especialmente útiles para evaluar la extensión y la localización de las lesiones, así como para determinar su actividad y respuesta al tratamiento. En las imágenes STIR podemos observar el LCR brillante y la médula espinal y la grasa, gris oscuro. Las lesiones aparecen como regiones claras en la médula espinal, tal y como podemos ver en la Figura 66.

Anexo I: Detección automática de lesiones en médula cervical



Figura 66. Imagen 2D STIR. El LCR aparece brillante, la grasa negra y la médula espinal gris oscura. Señaladas en amarillo, observamos la presencia de lesiones como hiperintensidades en el nivel C1-C7.

Al ser el LCR tan brillante, las fluctuaciones y pulsaciones del canal espinal pueden causar artefactos importantes como el de Gibb o de truncamiento. El abrupto cambio de intensidad que hay en tan poca cantidad de píxeles como la existente entre el brillo del LCR y la oscuridad de la médula espinal, puede no ser detectado por el equipo de RM debido a su frecuencia espacial de muestreo, dando lugar a los artefactos mencionados anteriormente.

El tiempo de inversión correcto para suprimir un tejido con un cierto tiempo de recuperación T1 se define como:

$$TiempoInversion_{tejidoA} = \ln 2 * T1tiemporecuperacion_{tejidoA} \quad (31)$$

La secuencia STIR presenta una gran relevancia en la investigación y la práctica clínica relacionada con las lesiones de la médula espinal. Su sensibilidad para detectar lesiones en la médula espinal, incluyendo lesiones traumáticas, infecciosas, inflamatorias y tumorales, permite una detección temprana y precisa de las anomalías, lo que es crucial para el diagnóstico y el tratamiento adecuados. Además de en la detección de lesiones, la secuencia STIR también puede proporcionar información sobre sus características, ayudando a diferenciar entre lesiones benignas y malignas, evaluar su extensión y gravedad y proporcionar información relevante para la planificación del tratamiento. Esta detección y caracterización, permite y facilita el seguimiento de la progresión de las lesiones en la médula espinal a lo largo del tiempo, de manera que se pueda evaluar la respuesta al tratamiento y determinar si hay cambios en el

Anexo I: Detección automática de lesiones en médula cervical

tamaño, forma o intensidad de la señal de las lesiones. Esto es esencial para monitorear la progresión de la enfermedad y ajustar las estrategias de tratamiento según sea necesario. La secuencia STIR desempeña un papel importante en la investigación académica relacionada con la médula espinal, ya que, como hemos mencionado anteriormente, permite estudiar las características de las lesiones, la patofisiología subyacente y las respuestas terapéuticas. Los estudios que utilizan la secuencia STIR pueden contribuir al conocimiento científico en el campo de la neurociencia y proporcionar información valiosa para el desarrollo de nuevas terapias y enfoques de tratamiento.

Esta secuencia es, por lo tanto, esencial en la detección, caracterización y seguimiento de las lesiones en la médula espinal, y su importancia radica en su capacidad para mejorar el diagnóstico, la comprensión de la patología y la contribución a la investigación. Es importante destacar que la secuencia STIR es solo una de las muchas secuencias de RM utilizadas en la evaluación de la médula espinal. Dependiendo del caso clínico y de la sospecha diagnóstica, el clínico puede solicitar la adquisición de otras secuencias, como T1, T2, FLAIR o difusión, entre otras, para obtener una evaluación más completa y precisa de la médula espinal.

Técnica *Inversion-recovery*

La técnica de recuperación de inversión (*Inversion-recovery*) es una técnica utilizada en RM para mejorar la visualización de ciertos tejidos o características en una imagen. Se basa en la manipulación de la magnetización de los tejidos mediante la aplicación de un pulso de RF específico.

En la RM, los tejidos se excitan mediante la aplicación de un pulso de RF que perturba el equilibrio de los momentos magnéticos de los átomos en el cuerpo. Después del pulso de excitación, los momentos magnéticos se recuperan gradualmente a su estado de equilibrio, lo que genera una señal detectable que se utiliza para generar la imagen.

En la técnica de recuperación de inversión, se introduce un pulso de inversión antes del pulso de excitación. El pulso de inversión invierte temporalmente la dirección de los momentos magnéticos de los tejidos, de manera que los momentos magnéticos apuntan en dirección opuesta a la del campo magnético principal. Luego, se aplica el pulso de excitación.

La inversión de los momentos magnéticos tiene el efecto de oscurecer la señal de ciertos tejidos en la imagen final. Sin embargo, a medida que los momentos magnéticos se recuperan de la inversión, estos tejidos comienzan a producir una señal más fuerte en comparación con los

Anexo I: Detección automática de lesiones en médula cervical

tejidos que no fueron invertidos. Esto permite resaltar la presencia de ciertos tejidos o características en la imagen.

En general, el pulso de una secuencia IR puede ser dividido en dos partes; un pulso de inversión de RF y un módulo de secuencia de pulso autónomo, al cual nos referimos como secuencia huésped (Bernstein et al., 2004).

El ciclo de repetición de la secuencia de pulso comienza con el pulso RF. Esto ocasiona que la magnetización longitudinal se invierta, es volteada por un pulso de RF de 180°. Posteriormente, esta se recupera sin desarrollar magnetización transversal, pasando por un punto nulo donde su valor es cero. El pulso de inversión puede ser selectivo para cada corte o invertir un volumen entero. Los pulsos de inversión suelen ir seguidos inmediatamente por gradientes “spoiler” que eliminan cualquier magnetización no deseada, con el fin de evitar interferencias con la adquisición de futuras imágenes y artefactos asociados. En la Figura 67 podemos observar la recuperación de la magnetización.

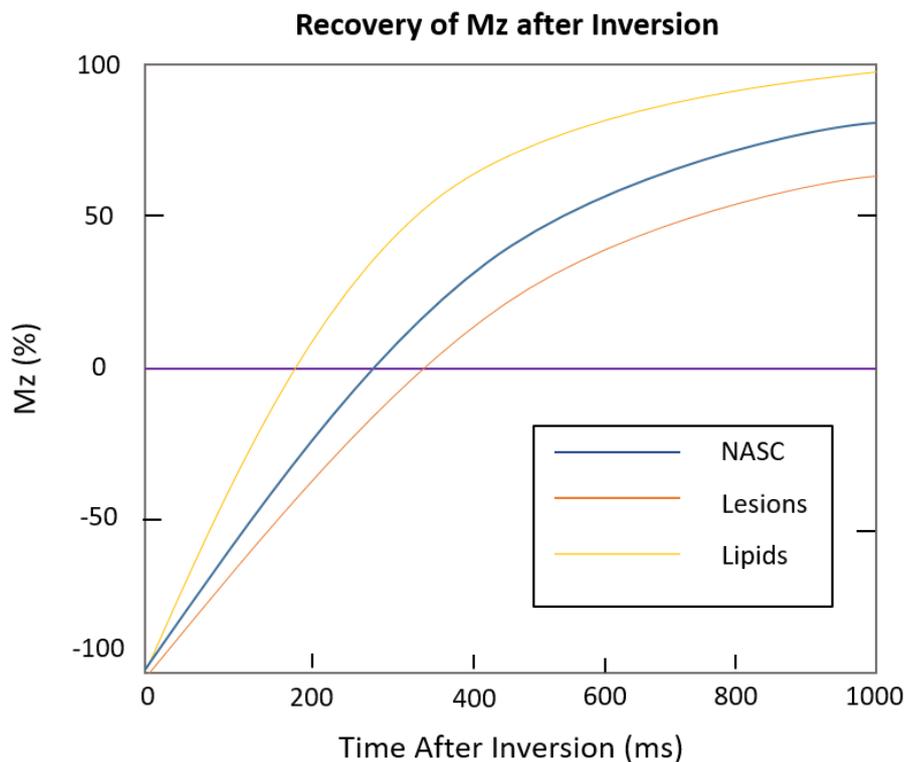


Figura 67. Curvas de recuperación T1 de lípidos, tejido de médula espinal sana (NASC) y tejido correspondiente a lesiones. La línea discontinua indica el punto nulo como tiempo tras la inversión en el que la magnetización longitudinal de los lípidos es cero.

Tras un periodo de tiempo específico, se aplica un pulso de excitación de 90°. La cantidad de tiempo entre la inversión y la excitación define el parámetro de imagen conocido como tiempo

de inversión. Este parámetro es de crucial importancia con respecto a la ponderación de la imagen. Se pueden encadenar dos o tres pulsos de inversión para suprimir la señal de dos o tres tipos de tejido. La inversión puede ser también espectralmente selectiva, de manera que, por ejemplo, sólo los lípidos sean invertidos.

La excitación marca el inicio de la secuencia huésped. Cualquiera que sea la cantidad de magnetización disponible en este momento, será volteada en el plano transversal donde puede ser medida para obtener imágenes. Esto será así independientemente de si la magnetización estaba en la dirección z positiva o negativa en la excitación. La secuencia huésped puede ser de varios tipos de secuencias, incluyendo secuencias FSE y de eco-gradiente como la secuencia MPRAGE. Hay muchas opciones, ya que la recuperación de inversión se puede utilizar junto a otras muchas técnicas de imagen de RM.

4 Registro de secuencias de imágenes médicas

El alineamiento o registro de imágenes es una técnica muy importante en el campo de las imágenes médicas. En muchas situaciones, para estudiar el diagnóstico de un paciente, es necesario analizar varios tipos de imágenes las cuales pueden o no haber sido adquiridas por diferentes tipos de dispositivos como: escáneres de rayos X, resonancia magnética (MRI), tomografía computarizada (CT), ultrasonido, etc. Cada uno de estos sensores puede dar diferente información del paciente, por lo que mediante el registro es posible combinar en una misma imagen (fusionada) las características propias proveniente de varios sensores (multimodal) o de un mismo sensor (monomodal), donde pueden variar puntos de vista o lapsos de tiempo diferentes. El registro de imágenes permite combinar la información de estas y poder visualizarlas en una única imagen con el objetivo de facilitar el examen médico.

Registrar un par o varias imágenes consiste en “alinearlas” de tal manera que cada una de las imágenes quede orientada de la misma forma que la imagen de referencia, con esto se puede combinar la información de ambas en un mismo espacio y en una misma representación visual. Lo que se pretende es transformar las imágenes a un sistema de coordenadas común de manera que los píxeles representen puntos biológicos homólogos.

Durante los últimos años ha habido un creciente interés en relacionar información entre diferentes imágenes médicas, interés acentuado también por la cada vez mayor disponibilidad de escáneres, modalidades y poder de cálculo.

Muchas han sido las investigaciones enfocadas en desarrollar algoritmos que se dediquen a ejecutar este proceso de registro. Sin embargo, se han dedicado menos esfuerzos en vista a la accesibilidad, interoperabilidad y extensibilidad de estos algoritmos, ya que los códigos fuente

normalmente no son publicados, no son fáciles de reproducir o les falta documentación. Esto es un problema, ya que el registro de imágenes es un prerrequisito para una gran amplitud de tareas de análisis de imagen médica y un componente algorítmico clave para los estudios basados en imagen.

Para la tarea que nos concierne, empleamos el software de código abierto SimpleElastix, el cual, como una librería de Python, nos permite realizar el registro de imágenes médicas, configurando, testeando y comparando distintos métodos mediante las líneas de código.

4.3 Antecedentes matemáticos

El concepto de registro está representado de manera simplificada en la Figura 68. Un procedimiento estándar de registro de imágenes implica dos imágenes de entrada: Una es conocida como la imagen fija $I_F(x)$ y otra como la imagen móvil $I_M(x)$. x indica la posición d -dimensional en la imagen. En nuestro caso, al estar registrando la zona cervical escaneada por una máquina de RM, tendremos $d = 3$ y x tendrá tres componentes, uno para cada dimensión espacial. También se pueden registrar series temporales, las cuales tienen cuatro dimensiones (3D+tiempo).

Una transformada $T(x)$ representa el mapeo espacial de puntos desde la imagen fija p hasta convertirlos en puntos en la imagen móvil q . Esto establece una correspondencia para cada píxel de la imagen fija con una posición en la de movimiento.

