



# Desarrollo de una herramienta basada en algoritmos de aprendizaje profundo para la evaluación automática de la musculatura de la zona lumbar a partir de imágenes de resonancia magnética

ESTUDIANTE: Germán Balerdi

LEGAJO: 8285

SUPERVISOR: Martín Alberto Belzunce

LUGAR DE TRABAJO: CEUNIM

FECHA: Diciembre 2023





# Agradecimientos

Quiero agradecer a todos los responsables de que en este documento esté presentando mi Proyecto Final de Ingeniería Biomédica.

Para empezar, quiero agradecer al profesor Alister Hart y al doctor Johann Henckel, que aportaron los datos e imágenes fundamentales para el desarrollo del trabajo, y también al CEUNIM, que proveyó de los recursos computacionales necesarios para entrenar los modelos de aprendizaje profundo.

En términos personales, y en el periodo que corresponde a esta última etapa de mi carrera universitaria, quiero agradecer a Martín Belzunce, que desde un principio me presentó un proyecto con todos los elementos disponibles, que me guío, corrigió y dedicó de su tiempo para que este trabajo pudiese llevarse a cabo. También quiero agradecerle a Agustina Portu, quien me aconsejó a lo largo de todo el proyecto y organizó la evaluación del presente trabajo. Y por último al Departamento de Ingeniería Clínica del sanatorio Las Lomas que me brindó el tiempo y el espacio para poder trabajar en el proyecto.

En lo que concierne a mi carrera universitaria, primero quiero agradecerle a la UNSAM que me formó tanto como profesional y como persona, que permitió que conociera gente de calidad y brindó de todas las herramientas para que lo lograra. De la misma manera quiero agradecer a los compañeros con los que cursé y estudié, en especial a Lucila con quien preparar finales se volvió ameno y realizable.

En lo paralelo a la vida universitaria, le agradezco a mis amigos que siempre me escucharon y motivaron. También a mi novia Lucía, que me acompañó tanto en los buenos como en los malos momentos que transcurrí durante la carrera.

Por último, quiero agradecer principalmente a mi familia que me apoyó siempre y que, sin su trabajo y sin la educación que me brindaron, definitivamente, no estaría en esta instancia.

# Índice

## Contenido

<b>Agradecimientos .....</b>	3
<b>Índice .....</b>	4
<b>Glosario .....</b>	6
<b>Resumen .....</b>	7
<b>1 Introducción .....</b>	9
<b>1.1 Lugar de trabajo y contexto .....</b>	9
<b>1.2 Objetivos .....</b>	11
<b>1.3 Estructura del informe .....</b>	12
<b>2 Marco Teórico .....</b>	13
<b>2.1 Calidad de musculatura lumbar como indicador de salud y estilo de vida .....</b>	13
<b>2.2 Secuencia DIXON en imágenes de resonancia magnética .....</b>	14
<b>2.3 Introducción al Deep learning .....</b>	16
<b>2.3.1 Redes neuronales artificiales .....</b>	17
<b>2.3.2 Función de pérdida y optimizador .....</b>	19
<b>2.3.3 Set de datos .....</b>	20
<b>2.4 Redes neuronales convolucionales .....</b>	22
<b>2.4.1 Convolución .....</b>	23
<b>2.4.2 Max Pooling .....</b>	24
<b>2.4.3 Redes neuronales convolucionales de clasificación .....</b>	25
<b>2.5 Arquitectura U-Net .....</b>	26
<b>3 Redes neuronales convolucionales para la segmentación automática de los músculos lumbares .....</b>	29
<b>3.1 Datos .....</b>	29
<b>3.1.1 Adquisición de datos .....</b>	29
<b>3.1.2 Construcción del set de datos .....</b>	29
<b>3.1.3 División del set de datos en entrenamiento y validación .....</b>	34