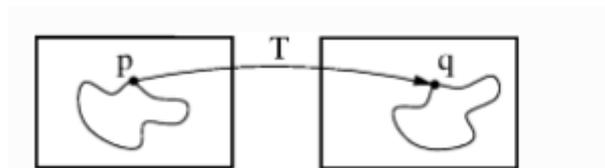


Figura 68. El registro de imágenes es el proceso de deformar los puntos de la imagen móvil, p , a los puntos biológicamente correspondientes, q , en el dominio de la imagen fija.

Una métrica de similitud nos da una medida de cómo de buena es la coincidencia entre la imagen fija y la móvil. Esta medida crea un criterio cuantitativo para ser optimizado por un optimizador sobre el espacio de búsqueda definido por los parámetros de la transformada. Es decir, el registro se formula como un problema de optimización en el que una función de coste C se minimiza con respecto a T . Matemáticamente, $I_M(x)$ es deformada para coincidir con $I_F(x)$ encontrando una transformación de coordenadas $T(x)$ que haga que $I_M(T(x))$ este espacialmente alineada con $I_F(x)$. Esto implica que el optimizador ajusta los parámetros de la transformada de manera que minimiza la diferencia entre las dos imágenes.

Anexo I: Detección automática de lesiones en médula cervical

La métrica es un componente clave en este proceso, usa información de la imagen fija y de la móvil para calcular un valor de similitud. La derivada de este valor nos dice en qué dirección deberíamos mover la imagen móvil para conseguir un mejor alineamiento. La imagen móvil se mueve a pequeños pasos y este proceso se repite hasta que cumplimos el criterio de convergencia. La métrica puede utilizar intensidades de píxeles, posiciones de puntos, características de imagen pre-calculadas o lo que queramos optimizar.

En SimpleElastix hay disponibles diferentes transformadas, métricas y optimizadores. La elección del método correcto para un problema en particular suele ser un proceso de prueba-error incluso para los más expertos.

4.4 Componentes de registro

A continuación, introduciremos la terminología común y algunas de las opciones para distintos tipos de componentes de registro.

4.4.1 Image Pyramids

Una estrategia piramidal multi-resolución mejora el rango de captura y robustez del registro. Hay tres tipos de *pyramids* disponibles en SimpleElastix: *SmoothingImagePyramid*, *RecursiveImagePyramid* y *ShrinkingImagePyramid*. *SmoothingImagePyramid* suaviza las imágenes con un filtro Gaussiano a diferentes escalas. *RecursiveImagePyramid* suaviza y submuestra la imagen. *ShrinkingImagePyramid* simplemente submuestra la imagen.

Para definir la estrategia de multi-resolución hay que definir dos parámetros: *NumberOfResolutions* e *ImagePyramidSchedule*. Este último define la cantidad de desenfoque y/o submuestra en cada dirección x,y,z . Si sólo definimos *NumberOfResolutions*, se usará un *ImagePyramidSchedule* por defecto que suavizará la imagen fija por un factor de 2 en cada dimensión.

4.4.2 Máscaras

Cuando lo que nos interesa es alinear subestructuras en vez de la anatomía global, el problema se complica. Por ejemplo, las imágenes TC en 4D de los pulmones varían considerablemente por el movimiento respiratorio, por lo que no nos interesaría alinear las cajas torácicas, que son más estáticas, ya que los pulmones se solaparían. Una opción sería recortar la imagen, pero esto nos limita la región de interés (ROI) a sólo un cuadrado (2D) o un cubo (3D). Si lo que necesitamos es una ROI de forma irregular, podemos utilizar máscaras. Una máscara, como ya sabemos, es una imagen binaria donde los valores de intensidad de interés se señalan con 1's, dejando el resto a 0's.

Anexo I: Detección automática de lesiones en médula cervical

Para el registro de imágenes, se recomienda utilizar una máscara cuando:

- La imagen contiene un borde artificial que no tiene importancia real para nosotros, ya que el registro podría estar tentado a alinear esos bordes, pasando por alto los que nos interesan de verdad. Un ejemplo es el borde del haz cónico en las imágenes de ultrasonidos.
- La imagen contiene estructuras cercanas a la ROI que pueden alterar el registro de las estructuras de dentro de la ROI, como en el ejemplo de los pulmones y la caja torácica.

Las máscaras se pueden utilizar como imágenes fijas o móviles. Una máscara de imagen fija es suficiente para centrar el registro en una ROI, ya que las posiciones de muestra se extraen de la imagen fija. Normalmente, sólo se utiliza una máscara como imagen móvil cuando la imagen móvil tenga valores de gris molestos cerca de la ROI.

4.4.3 Transformadas

La elección de la transformada es esencial para que el registro sea satisfactorio, ya que, refleja el tipo de transformación que queremos y restringe el espacio de solución.

El número de parámetros de la transformada corresponde a los grados de libertad (DOF, degrees of freedom) de la transformación. Este número suele ser de 3 DOFs para translación, 12 para una deformación affine 3D y entre cientos y miles DOFs para campos de deformación b-spline y métodos no paramétricos.

El número de DOFs es directamente proporcional a la complejidad de cálculo del problema de optimización, la cual afecta al tiempo de ejecución y a la probabilidad de convergencia a una solución óptima y un buen resultado de registro. Se recomienda como buena práctica empezar con transformadas simples e ir evolucionando hacia transformadas más complejas si fuese necesario.

Algunas transformadas comunes son (en orden de complejidad en aumento) traslación, rígida (rotación, traslación), Euler (rotación, traslación), affine (rotación, traslación, escalado, corte), b-spline (no rígida), Spline-kernel (no rígida) y combinaciones ponderadas de cualquiera de ellas.

En SimpleElastix, la transformada se define desde la imagen fija a la imagen móvil. Resulta contradictorio que la transformada se defina en esta dirección ya que es la imagen móvil la que queremos transformar. Sería más lógico mapear cada píxel de la imagen móvil a su nueva posición en la imagen fija, pero de ser así, dos píxeles de la imagen móvil podrían ser mapeados a la misma posición de la nueva cuadrícula fija y algunos píxeles de la imagen fija podrían

Anexo I: Detección automática de lesiones en médula cervical

directamente no ser mapeados. El convenio elegido nos permite iterar sobre la imagen fija y escoger un píxel de la imagen móvil para cada uno de sus píxeles.

4.4.4 Métricas

La métrica de similitud de semejanza mide el grado de similitud entre la imagen móvil y la fija. La métrica toma valores de intensidad de la imagen fija y de la que ha sido transformada y evalúa si el valor es apto.

La selección de una métrica adecuada depende mucho del problema de registro a ser resuelto. Algunas métricas son sólo válidas para comparar imágenes de la misma modalidad, otras pueden manejar comparaciones entre modalidades. No hay unas reglas claras a la hora de cómo elegir la métrica y normalmente requiere de varias pruebas hasta encontrar la que mejor nos va a nuestro problema.

Algunas métricas son:

- Mean Squared Difference (SSD), calcula las diferencias de intensidad cuadrática media por píxel entre la imagen fija y la móvil. El valor óptimo de esta métrica es cero. Esta métrica se basa en la suposición de que la intensidad que está representando los mismos puntos homólogos debe de ser la misma en ambas imágenes y ser sólo válida para dos imágenes con las mismas distribuciones de intensidad, como por ejemplo en imágenes de la misma modalidad.
- Normalized Correlation Coefficient (NCC) calcula la correlación cruzada de píxeles normalizada por la raíz cuadrada de la autocorrelación de las imágenes. Esta métrica es invariante a diferencias lineales entre distribuciones de intensidad y es adecuada especialmente para registro de TC intra-modales donde las escalas de intensidad están siempre relacionadas por una transformada linear incluso entre escáneres.
- Mutual Information (MI) mide la información mutua entre dos imágenes y es más genérica. La MI es una medida de cuanta información nos da una variable aleatoria (la intensidad de imagen) de otra variable aleatoria (la intensidad de otra imagen distinta). Esto corresponde a medir la dependencia de las distribuciones de densidad de probabilidad (PDF) de las intensidades de las imágenes fijas y móviles sin tener que especificar la forma actual de la dependencia. Esta métrica es adecuada para pares de imágenes multi y mono modales.

Anexo I: Detección automática de lesiones en médula cervical

- Kappa similarity metric, mide la superposición de estructuras segmentadas y está desarrollada específicamente para registrar imágenes binarias (segmentaciones). Aun así, en algunos casos, es mejor convertir las imágenes binarias a un mapa de distancia y aplicar una de las métricas anteriormente descritas.

4.4.5 Optimizadores

El optimizador estima los parámetros óptimos de la transformada de manera iterativa. Hay muchos optimizadores para emplear como Gradient Descent (GD), Robbins-Monroe (RM), Adaptive Stochastic Gradient Descent (ASGD), Conjugate Gradient (CG), Conjugate Gradient FRPR, Quasi-Newton LBFGS, RSGD “Cada parámetro aparte”, Simultaneous Perturbation (SP), CMAEvolutionStrategy y FullSearch que muestrea el espacio de búsqueda entero. Se recomienda, en caso de no tener ninguna preferencia, emplear Adaptive Stochastic Gradient Descent que requiere menos parámetros para ajustar y suele ser más robusto.

4.4.6 Muestreadores

El muestreador es el responsable de elegir las localizaciones en la imagen de entrada en las que la métrica será evaluada. En general, es suficiente con evaluar sólo unos cuantos voxels elegidos de manera aleatoria. SimpleElastix tiene diferentes estrategias como grid, aleatoria, coordenada aleatoria y muestreo completo, que se seleccionan utilizando el parámetro *ImageSampler*.

El muestreador de “grid” define una cuadrícula regular en la imagen fija y selecciona las coordenadas en ella. La imagen es submuestreada sin suavizar. El tamaño de la cuadrícula o factor de submuestreo es un parámetro de entrada. El muestreador aleatorio selecciona, como su nombre indica, de manera aleatoria, un número de vóxeles especificado por nosotros. Cada vóxel tiene la misma posibilidad de ser seleccionado y también puede ser seleccionado más de una vez. El muestreador de coordenadas aleatorias no está limitado a las posiciones de vóxel por lo que puede muestrear posiciones entre ellos. En ese caso, estos valores de gris se calcularían mediante interpolación.

4.4.7 Interpoladores

La métrica suele comparar valores de intensidad entre la imagen fija y los valores correspondientes en la imagen móvil una vez transformada. Cuando un punto se mapea de un espacio a otro por una transformada, puede ser mapeado a una posición que no pertenezca a

Anexo I: Detección automática de lesiones en médula cervical

una cuadrícula. Es aquí cuando la interpolación es necesaria para calcular los valores de intensidad de estas posiciones.

Existen varios métodos de interpolación que varían en calidad y velocidad.

- El `NearestNeighborInterpolator` devuelve el valor del vóxel más cercano espacialmente. Esta técnica no es de las de mayor calidad, pero también requiere pocos recursos. Este método se suele utilizar más en imágenes binarias.
- El `LinearInterpolator` devuelve una media ponderada de los vóxeles de alrededor utilizando las distancias como pesos. En `SimpleElastix` este método está optimizado y resulta rápido. En general, se puede utilizar junto con el muestreador de coordenadas aleatorias durante el proceso de optimización para un buen rendimiento en términos de tiempo.
- El `BSplineInterpolator` (o la versión más eficiente en memoria `BSplineInterpolatorFloat`) interpola valores de píxel usando aproximaciones b-spline del orden N definido por el usuario.

5 Metodología propuesta

A continuación, realizamos una breve explicación de las diferentes fases que componen la metodología a desarrollar para el objetivo de detección de lesiones en médula espinal. En la Figura 69 podemos ver un diagrama del procedimiento definido. Como en la mayoría de las labores de investigación, durante la elaboración de las distintas etapas, encontramos distintas problemáticas y disyuntivas con las que en un primer diseño de la planificación no se contaba, por lo que esto es sólo una primera aproximación.

Anexo I: Detección automática de lesiones en médula cervical

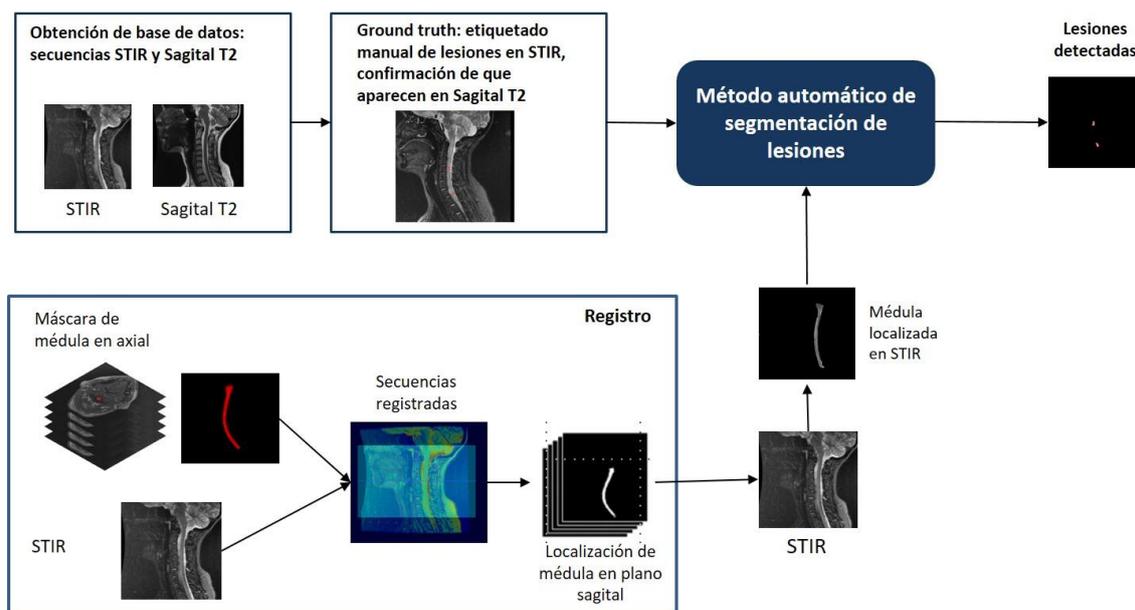


Figura 69. Diagrama de tareas para la segmentación automática de lesiones en médula cervical

5.1 Base de datos

Para la elaboración del algoritmo, se han descargado del PACS del hospital las secuencias STIR y Sagital T2 de los mismos 121 pacientes de EM que forman parte del conjunto de datos del proyecto de segmentación automática de médula espinal, tal como se recoge en la sección 3.1. Estas secuencias también están anonimizadas y su adquisición tuvo lugar en el mismo momento que las secuencias axiales 3D T1 empleadas en el modelo de segmentación automática de médula cervical.

5.2 Registro de imágenes

En el método de detección de lesiones propuesto, nos encontramos con una necesidad fundamental: realizar el registro entre la secuencia STIR y la axial 3D T1 empleada para la fase de segmentación automática de la médula. Esto lo realizamos con el fin de poder interconectar ambos objetivos del proyecto y ser capaces de utilizar la información de localización de la médula cervical obtenida en la segmentación y trasladar las posiciones que tenemos en el plano transversal al plano sagital en el que se encuentra la secuencia STIR, donde, como ya hemos mencionado, examinaremos las posibles lesiones.

En términos más genéricos, este proceso de registro es necesario como un paso de pre-procesado para alinear las imágenes médicas a un sistema de coordenadas común antes de analizarlas, en casos en los que queremos comparar escaneos de varios pacientes o del mismo, pero tomados en distintas sesiones, condiciones o modalidades de imagen.

Anexo I: Detección automática de lesiones en médula cervical

Este registro tiene como objetivo maximizar la coherencia espacial, tal como explicamos en el apartado 4 de este anexo, pues la secuencia STIR, no recoge la misma información que la axial 3D T1, ni en términos espaciales ni en términos de intensidad de imagen. Vamos a explicar mejor este aspecto de un modo visual observando la Figura 70 y la Figura 71.

Cuando observamos las reconstrucciones 3D de ambas secuencias, vemos que no recogen la misma información espacial anatómica del paciente. Mientras la secuencia axial 3D T1 recoge la sección cervical incluyendo todo el cuerpo del paciente, la secuencia STIR recoge información de una pequeña sección centrada en el canal medular del sujeto, focalizando su atención en esta estructura. Por lo tanto, como resulta intuitivo al observar las imágenes, la secuencia axial nos proporciona más información en el plano transversal (útil para el objetivo de segmentación automática de la médula espinal) y la secuencia STIR en el plano sagital (útil para el objetivo de detección automática de lesiones).

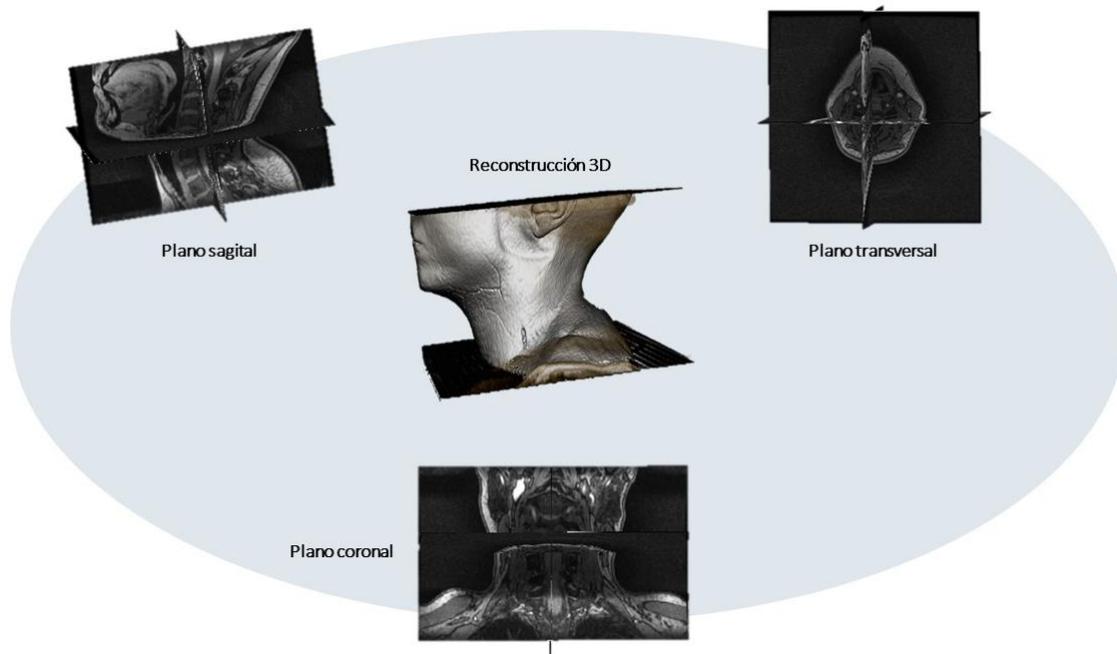


Figura 70. Reconstrucción 3D secuencia axial y planos que la componen

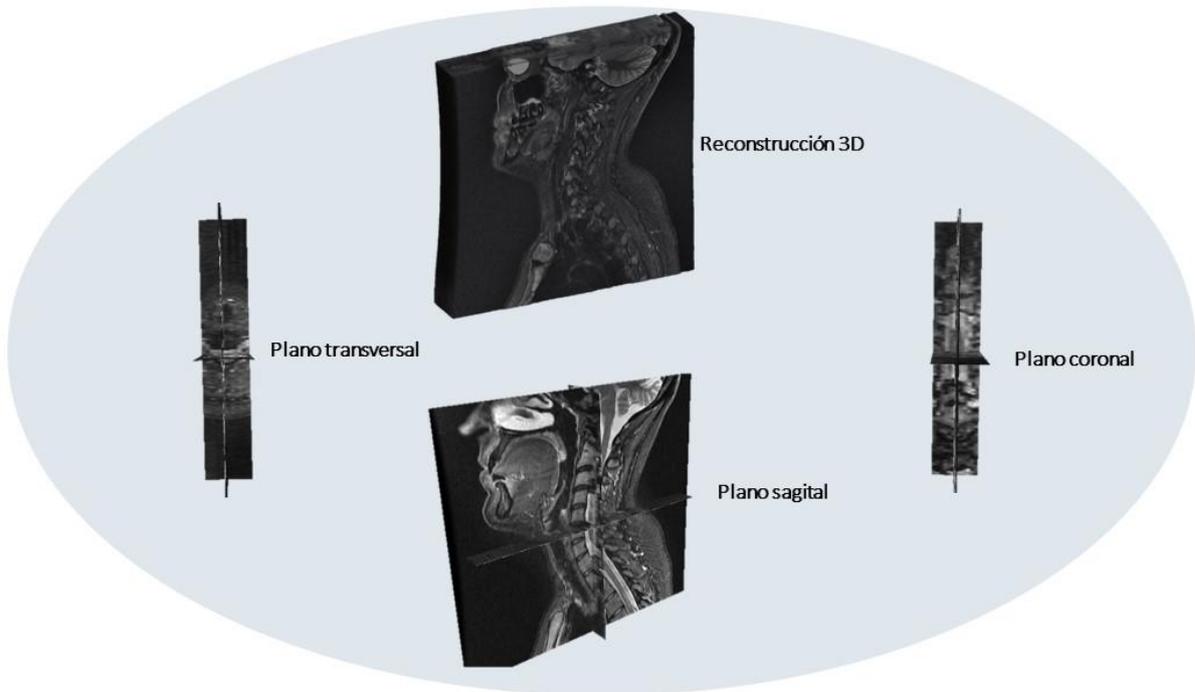


Figura 71. Reconstrucción 3D secuencia STIR y planos que la componen

Como hemos comentado anteriormente, nos centramos en las lesiones que se encuentran en la médula cervical, por lo que, para facilitar la identificación, nos interesa inspeccionar sólo esa zona y evitar posibles confusiones con otras regiones que no son de interés y pueden presentar intensidades semejantes a las de las lesiones. Es por ello por lo que queremos extraer información de localización de la segmentación automática de médula espinal realizada previamente en el plano transversal, de forma que las coordenadas de las posiciones que identificamos como médula las trasladaremos al plano sagital para aprovechar la información espacial.

Esto lo realizaremos mediante registro de imágenes, tal como mencionamos en la sección 4 de este anexo, de manera que transformaremos las secuencias con las que estamos trabajando (axial 3D T1 y STIR) a un sistema de coordenadas común para que los píxeles representen puntos biológicos homólogos. En nuestro caso realizamos el registro entre nuestras imágenes de entrada utilizando la información de la máscara de la médula espinal que obtenemos en el plano transversal tras la segmentación automática, pues lo que nos interesa es alinear subestructuras en vez de la anatomía global. Por lo tanto, y siguiendo la terminología que empleamos en la sección 4 donde hablamos de los fundamentos del registro de imagen, nuestra imagen fija será

Anexo I: Detección automática de lesiones en médula cervical

la secuencia STIR y nuestra imagen móvil la máscara de la médula espinal obtenida de la secuencia axial 3D mediante nuestro algoritmo de segmentación automática.

5.2.8 Ejemplo de aplicación

Algunas pruebas de registro de imágenes han sido ya ejecutadas mediante la biblioteca de Python SimpleElastix.

Recordamos que, la necesidad principal de este registro es poder localizar en la secuencia STIR la información de imagen perteneciente a la médula y ser capaces de identificar las posibles lesiones en ella. Para ello utilizamos la información espacial que ya tenemos en el plano transversal, la cual hemos conseguido mediante el algoritmo de segmentación automática de médula espinal cervical desarrollado en esta tesis.

A continuación, mostramos un ejemplo de registro de imagen entre una secuencia STIR y la máscara de la médula en el plano transversal de nuestra colección de estudios. La Figura 72 muestra los cortes de imagen que componen una secuencia STIR en el plano sagital. Superpuesto en un mapa de color más claro, observamos la información de localización espacial de la médula espinal cervical que obtenemos al hacer el registro entre la secuencia axial 3D y la STIR. Esta información no es del todo precisa, ya que, por ejemplo, en la Figura 72c o Figura 72d vemos como, tras el registro, se identifican como píxeles pertenecientes a la médula regiones que ya no pertenecen a ella y son parte del LCR o estructuras más distales.