<b>3.2 Componentes esenciales del entrenamiento .....</b>	35
<b>3.2.1 Métricas de evaluación.....</b>	35
<b>3.2.2 Función de pérdida y optimizador .....</b>	35
<b>3.3 Red neuronal convolucional 2D .....</b>	36
<b>3.3.1 Mono clase.....</b>	36
<b>3.3.2 Multiclasificación.....</b>	40
<b>3.4 Red neuronal convolucional 3D.....</b>	50
<b>4. Evaluación de la calidad muscular en sujetos sanos sedentarios y activos.....</b>	56
<b>4.1 Procesamiento de imágenes .....</b>	56
<b>4.2 Resultados Cuantitativos .....</b>	57
<b>4.2.1 Fracción de tejido graso volumétrico .....</b>	57
<b>4.2.2 Tamaño de musculatura de zona lumbar.....</b>	61
<b>5. Discusión y conclusiones.....</b>	67
<b>5.1 Discusión .....</b>	67
<b>5.2 Conclusiones .....</b>	68
<b>Bibliografía.....</b>	71

# Glosario

*P: Músculo Psoas*

*CL: Músculo Cuadrado Lumbar*

*ES: Músculo Erector Spinae*

*M: Músculo Multifidus*

*I: Músculo Iliaco*

*IP: Imagen Dixon In Phase*

*OP: Imagen Dixon Out of Phase*

*F: Imagen Dixon de Grasa*

*W: Imagen Dixon de Agua*

*FF: Imagen de fracción de tejido graso*

*CSA: Sección de área transversal*

*L4: Altura de cuarta vértebra lumbar*

*RMN: Resonancia magnética nuclear*

*STD: Desviación estándar*

*IQR: Rango intercuartílico*

*FOV: Campo de visión*

*RNC: Red neuronal convolucional*

*RNA: Red neuronal Artificial*

## Resumen

Múltiples estudios vinculan el deterioro de la calidad de los músculos de zona lumbar con la lumbalgia, una dolencia de espalda baja que es la principal causa de discapacidad laboral y produce costos económicos y sociales graves. Este deterioro puede deberse al sedentarismo o afecciones como la sarcopenia que implican la pérdida de masa muscular y fuerza con la vejez. Fisiológicamente, el deterioro se presenta como una reducción del tamaño del tejido muscular o un aumento en la infiltración de tejido graso.

Generalmente, las evaluaciones de calidad muscular basadas en imágenes consisten en un diagnóstico descriptivo y subjetivo sobre imágenes de resonancia magnética o tomografía axial computada realizado por un médico radiólogo en el cual se informan los niveles generales de infiltración de tejido graso y si hay presencia de atrofia. Sin embargo, existen técnicas de resonancia magnética, como por ejemplo la secuencia Dixon, que permiten adquirir imágenes cuantitativas del nivel de grasa intramuscular a nivel de voxel. Para evaluar un músculo en particular, es necesario segmentar el tejido de interés en los múltiples cortes que componen la imagen, demandando mucho tiempo de un profesional entrenado e impidiendo que esta técnica sea utilizada. Por esta razón, proponemos utilizar métodos de aprendizaje profundo para la segmentación automática de los músculos de la zona lumbar.

En este trabajo, se entrena una red neuronal convolucional con arquitectura U-Net 3D capaz de segmentar los músculos Psoas (P), Iliaco (I), Cuadrado Lumbar (CL), Erector Spinae y Multífido (ES+M) a partir de imágenes Dixon tridimensionales. A su vez se entrena un modelo 2D capaz de segmentar estos mismos músculos en un único corte axial a la altura de la cuarta vértebra lumbar (L4). De esta manera, a partir de segmentaciones automáticas se podrá realizar evaluaciones de calidad muscular en términos de fracción de tejido graso y tamaño. El entrenamiento se realiza con un set de datos compuesto por 26 imágenes de la región lumbar manualmente segmentadas para el modelo 3D y 58 imágenes de un corte axial a la altura de L4 manualmente segmentadas para el modelo 2D que luego son aumentadas artificialmente.

Los modelos entrenados fueron evaluados con el puntaje Dice en el set de validación. La mediana de los resultados del modelo 3D para los músculos Psoas y los conjuntos ES+M fueron superiores a 0.9, mientras que, para los Cuadrados lumbares e

Iliacos, mayores a 0.85. En el caso del modelo 2D, se obtuvieron puntajes mayores a 0.95 para los Psoas y ES+M y mayores a 0.92 para los cuadrados lumbares.

A partir de estos modelos, se evaluó la calidad muscular en términos de fracción de tejido graso y tamaño en 118 sujetos pertenecientes a un grupo de personas sedentarias y a otro de ciclistas recreacionales. Empleando la prueba de Mann-Whitney, encontramos que el tamaño normalizado por peso de la musculatura de hombres es mayor que el de las mujeres, con una diferencia significativa en todos los músculos ( $p<0.05$ ). Por parte de la infiltración de tejido graso, las mujeres poseen una fracción mayor en todos los músculos ( $p<0.05$ ) excepto por el músculo Iliaco ( $p=0.10$ ). De esta misma manera, también realizamos las evaluaciones entre ciclistas y sujetos sedentarios de ambos sexos y determinamos que la musculatura es más grande y magra para los ciclistas ( $p<0.05$ ). Por otro lado, usando regresiones lineales entre las métricas obtenidas e información demográfica y antropométrica se determinó que la edad es el factor que más influye en la infiltración de tejido graso mientras que la altura y el peso influyen en el tamaño. Por último, se llevó a cabo una regresión lineal entre las métricas de salud muscular bidimensionales y tridimensionales y se concluyó que es viable estimar la calidad muscular de todo el conjunto ES+M a partir de un solo corte a la altura de L4, no así para el resto de los músculos.

En conclusión, en este trabajo presentamos una novedosa herramienta para evaluar de forma automática la calidad de los músculos de la zona lumbar a partir de imágenes de resonancia magnética. Adicionalmente, empleando esta herramienta se obtuvieron resultados originales de valores de referencia de tamaño e infiltración graso para este grupo muscular, los cuales contribuirán a la investigación clínica y a comprender mejor las características de estos. Por último, pudimos determinar que evaluar estos músculos a partir un único corte axial como se hace muchas veces en la práctica, solo es adecuado para el grupo ES+M.

# 1 Introducción

Varios estudios relacionan la disminución de la calidad de los músculos de la región lumbar con la lumbalgia, una afección de la parte baja de la espalda que es la principal causa de discapacidad laboral y conlleva implicaciones económicas y sociales graves. Este deterioro puede estar relacionado con el sedentarismo o condiciones como la sarcopenia, que implica la pérdida de masa muscular y fuerza asociada al envejecimiento. Desde una perspectiva fisiológica, este deterioro se manifiesta como una disminución en el tamaño del tejido muscular o un aumento en la infiltración de tejido graso.

Para evaluar la calidad de determinados músculos, se suele recurrir a evaluaciones manuales realizadas por médicos o fisioterapeutas, o alternativamente a través de imágenes médicas. En este último caso, médicos radiólogos realizan evaluaciones cualitativas del tejido muscular en las imágenes que están diagnosticando. Esto se debe a que múltiples dolencias y patologías pueden estar vinculadas con el nivel de atrofia en la región afectada [1]. Este diagnóstico generalmente consiste en informar descriptivamente o clasificar utilizando escalas como la de Goutallier para el nivel de infiltración graso [2] [3]. Consecuentemente, esta evaluación resulta cualitativa y subjetiva al ojo del profesional que la realiza.

Realizar una evaluación cuantitativa de calidad muscular de manera no invasiva es una tarea desafiante que involucra la adquisición de imágenes médicas, segmentar las imágenes médicas (proceso que consiste en separar el área que ocupa el tejido de interés dentro de la imagen) y analizar sus características. Teniendo en cuenta que las imágenes de resonancia magnética nuclear (RMN) o tomografía axial computada están compuestas por un conjunto de múltiples cortes, segmentar todo el tejido de forma manual implica mucho tiempo de un profesional capacitado. Es por eso que en este trabajo se busca desarrollar una herramienta capaz de automatizar esta tarea y evaluar el tamaño y la fracción de tejido graso como métricas de calidad muscular a partir de imágenes de resonancia magnética de región lumbar adquiridas con secuencia DIXON.