Anexo I: Detección automática de lesiones en médula cervical

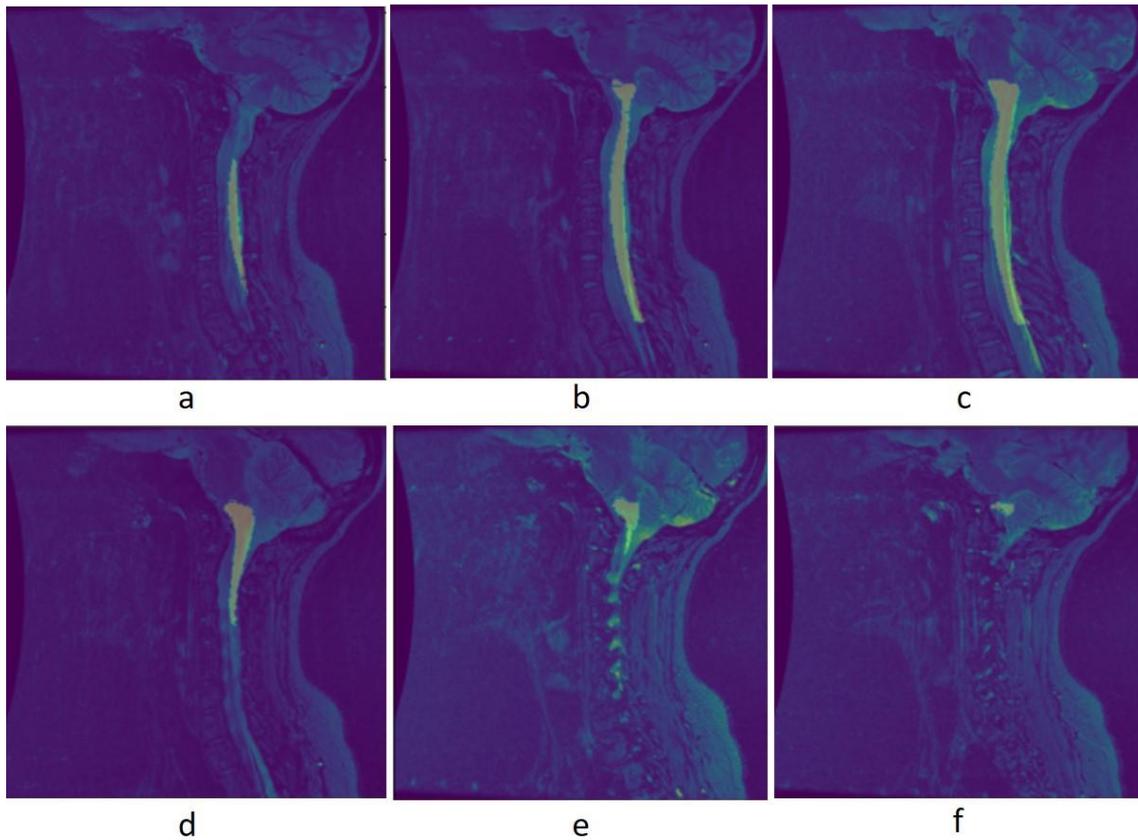


Figura 72. Cortes pertenecientes a secuencia STIR con información superpuesta de localización de médula cervical procedente de secuencia axial 3D.

En la Figura 73, observamos otro ejemplo de registro. En la columna de la izquierda vemos la máscara de médula espinal cervical que obtenemos en el plano sagital procedente de la información espacial en axial. En la columna de la derecha observamos las imágenes STIR donde se quiere aplicar la máscara para focalizar la búsqueda de lesiones en esa región.

Anexo I: Detección automática de lesiones en médula cervical

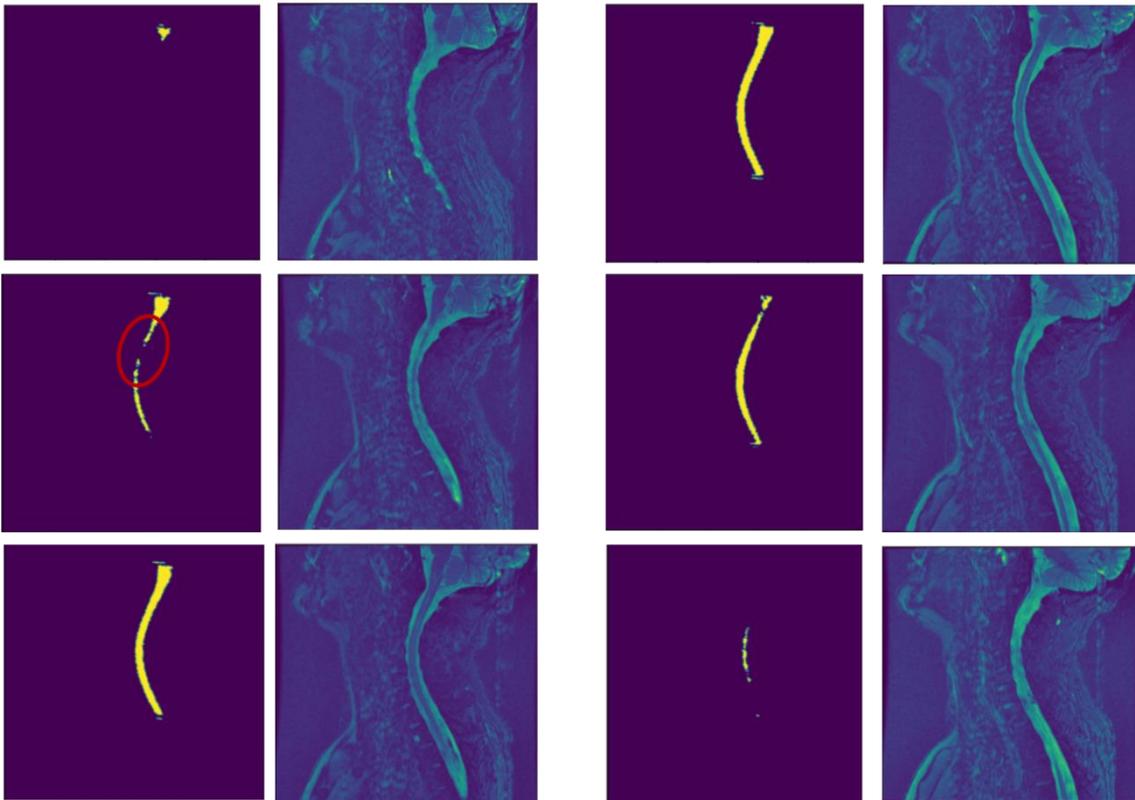


Figura 73. Ejemplo de máscara de médula espinal generada en sagital (columna izquierda) tras el registro entre máscara en transversal y secuencia STIR (columna derecha).

Como podemos observar también en este último ejemplo, este registro no genera máscaras ideales, pues en rojo vemos rodeada una región en la que se generan espacios intermedios que no corresponden con la estructura real de la médula.

Es aquí donde comenzamos a encontrarnos con dificultades derivadas de la no “idoneidad” y de la complejidad del campo del registro de imágenes médicas, pues al ser muy heterogéneo, encontrar un método automático que nos proporcione una misma angulación entre secuencias STIR y axial 3D, en este caso, es una complicada tarea.

Una de las pruebas que se realizó con el fin de encontrar una solución a la problemática indicada en el registro que acabamos de reflejar en la Figura 73, es aplicar técnicas de imagen morfológicas como la dilatación 3D, añadiendo así píxeles a los límites de los objetos de una imagen. El resultado principal de la dilatación morfológica es que los objetos sean más visibles y se rellenen los pequeños huecos. Con este paso, solventamos el problema de los espacios en la máscara, pero, por el contrario, la máscara se agranda en general, por lo que ya no corresponderá sólo a la médula cervical, sino que también recogerá información del líquido circundante, lo que nos afecta al posterior análisis de la detección de lesiones, ya que, el LCR

Anexo I: Detección automática de lesiones en médula cervical

presenta niveles de intensidad semejantes a las hiperintensidades que tienen las posibles lesiones, tal como mostramos en la Figura 74.

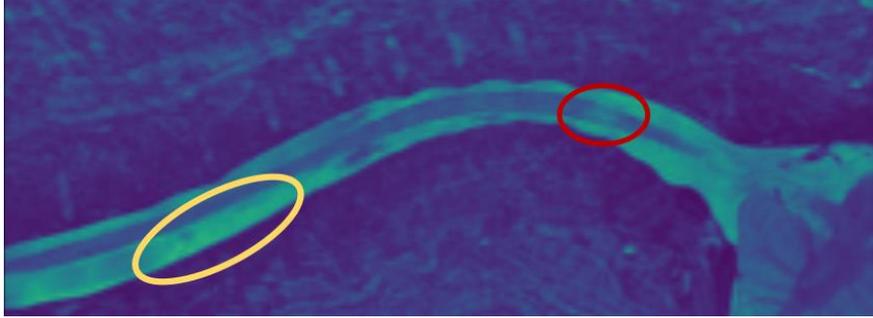


Figura 74. Similitud de intensidad entre LCR (región en amarillo) y lesión (región en rojo)

Para lograr que la máscara medular tenga una forma más estrecha, la cual se adapte sólo a la región en la que nos interesa buscar lesiones, aplicamos una erosión morfológica para eliminar los píxeles pertenecientes al LCR. Este proceso de dilatación, seguida de una operación de erosión, se conoce como cierre morfológico y es útil para casos como el nuestro, en el que interesa rellenar huecos pequeños en una imagen a la vez que se conservan la forma y el tamaño de otros huecos y objetos grandes. En la Figura 75 podemos observar, en el conjunto de columnas de la izquierda, la máscara tras la dilatación y su imagen STIR correspondiente, y en las de la derecha, las mismas máscaras y su correspondiente imagen en STIR pero tras las operaciones de dilatación y erosión (cierre).

Anexo I: Detección automática de lesiones en médula cervical

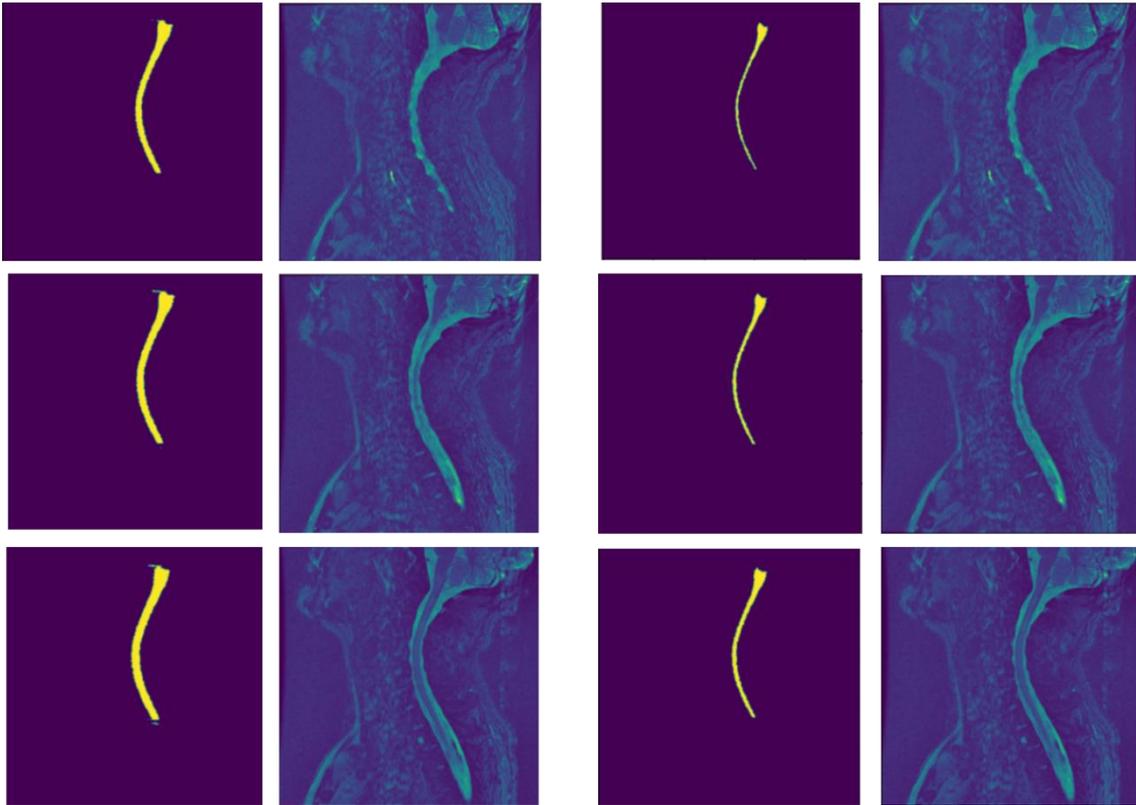


Figura 75. A la izquierda, máscaras generadas por el registro tras aplicar dilatación, a la derecha, mismas máscaras tras añadir también erosión.