## 1.1 Lugar de trabajo y contexto

Este trabajo se realiza en los laboratorios de análisis y procesamiento de imágenes del centro universitario de imágenes médicas (CEUNIM) situado dentro de la Universidad Nacional de San Martín (UNSAM). En estos se cuenta con computadoras de escritorio de

alto desempeño las cuales permiten desarrollar herramientas basadas en algoritmos iterativos o modelos de aprendizaje profundo.

El trabajo se inserta en una línea de investigación liderada por Dr. Ing. Martín Alberto Belzunce que busca estudiar métricas de calidad muscular basadas en imágenes de distintos grupos musculares en diferentes grupos de estudio [4] [5] [6] [7] [8]. Teniendo en cuenta el objetivo general de la línea de investigación, intentaremos crear una herramienta automática de evaluación cuantitativa que ayude a estudiar la calidad muscular de la zona lumbar y correlacionarla con la actividad física, el envejecimiento, patologías y dolencias vinculadas con esta región. Este conjunto muscular está compuesto por los músculos Psoas, Cuadrado Lumbar, Iliaco, Multifido y Erector Spinae y resulta relevante estudiarlo debido que proveen estabilidad postural y tracción para la marcha, soportan los movimientos de las extremidades superiores y ejercen los movimientos de la columna vertebral al mismo tiempo que la protege. Así mismo, el deterioro de estos músculos puede causar patologías severas como la camptocormia, la cual consiste en una cifosis lumbar crónica que afecta principalmente a pacientes de mayor edad, o la lumbalgia que es una dolencia que afecta al 80% de la población en algún momento de su vida y es la principal causa de discapacidad laboral implicando costos económicos y sociales graves [9] [10] [11] [12] [13] [14] [15].

Hoy en día, para evaluar cuantitativamente la composición de un tejido muscular, se puede tomar muestras mediante una biopsia o utilizando imágenes de resonancia magnética adquiridas con la secuencia DIXON. Esta última permite crear imágenes con excelente contraste entre el tejido graso y el tejido muscular, y nos permite obtener un valor cuantitativo de fracción de grasa para cada uno de los vóxeles de la imagen. No obstante, para evaluar un tejido muscular particular dentro de la imagen, es necesario realizar una segmentación de la región ocupada por el mismo. Esto consiste en delimitar el área de interés y crear una etiqueta binaria de la región. Al aplicar la etiqueta sobre la imagen original podremos extraer únicamente la información relevante. Ante la falta de herramientas tecnológicas que ayuden con esta tarea, la misma implica que un profesional de la salud con conocimientos anatómicos del músculo segmente la imagen manualmente. Esto se vuelve realmente extenso si pretende segmentar el volumen entero ya que dependiendo del tamaño de voxel, la región puede abarcar múltiples cortes axiales. Consecuentemente, automatizar esta tarea resulta de gran interés ya que

permite extraer información de múltiples pacientes en muy poco tiempo y sin necesitar de un profesional capacitado.

Dado que en los últimos años el uso de redes neuronales convolucionales (RNC) en el procesamiento de imágenes ha proliferado debido a sus excelentes resultados, se propone la implementación de una RNC capaz de segmentar automáticamente los músculos de la zona lumbar. La RNC tendrá una arquitectura U-Net que se destaca en la segmentación de imágenes médicas [16]. De esta manera se pueden crear etiquetas para cada uno de los músculos de interés y automatizar la evaluación de la calidad muscular en términos de tamaño (volumen y sección transversal) e infiltración de tejido graso (FF).

## 1.2 Objetivos

El objetivo de este trabajo consiste en construir una herramienta basada en algoritmos de aprendizaje automático o *deep learning* que sea capaz de segmentar automáticamente los músculos de la zona lumbar en imágenes de RMN adquiridas con la secuencia Dixon, y así poder evaluar el tamaño y la fracción de tejido graso en un grupo de 118 participantes y obtener valores de referencia.

Para llevar a cabo el objetivo general, se debe cumplir con los siguientes objetivos secundarios:

- Generar un set de datos compuesto por imágenes de resonancia magnética adquiridas con secuencia DIXON y segmentaciones manuales a partir de un grupo muestral de sujetos de distintas características antropométricas.
- Entrenar una red neuronal convolucional con arquitectura U-Net bidimensional (2D) capaz de segmentar correctamente los músculos presentes a la altura de la cuarta vértebra lumbar (L4) en imágenes DIXON.
- Entrenar una red neuronal convolucional con arquitectura U-Net tridimensional (3D) capaz de segmentar correctamente los músculos presentes en la zona lumbar a partir de imágenes DIXON.
- Utilizar los modelos entrenados para calcular el tamaño de los músculos de manera volumétrica y en una sección transversal (CSA, por sus siglas en inglés) a la altura de L4.
- Utilizar los modelos previamente entrenados para calcular de manera automática la fracción de infiltración de grasa volumétrica y en CSA a la altura de L4.

- Generar valores de referencia de volumen, CSA en L4 y FF para sujetos sanos.
- Estudiar la precisión de las métricas de salud muscular de sección transversal como sustitutas de las métricas volumétricas.

### **1.3 Estructura del informe**

Este informe se compone de 5 capítulos, cada uno con múltiples subcapítulos. Este capítulo (capítulo 1) contextualiza el trabajo y presenta los objetivos generales y secundarios del mismo. El capítulo 2 presenta los fundamentos teóricos de la musculatura lumbar y las patologías relacionadas con el deterioro de la calidad muscular, la secuencia Dixon y los algoritmos de *deep learning* utilizados. El tercer capítulo especifica los pasos y las evaluaciones realizadas durante el entrenamiento de los modelos de segmentación automática 2D y 3D. En el cuarto capítulo están presentes los valores de calidad muscular obtenidos al utilizar los modelos previamente entrenados en las imágenes del grupo de estudio. Finalmente, en el último capítulo se concluye los resultados obtenidos.