Esta máscara que obtenemos tras el cierre morfológico, es la que utilizamos para obtener la segmentación de médula cervical en el plano sagital donde nos interesa detectar las lesiones enmascarando la secuencia STIR, como mostramos en la Figura 76.

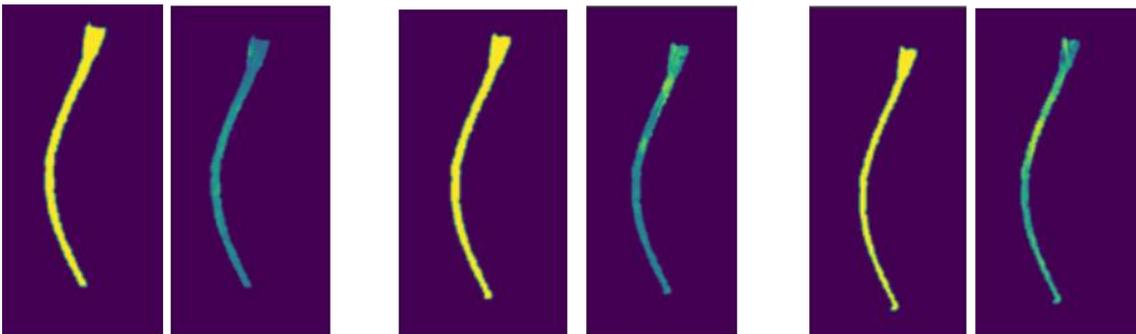


Figura 76. A la izquierda, máscara final tras operación de cierre, a la derecha, imagen de intensidad resultante tras aplicar la máscara a la secuencia STIR.

Como hemos indicado anteriormente, esta técnica necesita de un mayor estudio y mejora, planteando como objetivo en nuestras líneas futuras el desarrollo de un método de registro más generalizable a todos los casos intentando aplicar el menor pos procesado individualizado.

Anexo I: Detección automática de lesiones en médula cervical

Estos ejemplos son reflejados con el mero objetivo de explicar de una manera más práctica y visual lo que se pretende conseguir y qué técnicas se han probado hasta ahora para ello.

5.3 Detección de lesiones

5.3.9 Introducción

Al inicio del presente anexo, en el apartado 2, recogemos la importancia de detectar lesiones en médula espinal cervical, pues la detección de éstas ayudará a confirmar el diagnóstico de la enfermedad y a diferenciarla de otras afecciones que pueden presentar síntomas similares. Su tamaño y ubicación en la médula cervical pueden proporcionar información sobre la gravedad de la enfermedad y su impacto en la función neurológica del paciente, lo que es de utilidad a los clínicos para determinar el pronóstico y planificar el tratamiento más adecuado. Para ello, en la rutina hospitalaria, los especialistas emplean imágenes de RM convencionales, como las descritas en la sección 3 (STIR o sagital potenciada T2) para cuantificar de manera no invasiva la progresión tanto en tiempo como en espacio de la carga lesional (Filippi & Rocca, 2007). El estudio de estas lesiones ha ganado una merecida importancia dado su potencial valor para el diagnóstico y pronóstico de la enfermedad (Arrambide et al., 2017; Sombekke et al., 2013). Debido a ello, la segmentación de la médula espinal y las lesiones de EM que contiene (lesiones intramedulares) se ha convertido en un procedimiento común para evaluar cuantitativamente la integridad estructural de esta parte del SNC en pacientes de EM. Por consiguiente, y junto al hecho de que, como tarea manual, la segmentación de la médula espinal y las lesiones intramedulares resulta laboriosa, lenta de ejecutar y llena de variabilidad entre evaluadores, se torna clara la necesidad de una herramienta de segmentación automática y robusta que se encargue de esta labor.

Varios son los métodos de segmentación automática que se han propuesto en los últimos años, los primeros incluyendo enfoques con técnicas de contornos activos, basados en superficies (De Leener et al., 2015; Koh et al., 2010) y métodos basados en atlas (Carbonell-Caballero et al., 2006; M. Chen et al., 2013; Pezold et al., 2015; Tang et al., 2013). Aunque algunos de estos métodos han demostrado tener un buen funcionamiento (De Leener et al., 2016), normalmente requieren de interacción humana, como por ejemplo al tener que indicar manualmente una región de interés, o están limitados a un contraste y resolución específicos. La segmentación automática de la médula espinal o de las lesiones intramedulares, es una tarea difícil de implementar dada la gran variedad de formas, longitudes y patologías presentes en la médula que, a su vez, podemos encontrar recogidas en imágenes de distintas dimensiones, resoluciones, orientaciones, contrastes y artefactos.

Anexo I: Detección automática de lesiones en médula cervical

Dentro del estado del arte de la segmentación de lesiones en la enfermedad de EM, en las dos últimas décadas se han desarrollado algunos métodos para la detección en conjuntos de datos de cerebro (García-Lorenzo et al., 2013; Lladó et al., 2012), aunque todavía sigue siendo una tarea con retos en el horizonte y aspectos que optimizar (Meier et al., 2018; Roy et al., 2018; Valverde, Cabezas, et al., 2017; Valverde, Oliver, et al., 2017). Aunque estos métodos, funcionan en cerebro, no son transferibles a médula espinal debido principalmente a su particular morfología. Además, los métodos tradicionales basados en intensidad mencionados anteriormente, son complicados de aplicar a médula espinal debido a factores como el ruido *bias-field* (bastante frecuente y difícil de corregir), la confusión entre las intensidades de las lesiones con las de otras estructuras, los artefactos y los efectos de volumen parcial donde varias estructuras pueden contribuir a la señal de los vóxeles pertenecientes a los bordes (como el líquido cerebroespinal y la médula).

Los métodos basados en CNN, han demostrado ser muy robustos a la existencia de variabilidad entre imágenes. En particular, y tal como introducimos en la sección 1.3.3.1.4 de esta tesis, la arquitectura U-net marcó en 2015 un antes y un después en la comunidad de la segmentación de imagen médicas. El buen funcionamiento de la arquitectura U-net proviene de su diseño con dos rutas diferenciadas: una de compresión que captura las características de la imagen a diferentes niveles de abstracción y otra simétrica de expansión para recuperar la información espacial con la ayuda de las conexiones de salto entre ambas rutas. Sin embargo, aunque la arquitectura U-net muestra buenos resultados en una gran variedad de tareas en imagen médica, utilizarla en conjuntos de datos muy desbalanceados, como los que nos encontramos en tareas de segmentación de médula espinal o lesiones intramedulares (pocos píxeles pertenecientes a la ROI en comparación al resto de la imagen) genera dificultades de implementación (Buda et al., 2017; Sudre et al., 2017).

Además, la calidad de las imágenes de RM que recogen la médula espinal están especialmente afectadas por artefactos procedentes del movimiento del paciente o de ciclos respiratorios o cardíacos (Kearney et al., 2015) que pueden introducir incertidumbre sobre si las hiperintensidades son verdaderamente lesiones o simplemente artefactos. Incluso para los especialistas clínicos, después de determinar la existencia de una lesión, es complicado delinear los bordes de ella si están presentes estos artefactos y efectos de volumen parcial, y todavía lo es más en lesiones difusas, aunque no exista la presencia de dichos artefactos. La variabilidad entre evaluadores en analizar la implicación que tiene la enfermedad en los volúmenes de RM de la médula espinal puede estar provocado por estos factores y puede tener importantes consecuencias clínicas si afecta al tratamiento elegido para el paciente de EM.

Anexo I: Detección automática de lesiones en médula cervical

Con el fin de guiar a radiólogos y neurólogos en esta compleja tarea de detección de lesiones en la médula espinal y reducir la variabilidad entre evaluadores, existen implementados algunos métodos automáticos (Gros et al., 2019). Los métodos basados en DL forman parte del contenido más actual dentro del estado del arte, sin embargo, tienen la desventaja de que estos métodos dependen del aprendizaje adquirido en un conjunto de imágenes donde las lesiones han sido delineadas por expertos, por lo que, la variabilidad entre evaluadores sigue pudiendo afectar a estos modelos de dos formas: primero, los modelos pueden aprender a recrear patrones de segmentación en vez de las segmentaciones “reales”, proporcionándonos malos resultados (Karimi et al., 2019), y segundo, para evaluar el modelo sus predicciones son comparadas con el *ground truth* señalado por expertos, los cuales también pueden padecer de variabilidad y originar una sub o sobrestimación del funcionamiento del modelo.

Se ha demostrado correlación entre los hallazgos en las imágenes RM, en forma de captación de Gadolinio por las lesiones, con la actividad inflamatoria, utilizándose como un biomarcador de progresión dada la buena correlación con los hallazgos neuropatológicos (Arnold & Matthews, 2002; Lee et al., 1999; Lycklama À Nijeholt & Barkhof, 2003; Smith et al., 1993). Esto ha llevado a la aceptación de la RM como un marcador en ensayos clínicos de tratamientos antiinflamatorios para la EM. Las mediciones de la carga lesional y de la actividad inflamatoria de la enfermedad, a través de técnicas cuantitativas de RM se utilizan cada vez con mayor frecuencia, tanto en clínica como en investigación, pues se ha observado una correlación entre los hallazgos en la RM y la actividad inflamatoria, la cual se ha utilizado como un biomarcador de progresión, con una buena correlación a los hallazgos neuropatológicos.

Aunque la segmentación de lesiones de la EM se puede hacer de forma manual corte por corte por un especialista, resulta una labor muy tediosa y que implica una gran dedicación temporal. Además, se pueden producir errores intraevaluador (mismo evaluador), o interevaluador (diferentes evaluadores). Es por ello que, con la propuesta de un método de detección automática de lesiones en médula espinal cervical en EM, intentamos abordar el problema de la gran cantidad de tiempo aplicada por el evaluador y mejorar la reproducibilidad de la medición de los volúmenes de las lesiones de EM.

5.3.10 Ejemplo de aplicación

En el apartado anterior de registro entre secuencias, hemos comentado la necesidad de trasladar la información de localización de la médula cervical del plano de la secuencia axial al de la secuencia STIR.

Anexo I: Detección automática de lesiones en médula cervical

Siguiendo el ejemplo de aplicación anterior, continuamos con la fase de detección de lesiones, pero emplearemos técnicas tradicionales basadas en intensidad. Como mencionamos en el ejemplo de aplicación del registro, esto lo reflejamos solamente con fines informativos e ilustrativos para mostrar lo que se pretende conseguir mediante algoritmos de aprendizaje automático.

Con el fin de eliminar la posible información de intensidad perteneciente al LCR que quedase tras la operación de cierre morfológico, la cual podemos observar en la imagen superior de la Figura 77, calculamos el valor medio de gris de los píxeles pertenecientes a la región señalada por la máscara (*thr*) y lo utilizamos como umbral para quedarnos con aquellos píxeles que están por debajo de cierto nivel de intensidad (médula sin LCR). Empleando este umbral ($1.2 \cdot thr$), volvemos a extraer la imagen perteneciente a la médula cervical con el fin de eliminar aquellas intensidades mayores propicias a ser del LCR, tal como se muestra en la imagen inferior de la Figura 77.