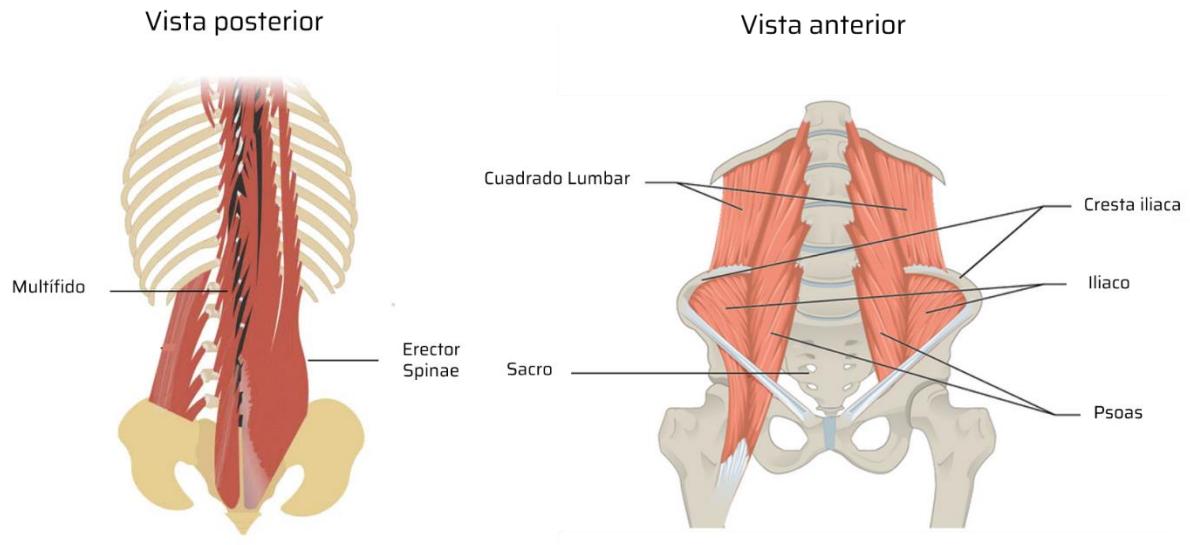
## **2 Marco Teórico**

El siguiente capítulo presenta los fundamentos teóricos necesarios para contextualizar y desarrollar este trabajo.

### **2.1 Calidad de musculatura lumbar como indicador de salud y estilo de vida**

La región lumbar de la columna vertebral está ubicada en la parte baja de la espalda, entre la región torácica y la región sacra. Dentro de esta existen 5 vértebras y se pueden observar 5 pares de músculos que rodean la columna. En la Figura 2.1 se presenta un esquemático de los músculos en cuestión.

- Psoas (P): Comienza en los laterales de la primera vértebra lumbar (L1) y finaliza en el trocánter menor de cada uno de los huesos femorales. Tiene una forma fusiforme y cilíndrica en todo su recorrido alcanzando su máximo espesor a la altura de la unión sacroiliaca. Las funciones biomecánicas consisten principalmente en la flexión de las articulaciones de la cadera y en conjunto con otros músculos ayuda a elevar el tronco desde la posición supina y elevar la pierna hacia el torso [17].
- Iliaco (I): El músculo iliaco origina en la parte superior de la fosa iliaca y finaliza en el trocánter inferior del fémur. Tiene una forma plana y triangular imitando la forma de la estructura en la que está anclado. Su función es estabilizar la pelvis y flexionar la cadera durante la marcha [18].
- Cuadrado Lumbar (CL): Ubicado en la región posterior, lateral e inferior de la columna vertebral e insertado entre la decimosegunda costilla y la cresta iliaca. Su forma es aplanada y cuadrilátera. Su función principal es de flexionar lateralmente la columna vertebral pero también actúa como comunicante de las fuerzas ejercidas por músculos adyacentes [19].
- Multífido (M): Presente en la parte posterior de la columna vertebral desde el cráneo hasta el sacro y encapsulado por una fascia. Tiene la función de estabilizar la columna y tienen un papel importante en el balance y postura [20].
- Erector Spinae (ES): Es un conjunto de tres músculos que recorren toda la columna y que están principalmente desarrollados en la parte lumbar. Tienen un rol importante en el movimiento de la caja torácica y la flexión de la parte superior de la columna vertebral [20].



*Figura 2.1. Esquemático de los músculos de zona lumbar. En el lado izquierdo de la vista posterior se observa el músculo Multífido adyacente a la columna, del lado derecho el, conjunto Erector Spinae. En la vista anterior se presentan los cuadrados lumbares, los músculos Iliacos y Psoas. Imagen adaptada de [21] [22].*

Ya que las funciones que efectúan estos músculos se ven disminuidas con la edad y estilos de vida sedentarios, es consecuente evaluar la calidad muscular de este conjunto como un indicador de calidad de vida.

Múltiples trabajos asocian la infiltración intramuscular de tejido graso y atrofia con el deterioro de los discos vertebrales o la lumbalgia (dolor de espalda baja), una de las principales causas de discapacidad laboral y que afecta al 80% de la población en algún momento de su vida. Este deterioro puede deberse también a la sarcopenia la cual refiere a la pérdida de masa y funcionalidad muscular involuntaria debido a la edad [1] [9] [10] [23] [24].

## 2.2 Secuencia DIXON en imágenes de resonancia magnética

La técnica Dixon es una secuencia de resonancia magnética basada en desplazamiento químico (*chemical shift*) frecuentemente utilizada en la evaluación de tejido musculoesquelético. A diferencia de otras técnicas de supresión de tejido graso, permite crear imágenes de grasa, agua y fracción de grasa después de realizar la adquisición. El principio de esta técnica consiste en, aprovechando que las moléculas de grasa y agua precesan a distintas frecuencias, adquirir imágenes en las cuales las señales de tejido graso y magro se encuentran en fase (IP) y en contra fase (OP). Para ello, por cada tiempo de repetición (TR), se toman dos adquisiciones con tiempos de eco (TE)

específicos (depende de la intensidad de campo magnético). En la Figura 2.2 los tiempos de eco necesarios son 2.3ms y 4.6ms que corresponden al caso de un resonador de 1.5T [25].

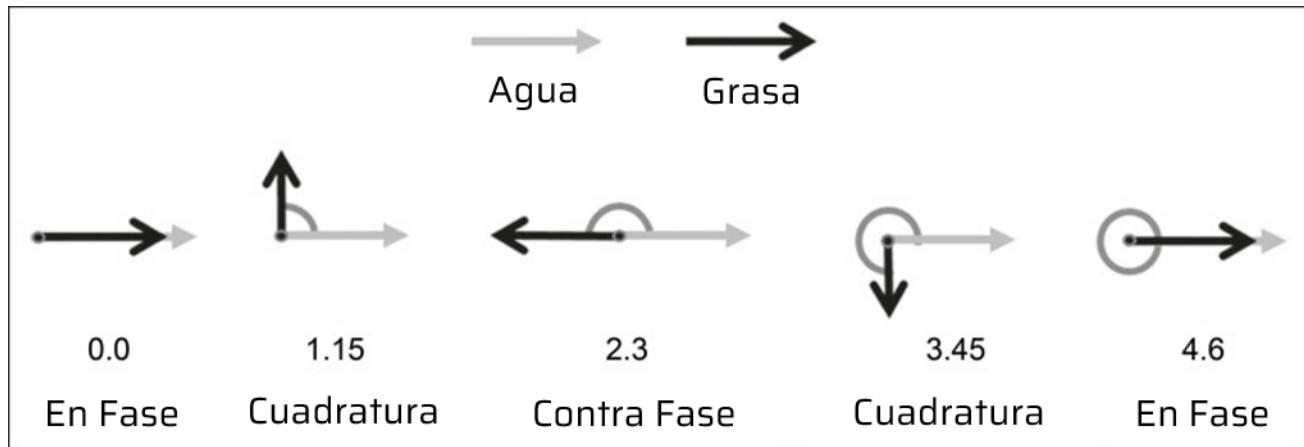


Figura 2.2. Representación de vectores de magnetización para señales de agua en adquisiciones para distintos tiempos de eco (ms) en un scanner de 1.5 T. Figura adaptada de [25].