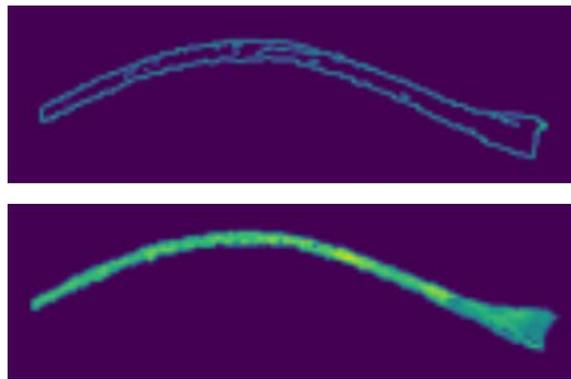
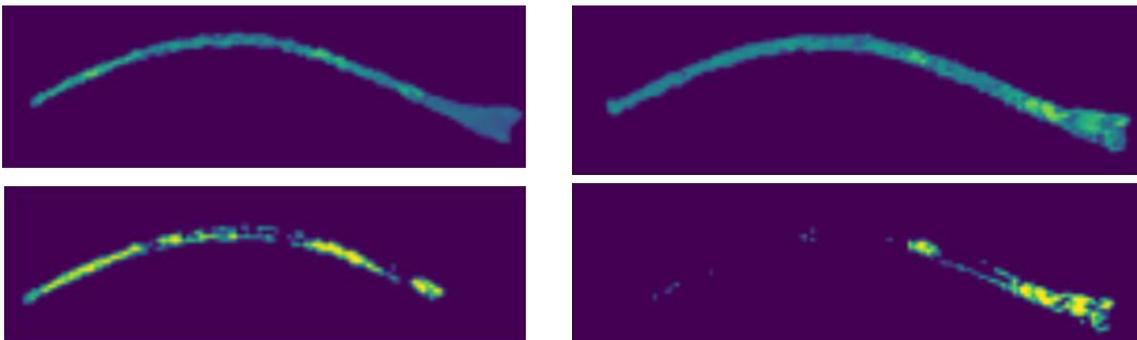


Figura 77. Arriba: detección de píxeles que consideramos divisores entre médula y LCR, abajo: canal medular sin dichos píxeles en el borde.

Así se pretende tener la información de la médula lo más limpia posible, sin restos de LCR u otras estructuras para evitar posibles confusiones. Seguidamente pasamos a detectar las zonas hiperintensas (potenciales lesiones), mediante técnicas de clustering empleadas en tareas de segmentación para agrupar píxeles similares y etiquetarlos como el mismo segmento.



Anexo I: Detección automática de lesiones en médula cervical

Figura 78. Arriba: médulas tras eliminar LCR, abajo: zonas hiperintensas detectadas mediante clustering.

Como podemos observar en la Figura 78, la detección de lesiones es poco precisa pues en la imagen hay valores de gris distribuidos de forma difusa y los conjuntos de píxeles hiperintensos no están agrupados de forma definida. Queda pues demostrada la necesidad de desarrollo de un algoritmo de IA lo suficientemente preciso como para conseguir diferenciar estos píxeles pertenecientes a lesiones intramedulares, una complicada tarea debido a la baja resolución de las imágenes y el alto desbalanceo de etiquetas en esta problemática, donde menos del 1% de los vóxeles de la secuencia STIR o Sagital T2 nos aportan información lesional.

5.4 Correlación entre lesiones y enfermedad

Como objetivo final, se pretende hacer un estudio entre las lesiones detectadas en los 121 pacientes que conforman nuestro estudio y la información clínica sobre sus síntomas y fenotipado. Como ya hemos mencionado, el diagnóstico de la esclerosis múltiple generalmente se basa en la presencia de lesiones típicas en las imágenes de RM, tanto a nivel cerebral como intramedular, combinadas con síntomas clínicos característicos y otras pruebas para descartar otras enfermedades que puedan causar síntomas similares.

Es importante tener en cuenta que la EM es una enfermedad heterogénea, lo que significa que puede manifestarse de diferentes maneras en diferentes personas. Algunos pacientes pueden tener más lesiones en la médula espinal, mientras que otros pueden tener más lesiones en el cerebro o incluso lesiones mínimas en la RM, lo que dificulta el diagnóstico en muchos casos, motivación justificada y más que suficiente para la ejecución de este proyecto y la facilitación al ámbito clínico de toda herramienta de apoyo para el diagnóstico de la enfermedad. Tanto la metodología ya desarrollada para segmentar médula espinal y estudiar la atrofia, como la propuesta para segmentar lesiones en la médula espinal, serán un componente importante para el estudio de la EM y que siempre se deberá utilizar junto con otros hallazgos clínicos para establecer un diagnóstico preciso en pro del bienestar del paciente.

Anexo II: Contribuciones científicas

Anexo II: Contribuciones científicas

A continuación, muestro un breve repaso de las contribuciones científicas aportadas durante esta tesis doctoral, pues estas pueden y se pretende que conduzcan a avances tecnológicos y beneficios para la sociedad, como en nuestro caso, a la ayuda a la atención médica. Es importante el desarrollo de estas actividades durante la tesis, ya que realizar contribuciones científicas nos ayuda a la mejora de habilidades avanzadas de investigación, como el diseño de experimentos, la recopilación y el análisis de datos, la interpretación de resultados y la síntesis de información. Estas habilidades son valiosas no solo para la tesis doctoral, sino también para futuras investigaciones y carreras profesionales. Además, estas publicaciones y presentaciones, recogidas en la Tabla 11, son una forma importante de compartir el trabajo y obtener reconocimiento en la comunidad académica y científica.

En cuanto a resultados en el campo de la divulgación, durante estos años, hemos participado en los congresos internacionales como:

- *European Congress of Radiology (ECR 2020)*
- *18th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC 2020)*
- *8th European Medical and Biological Engineering Conference, Slovenia.*

Con respecto a publicaciones en revista, nuestro paper fue publicado la prestigiosa revista “Journal of Digital Imaging” (JDI) bajo el título “Automated Cervical Spinal Cord Segmentation in Real-World MRI of Multiple Sclerosis Patients by Optimized Hybrid Residual Attention-Aware Convolutional Neural Network” a cuya publicación se puede acceder en el siguiente enlace:

<https://rdcu.be/c50cf>

Tabla 11. Síntesis de actividad divulgativa.

	Fuente	Nombre	Autores	Año	Tipo
<i>European Congress of Radiology, ECR 2020</i>	Congreso	Automatic MR spinal cord segmentation by hybrid residual attentionaware convolutional neural networks and learning rate optimization on real world data	A. Bueno Gómez, A. Alberich-Bayarri, I. Bosch-Roig, J. Carreres-Polo, V. Casanova	2020	Presentación de investigación
<i>18th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC 2020)</i>	Congreso	Automatic MR spinal cord segmentation by hybrid residual attention-aware convolutional neural networks and learning	A. Bueno Gómez, A. Alberich-Bayarri, I. Bosch-Roig, J. Carreres-Polo	2020	Presentación de investigación

Anexo II: Contribuciones científicas

		rate optimization on real world data			
8th European Medical and Biological Engineering Conference, Slovenia	Congreso	Automatic MR Spinal Cord Segmentation by Hybrid Residual Attention-Aware Convolutional Neural Networks and Learning Rate Optimization on Real World Data	A. Bueno Gómez, A. Alberich- Bayarri, I. Bosch- Roig, J. Carreres- Polo	2021	Presentación de investigación
Journal of Digital Imaging	Revista	Automated Cervical Spinal Cord Segmentation in Real- World MRI of Multiple Sclerosis Patients by Optimized Hybrid Residual Attention- Aware Convolutional Neural Network	A. Bueno Gómez, I. Bosch-Roig, A. Rodríguez Ortega, A. Jiménez Pastor, J. Carreres-Polo, Matías Fernández, L. Marti-Bonmati, A. Alberich-Bayarri	2022	Original paper

Anexo III: Índice de abreviaturas

25FWT	25-Foot Walking Test	NAIMS	National Institutes of Health and the North American Imaging in Multiple Sclerosis Cooperative
9HPY	Hole Peg Test	NIFTI	Neuroimaging Informatics Technology Initiative
ASNR	American Society of Neuroradiology	NMSS	National Multiple Sclerosis Society
BN	Batch Normalization	OCB	Bandas Oligoclonal
CDMS	Clinically Definite Multiple Sclerosis	P25	Percentil 25
CIS	Síndrome Clínicamente Aislado	P50	Percentil 50
CLR	Cyclic Learning Rate	P75	Percentil 75
CMSC	Consortium of Multiple Sclerosis Centers	PACS	Picture Archiving and Communication System
CNN	Convolutional Neural Network	QIBs	Quantitative Imaging Biomarkers
DL	Deep Learning	RA-UNet	Residual Attention-aware UNet
DSC	Sørensen-Dice Coefficient	ReLU	Rectified Linear Unit
EDSS	Escala de Discapacidad Extendida de Kurtzke	RF	Radio Frecuencia
EM	Esclerosis Múltiple	RIS	Síndrome Radiológicamente Aislado
EMPP	Esclerosis Múltiple Primaria Progresiva	RM	Resonancia Magnética
EMPR	Esclerosis Múltiple Progresiva Recurrente	RNN	Redes Neuronales Recurrentes
EMRR	Esclerosis Múltiple Recurrente Remitente	ROI	Region of Interest
EMSP	Esclerosis Múltiple Secundaria Progresiva	RSNA	Radiological Society of North America
FA	Flip ANgle	RWD	Real-World Data
FN	Falso Negativo	SF	Sistema Funcional
FOV	Field of View	SNC	Sistema Nervioso Central
FP	Falso Positivo	SNR	Relación Señal-a-Ruido
FTL	Focal Tversky Loss	STIR	Short Tau Inversion Recovery
GAN	Redes Adversarias Generativa	TC	Tomografías Computarizadas
GPU	Graphical Processing Unit	TE	Tiempo de Eco
IA	Inteligencia Artificial	TL	Tversky Loss
IgG	Inmunoglobina G	TR	Tiempo de Relajación
IoU	Intersección sobre la Unión	UCCA	Upper Cervical Cord Area
LCR	Líquido cefalorraquídeo	VN	Verdadero Negativo
MCC	Matthews Correlation Coefficient	Voxel	VOlume pixEL
MS	Machine Learning	VP	Verdadero Positivo

Tabla de ilustraciones

FIGURA 1. REPRESENTACIÓN GRÁFICA DE LA DESMIELINIZACIÓN. (R. H. MILLER ET AL., 2017).....	3
FIGURA 2. TRAYECTORIA DE LA ENFERMEDAD DE EMRR. PRESENCIA DE RECAÍDAS CON LEVES REMISIONES ENTRE ELLAS. LAS RECAÍDAS PUEDEN DURAR ENTRE 24-48 HORAS.	7
FIGURA 3. TRAYECTORIA DE LA EMSP, SIMILAR A LA FASE TEMPRANA-MEDIA DE LA EMRR, MOSTRANDO UNA PÉRDIDA MÁS CONTINUA DE LAS FUNCIONES FÍSICAS Y COGNITIVAS.	9
FIGURA 4. TRANSCURSO DE EMPP. NO MUESTRA RECAÍDAS, PERO SI UNA PÉRDIDA GRADUAL DE LAS FUNCIONES FÍSICAS Y COGNITIVAS CON LOS AÑOS. PUEDE HABER PRESENCIA DE LEVES PERIODOS DE ESTABILIDAD O MEJORAS, PERO POCO FRECUENTE.	10
FIGURA 5. CURSO CLÍNICO DE LA EM (FILIPPI ET AL., 2018). SI NO SE TRATA, LA EMRR PUEDE EVOLUCIONAR A EMSP, DONDE LOS NIVELES DE DISCAPACIDAD AUMENTAN CONTINUAMENTE SIN PERIODOS DE RECAÍDA NI REMISIONES.....	12
FIGURA 6. IZQUIERDA: IMAGEN COMPONENTE DE SECUENCIA TRANSVERSAL 3D POTENCIADA EN T1. DERECHA: AMPLIACIÓN DE LA REGIÓN PERTENECIENTE A LA MÉDULA.....	17
FIGURA 7. EJEMPLO ILUSTRATIVO SE IMÁGENES AXIALES 2D COMPONENTES DE LA SECUENCIA AXIAL 3D.	18
FIGURA 8. EFECTO DEL VOLUMEN PARCIAL EN UNA REGIÓN DE LA IMAGEN (IZQUIERDA) Y MISMA REGIÓN AUMENTADA (DERECHA).	20
FIGURA 9. ARTEFACTO DE RUIDO BIAS-FIELD.	21
FIGURA 10. SE MUESTRA LA RÁPIDA DECAÍDA DE LAS TASAS DE ERROR GANADORAS DEL ILSVRC (LÍNEA ROJA) Y EL CRECIENTE NÚMERO DE PARTICIPACIONES QUE EMPLEAN GPU CADA AÑO (BARRAS EN VERDE) (NVIDIA AND IBM CLOUD SUPPORT IMAGE NET LARGE SCALE VISUAL RECOGNITION CHALLENGE NVIDIA TECHNICAL BLOG, N.D.).	40
FIGURA 11. EJEMPLO DE CLASIFICACIÓN DE MÉDULA ESPINAL. UNA IMAGEN DE ENTRADA 2D AXIAL SE PROCESA A TRAVÉS DE UNA CNN EN FORMA ESQUEMÁTICA. LA EXTRACCIÓN DE CARACTERÍSTICAS ES LLEVADA A CABO POR LAS CAPAS DE CONVOLUCIÓN PARA DESPUÉS PASAR LA CAPA DE MAX-POOLING Y APLICAR UN SUBMUESTRO DE FACTOR 2. ESTE PROCESO ES REPETIDO EN LA SEGUNDA CAPA DE CONVOLUCIÓN Y MAX-POOLING. POR ÚLTIMO, LA SALIDA DE LA SEGUNDA CAPA DE POOLING ES PASADA A UN VECTOR UNIDIMENSIONAL (FLATTEN) Y SE GENERA LA PROBABILIDAD DE CLASIFICACIÓN MEDIANTE LA CAPA TOTALMENTE CONECTADA.....	42
FIGURA 12. COMPARACIÓN DE LAS FUNCIONES DE ACTIVACIÓN. RELU (IZQUIERDA). LEAKY RELU (DERECHA). SIENDO EL EJE X LA ENTRADA A LA RED Y EL EJE Y LA SALIDA. (ACTIVATION FUNCTIONS: WITH REAL-LIFE ANALOGY AND PYTHON CODE - SHIKSHA ONLINE, N.D.)	46
FIGURA 13. COMPARACIÓN DE LAS FUNCIONES DE ACTIVACIÓN. SIGMOID (IZQUIERDA). TANGENTE HIPERBÓLICA (DERECHA). SIENDO EL EJE X LA ENTRADA A LA RED Y EL EJE Y LA SALIDA. (ACTIVATION FUNCTIONS: WITH REAL-LIFE ANALOGY AND PYTHON CODE - SHIKSHA ONLINE, N.D.)	47
FIGURA 14. PROPAGACIÓN HACIA DELANTE (IZQUIERDA) Y HACIA ATRÁS (DERECHA). LA PRIMERA CALCULA LA SALIDA DEL MODELO A TRAVÉS DE TODAS LAS CAPAS, LA SEGUNDA ACTUALIZA LOS PESOS DE LAS CAPAS PARA MINIMIZAR AL MÁXIMO LAS PÉRDIDAS EN LA SIGUIENTE ITERACIÓN.	49
FIGURA 15. PROCESO DE ENTRENAMIENTO DE UNA RED NEURONAL. LAS REDES NEURONALES APRENDEN EN UN PROCESO ITERATIVO PROPAGANDO LA INFORMACIÓN HACIA DELANTE Y HACIA ATRÁS MEDIANTE UNA FUNCIÓN DE PÉRDIDAS. LA RED SE CENTRA EN MINIMIZAR LA FUNCIÓN DE PÉRDIDAS ACTUALIZANDO SUS PESOS DURANTE LA PROPAGACIÓN HACIA ATRÁS, DE MANERA QUE LA NUEVA INFORMACIÓN DE SALIDA SE RETROALIMENTA A LA PÉRDIDA POR PROPAGACIÓN DIRECTA.	50
FIGURA 16. COMPONENTES DE MATRIZ DE CONFUSIÓN.	55
FIGURA 17. DEMOSTRACIÓN DEL COMPORTAMIENTO DE LAS MÉTRICAS EN TÉRMINOS DE DISTINTOS TAMAÑOS DE ROI EN COMPARACIÓN A LA IMAGEN COMPLETA. PODEMOS OBSERVAR VARIOS ESCENARIOS EN EL ÁMBITO DE LA SEGMENTACIÓN DE IMAGEN MÉDICA. ESCENARIOS: A) SIN SEGMENTAR (NO HAY PÍXELES ETIQUETADOS EN LA ROI EN COLOR ROJO), B) SEGMENTACIÓN COMPLETA (TODOS LOS PÍXELES ETIQUETADOS COMO ROI, TODOS EN COLOR ROJO), C) SEGMENTACIÓN ALEATORIA (ETIQUETADO DE LA ROI COMPLETAMENTE ALEATORIO), D) SIN ENTRENAR (DESPUÉS DE POCAS ÉPOCAS DE ENTRENAMIENTO), E) ENTRENADO (MODELO COMPLETAMENTE AJUSTADO, SOLO EN COLOR ROJO LA LESIÓN BUSCADA), F) GROUND-TRUTH.....	59
FIGURA 18. DIAGRAMA REPRESENTATIVO DE LA CONEXIÓN DE SALTO.	64

Tabla de ilustraciones

FIGURA 19. VISUALIZACIÓN DE POR QUÉ AÑADIR CONEXIONES RESIDUALES PUEDE AYUDAR CUANDO SE ENTRENA UN MODELO MUY PROFUNDO.	65
FIGURA 20. ARQUITECTURA DEL MÓDULO DE ATENCIÓN. EN LA PARTE SUPERIOR, EL MÓDULO DE ATENCIÓN COMPUESTO POR LA RAMA FUERTE Y LA RAMA DÉBIL. LA RAMA FUERTE APRENDE CARACTERÍSTICAS ORIGINALES DE LA IMAGEN MIENTRAS QUE LA RAMA DÉBIL SE CENTRA EN REDUCIR RUIDO Y MEJORAR LAS CARACTERÍSTICAS QUE SEÑALA LA MÁSCARA. EN LA PARTE INFERIOR, LA RAMA DÉBIL COMPUESTA POR UNA AGRUPACIÓN DE BLOQUES CODIFICADOR-DECODIFICADOR. LA P INDICA LA PROFUNDIDAD DE LAS CONEXIONES DE SALTO.	68
FIGURA 21. ARQUITECTURA U-NET (REPRESENTACIÓN DE EJEMPLO PARA LA MENOR RESOLUCIÓN, 32x32 PÍXELES). CADA CAJA AZUL CORRESPONDE A UN MAPA DE CARACTERÍSTICAS MULTICANAL. EL NÚMERO DE CANALES ESTÁ INDICADO ENCIMA DE CADA CAJA Y EL TAMAÑO X-Y EN LA PARTE INFERIOR IZQUIERDA. LAS CAJAS BLANCAS REPRESENTAN MAPAS DE CARACTERÍSTICAS COPIADOS Y LAS FLECHAS REPRESENTAN LAS DIFERENTES OPERACIONES.	70
FIGURA 22. ARQUITECTURA GENERAL BASADA EN BLOQUES RESIDUALES Y MECANISMOS DE ATENCIÓN. LOS BLOQUES RESIDUALES PERMITEN EL DISEÑO DE UNA RED MÁS PROFUNDA SOLVENTANDO EL PROBLEMA DE DESVANECIMIENTO DE GRADIENTE. EL MECANISMO DE ATENCIÓN RESUELVE EL PROBLEMA DE DESBALANCEO DE DATOS APRENDIENDO A FOCALIZAR EL APRENDIZAJE EN LAS LOCALIZACIONES RELEVANTES.	72
FIGURA 23. DISEÑO DEL BLOQUE RESIDUAL. EL MAPEO DE IDENTIDAD FUNCIONA COMO CONEXIÓN DE SALTO, LA SALIDA DEL MAPEO DE IDENTIDAD SERÁ AÑADIDA A LA SALIDA DE LOS BLOQUES APILADOS.....	72
FIGURA 24. REPRESENTACIÓN DE FTL VARIANDO LOS VALORES γ . CUANDO $\gamma = 1$, TENEMOS LAS PÉRDIDAS TVERSKY ESTÁNDAR. (XU ET AL., 2020)	76
FIGURA 25. TÉCNICA DE TASA DE APRENDIZAJE TRIANGULAR. LAS LÍNEAS DISCONTINUAS REPRESENTAN LA TASA DE APRENDIZAJE VARIANDO ENTRE LOS VALORES LÍMITE. EL PARÁMETRO DE ENTRADA STEPSIZE ES EL NÚMERO DE ITERACIONES EN MEDIO CICLO.	77
FIGURA 26. DIAGRAMA DE FLUJO UTILIZADO PARA DESARROLLAR LA HERRAMIENTA DE SEGMENTACIÓN AUTOMÁTICA DE MÉDULA CERVICAL PROPUESTA. PARTIENDO DE LA GENERACIÓN DEL CONJUNTO DE DATOS (PASO 1) Y PASANDO POR SU NORMALIZACIÓN (PASO 2) Y ETIQUETADO (PASO 3) HASTA LLEGAR AL ENTRENAMIENTO Y VALIDACIÓN DE LA ARQUITECTURA PROPUESTA (PASO 4).	84
FIGURA 27. INTERFAZ GRÁFICA DE ITK-SNAP. VENTANA DIVIDIDA POR UNA SECCIÓN DE HERRAMIENTAS DISPONIBLES A LA IZQUIERDA, Y UNA CUADRÍCULA COMPUESTA POR 4 IMÁGENES DEL MISMO VOLUMEN, 3 VISTAS EN 2D Y UNA EN 3D.	87
FIGURA 28. IMAGEN CORPORATIVA DE ANACONDA (WELCOME! — ANACONDA DOCUMENTATION, N.D.)	88
FIGURA 29. IMAGEN CORPORATIVA DE ITK-SNAP (ITK-SNAP HOME, N.D.)	89
FIGURA 30. REPRESENTACIÓN DE TENSOR 3D DE IMÁGENES 2D.....	91
FIGURA 31. DIAGRAMA DE BLOQUES DEL MODELO DE SEGMENTACIÓN ATTENTION U-NET. LAS PUERTAS DE ATENCIÓN FILTRAN LAS CARACTERÍSTICAS PROPAGADAS A TRAVÉS DE LAS CONEXIONES DE SALTO. IMAGEN EXTRAÍDA DEL PAPER ORIGINAL DE ATTENTION U-NET (OKTAY ET AL., 2018).	93
FIGURA 32. ESTRUCTURA DEEP U-NET. IMAGEN EXTRAÍDA DEL PAPER ORIGINAL (LI ET AL., N.D.)	94
FIGURA 33. ESTRUCTURA DE U-NET ++. IMAGEN EXTRAÍDA DEL PAPER ORIGINAL (Z. ZHOU ET AL., 2018).....	95
FIGURA 34. PROCEDIMIENTO A SEGUIR EN LA TÉCNICA DE LEARNING RATE FINDER (BUSCADOR DE TASA DE APRENDIZAJE)....	98
FIGURA 35. PÉRDIDAS EN FUNCIÓN DE LA TASA DE APRENDIZAJE PARA LOCALIZAR LOS LÍMITES ÓPTIMOS DE CLR. <i>maxlr</i> : LAS PÉRDIDAS EMPIEZAN A AUMENTAR, <i>baselr</i> : LAS PÉRDIDAS COMIENZAN A BAJAR.....	98
FIGURA 36. TASA DE APRENDIZAJE VARIANDO ENTRE LOS LÍMITES ESTABLECIDOS. 'TRAINING ITERATIONS' ES EL NÚMERO DE ACTUALIZACIONES DE PESOS POR ÉPOCA, ES DECIR, EL NÚMERO TOTAL DE MUESTRAS DE ENTRENAMIENTO DIVIDIDO POR EL TAMAÑO DE BATCH.....	99
FIGURA 37. REPRESENTACIÓN DE VÓXEL PERTENECIENTE A LA MÁSCARA DE SEGMENTACIÓN Y SUS 3 COMPONENTES DIMENSIONALES.	101
FIGURA 38. DIAGRAMA DE FLUJO PARA CALCULAR LA VOLUMETRÍA DE LA SEGMENTACIÓN DE MÉDULA ESPINAL CERVICAL.	102
FIGURA 39. PÉRDIDAS Y MCC DEL SET DE VALIDACIÓN DURANTE EL PROCESO DE ENTRENAMIENTO.	104
FIGURA 40. EVOLUCIÓN DEL VALOR DICE DURANTE LAS ÉPOCAS DEL ENTRENAMIENTO UTILIZANDO CLR Y UNA TASA DE APRENDIZAJE FIJA DE $1e - 4$. CON CLR SE ALCANZA EN MENOS ÉPOCAS EL VALOR MAYOR DICE.....	105
FIGURA 41. HISTOGRAMA DE COEFICIENTES DICE GENERADOS DURANTE EL ENTRENAMIENTO.	106
FIGURA 42. REPRESENTACIÓN COMPARATIVA DE DISTRIBUCIÓN DE VALORES DE LAS DISTINTAS MÉTRICAS DE EVALUACIÓN UTILIZADAS.....	108
FIGURA 43. HISTOGRAMA DE COEFICIENTES DICE EN TEST.	109

Tabla de ilustraciones

FIGURA 44: REPRESENTACIÓN COMPARATIVA DE DISTRIBUCIÓN DE VALORES DE LAS DISTINTAS MÉTRICAS DE EVALUACIÓN UTILIZADAS EN VALIDACIÓN.....	110
FIGURA 45. HISTOGRAMA DE DISTRIBUCIÓN DE COEFICIENTES DICE OBTENIDOS EN VALIDACIÓN.	111
FIGURA 46. MÉDULA CERVICAL EN 3D EN VARIAS PERSPECTIVAS. EN VERDE LOS VÓXELES COINCIDENTES ENTRE HERRAMIENTA PROPUESTA Y RADÍÓLOGO. EN ROJO LOS NO COINCIDENTES, LA MAYORÍA POR MAYOR LONGITUD DE SEGMENTACIÓN DE LA PORCIÓN MEDULAR.	112
FIGURA 47. EJEMPLO DE SEGMENTACIÓN EN PLANO TRANSVERSAL CON DICE ALTO. A) REAL WORLD DATA, B) GROUND TRUTH, C) MÁSCARA PREDICHA POR EL MODELO PROPUESTO, D) SUPERPOSICIÓN ENTRE GROUND TRUTH Y MÁSCARA PREDICHA (EN VERDE ZONAS COINCIDENTES, EN ROJO ZONAS NO COINCIDENTES).....	113
FIGURA 48. EJEMPLO DE SEGMENTACIÓN EN PLANO TRANSVERSAL CON DICE BAJO. EJEMPLO DE SEGMENTACIÓN CON DICE BAJO. A) REAL WORLD DATA, B) GROUND TRUTH, C) MÁSCARA PREDICHA POR EL MODELO PROPUESTO, D) SUPERPOSICIÓN ENTRE GROUND TRUTH Y MÁSCARA PREDICHA (EN VERDE ZONAS COINCIDENTES, EN ROJO ZONAS NO COINCIDENTES).....	114
FIGURA 49. EJEMPLO DE MENOR PRECISIÓN EN LA SEGMENTACIÓN. EN AMARILLO Y EN MORADO SE INDICAN LAS HETEROGENEIDADES PROVOCADAS POR LA MALA DIFERENCIACIÓN EN LA IMAGEN ENTRE LA MÉDULA Y LAS ESTRUCTURAS COLINDANTES.....	115
FIGURA 50. ANÁLISIS DE CASO DE VALIDACIÓN. A) SEGMENTACIÓN EN 3D, B) CORTE TRANSVERSAL EN EL QUE DIFIEREN SEGMENTACIÓN MANUAL Y AUTOMÁTICA, B1) AMPLIACIÓN DE B), EN ROJO LOS PÍXELES QUE AÑADE LA SEGMENTACIÓN AUTOMÁTICA FRENTE A LA MANUAL, C) CORTE TRANSVERSAL PERTENECIENTE A MAYOR LONGITUD DE SEGMENTACIÓN, C1) AMPLIACIÓN DE C).	116
FIGURA 51. ANÁLISIS DE CASO DE VALIDACIÓN. A) SEGMENTACIÓN 3D, ERROR DE LA HERRAMIENTA INDICADO EN AMARILLO, B) ÚLTIMO CORTE SUPERIOR EN EL QUE COINCIDEN SEGMENTACIÓN MANUAL Y HERRAMIENTA, C)-E) EXTENSIÓN LONGITUDINAL DE CORTES PERTENECIENTES A LA SEGMENTACIÓN MANUAL, F) CORTE CON ZONA ERRÓNEA, G) MISMO CORTE SIN SUPERPOSICIÓN DE MÁSCARA.....	117
FIGURA 52. ANÁLISIS DE CASO DE VALIDACIÓN. A) SEGMENTACIÓN EN 3D, B) CORTE TRANSVERSAL PERTENECIENTE A EXTENSIÓN DE LONGITUD DE SEGMENTACIÓN, B1) AMPLIACIÓN DE B), C) CORTE TRANSVERSAL EN EL QUE DIFIEREN SEGMENTACIÓN MANUAL Y AUTOMÁTICA, EN ROJO LOS PÍXELES QUE AÑADE LA SEGMENTACIÓN AUTOMÁTICA FRENTE A LA MANUAL, C1) AMPLIACIÓN DE C.	118
FIGURA 53. ANÁLISIS DE CASO DE VALIDACIÓN. A) CONFUSIÓN DE MÉDULA CON VÉRTEBRA POR LA HERRAMIENTA AUTOMÁTICA, B) MAYOR ADICIÓN DE PÍXELES A LA MÁSCARA POR LA SEGMENTACIÓN AUTOMÁTICA.....	119
FIGURA 54. HISTOGRAMAS DE DISTRIBUCIÓN DICE DE LAS CONFIGURACIONES MOSTRADAS EN LA TABLA 6, DESDE LA IMAGEN (A) HASTA LA (I), EN EL MISMO ORDEN DE APARICIÓN. EVOLUCIÓN PROGRESIVA DE LOS MODELOS, DISTRIBUCIONES MENOS DISPERSAS Y MÁS CENTRADAS EN VALORES ALTOS EN CADA ENSAYO.....	124
FIGURA 55. MUESTRA VISUAL DE PRUEBA CON ARQUITECTURA U-NET. TERCERA CONFIGURACIÓN DE LA TABLA 9.	125
FIGURA 56. MUESTRA VISUAL DE PRUEBA CON ARQUITECTURA U-NET. QUINTA CONFIGURACIÓN DE LA TABLA 9.	125
FIGURA 57. MUESTRA VISUAL DE PRUEBA CON ARQUITECTURA U-NET. OCTAVA CONFIGURACIÓN DE LA TABLA 9.	126
FIGURA 58. MUESTRA VISUAL CON NUESTRO MÉTODO.	126
FIGURA 59. DISTRIBUCIÓN DE VALORES DE SECCIÓN MEDIA ENTRE LAS DISTINTAS FORMAS CLÍNICAS QUE COMPONEN NUESTRA BASE DE DATOS.....	128
FIGURA 60. VALORES DE EDSS POR FORMA CLÍNICA.	129
FIGURA 61. RELACIÓN ENTRE SECCIÓN MEDIA Y EDSS EN LAS DIFERENTES FORMAS CLÍNICAS. LAS LÍNEAS MUESTRAN LA TENDENCIA LINEAL DE LOS DATOS Y SE MUESTRAN LOS DISTINTOS VALORES DE CORRELACIÓN DE PEARSON.....	130
FIGURA 62. DISTRIBUCIÓN DE VALORES DE SECCIÓN MEDIA POR FORMA CLÍNICA Y RELACIÓN CON EDSS MEDIA.....	131
FIGURA 63. SECUENCIAS PROCEDENTES DE LA MISMA ADQUISICIÓN. STIR (IZQUIERDA), MAYOR SENSIBILIDAD, PRESENCIA DE CLARAS HIPERINTENSIDADES FOCALES EN LA MÉDULA ESPINAL, SEÑALADAS EN ROJO. SAGITAL T2 (DERECHA), MAYOR ESPECIFICIDAD, FALTA DE CLARIDAD EN LAS SUPUESTAS LESIONES, LAS QUE SE INTUYEN, RODEADAS EN ROJO.....	161
FIGURA 64. IMAGEN SAGITAL POTENCIADA T2. LA MÉDULA CERVICAL APARECE COMO UNA ESTRUCTURA GRIS OSCURA RODEADA POR EL LCR Y GRASA INTRAESPINAL.	163
FIGURA 65. DISTINTOS CORTES SAGITALES COMPONENTES DE UNA SECUENCIA SAGITAL T2.....	163
FIGURA 66. IMAGEN 2D STIR. EL LCR APARECE BRILLANTE, LA GRASA NEGRA Y LA MÉDULA ESPINAL GRIS OSCURA. SEÑALADAS EN AMARILLO, OBSERVAMOS LA PRESENCIA DE LESIONES COMO HIPERINTENSIDADES EN EL NIVEL C1-C7.	165

Tabla de ilustraciones

FIGURA 67. CURVAS DE RECUPERACIÓN T1 DE LÍPIDOS, TEJIDO DE MÉDULA ESPINAL SANA (NASC) Y TEJIDO CORRESPONDIENTE A LESIONES. LA LÍNEA DISCONTINUA INDICA EL PUNTO NULO COMO TIEMPO TRAS LA INVERSIÓN EN EL QUE LA MAGNETIZACIÓN LONGITUDINAL DE LOS LÍPIDOS ES CERO.....	167
FIGURA 68. EL REGISTRO DE IMÁGENES ES EL PROCESO DE DEFORMAR LOS PUNTOS DE LA IMAGEN MÓVIL, P, A LOS PUNTOS BIOLÓGICAMENTE CORRESPONDIENTES, Q, EN EL DOMINIO DE LA IMAGEN FIJA.	169
FIGURA 69. DIAGRAMA DE TAREAS PARA LA SEGMENTACIÓN AUTOMÁTICA DE LESIONES EN MÉDULA CERVICAL.....	175
FIGURA 70. RECONSTRUCCIÓN 3D SECUENCIA AXIAL Y PLANOS QUE LA COMPONEN	176
FIGURA 71. RECONSTRUCCIÓN 3D SECUENCIA STIR Y PLANOS QUE LA COMPONEN	177
FIGURA 72. CORTES PERTENECIENTES A SECUENCIA STIR CON INFORMACIÓN SUPERPUESTA DE LOCALIZACIÓN DE MÉDULA CERVICAL PROCEDENTE DE SECUENCIA AXIAL 3D.....	179
FIGURA 73. EJEMPLO DE MÁSCARA DE MÉDULA ESPINAL GENERADA EN SAGITAL (COLUMNA IZQUIERDA) TRAS EL REGISTRO ENTRE MÁSCARA EN TRANSVERSAL Y SECUENCIA STIR (COLUMNA DERECHA).	180
FIGURA 74. SIMILITUD DE INTENSIDAD ENTRE LCR (REGIÓN EN AMARILLO) Y LESIÓN (REGIÓN EN ROJO).....	181
FIGURA 75. A LA IZQUIERDA, MÁSCARAS GENERADAS POR EL REGISTRO TRAS APLICAR DILATACIÓN, A LA DERECHA, MISMAS MÁSCARAS TRAS AÑADIR TAMBIÉN EROSIÓN.....	182
FIGURA 76. A LA IZQUIERDA, MÁSCARA FINAL TRAS OPERACIÓN DE CIERRE, A LA DERECHA, IMAGEN DE INTENSIDAD RESULTANTE TRAS APLICAR LA MÁSCARA A LA SECUENCIA STIR.	182
FIGURA 77. ARRIBA: DETECCIÓN DE PÍXELES QUE CONSIDERAMOS DIVISORES ENTRE MÉDULA Y LCR, ABAJO: CANAL MEDULAR SIN DICHS PÍXELES EN EL BORDE.....	186
FIGURA 78. ARRIBA: MÉDULAS TRAS ELIMINAR LCR, ABAJO: ZONAS HIPERINTENSAS DETECTADAS MEDIANTE CLUSTERING.	187