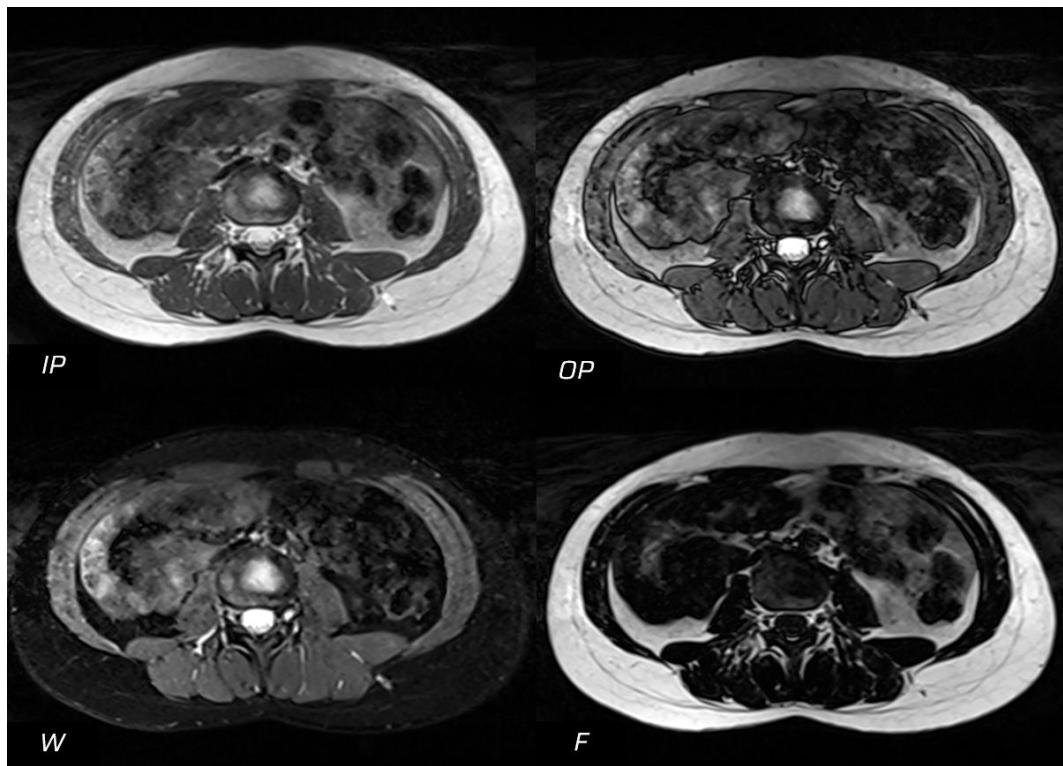


Figura 2.3 Conjunto de imágenes IP, OP, W y F adquiridas mediante una secuencia Dixon axial de RMN a la altura de la vértebra L4.

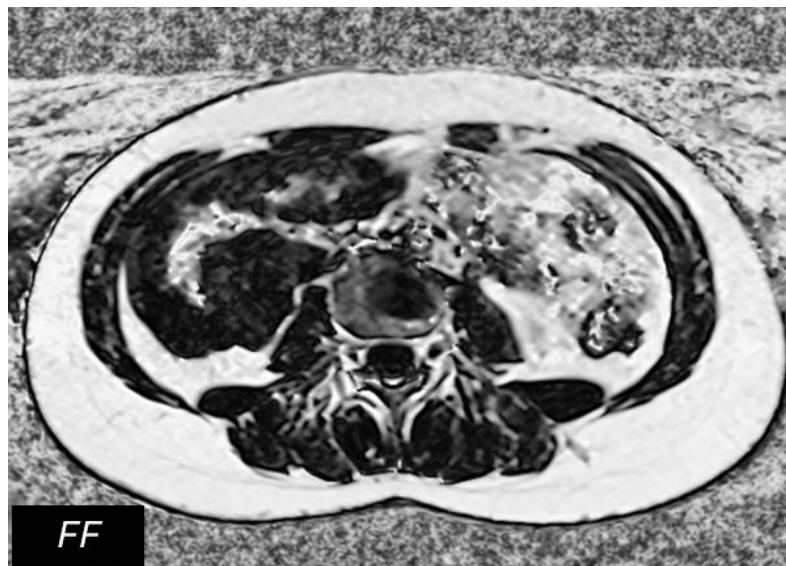
Las imágenes adquiridas inicialmente (presentadas en la figura 2.3) permiten crear las imágenes de señal de agua (W) (que representa el tejido muscular magro) y grasa (F) siguiendo las siguientes relaciones:

$$W = \frac{IP+OP}{2} \quad (1)$$

$$F = \frac{IP-OP}{2} \quad (2)$$

$$FF = \frac{F}{W+F} = \frac{IP-OP}{2 IP} \quad (3)$$

A partir de la fórmula 3 es posible construir una imagen de fracción de tejido graso (FF) (como la que se observa en la Figura 2.4) que contiene en cada voxel un valor entre 0 y 1 representante de la proporción de grasa existente en ese volumen.



*Figura 2.4 Imagen FF construida a partir de las imágenes W y F.*

Si bien se pueden construir imágenes de fracción de tejido graso a partir de las imágenes IP y OP, los resonadores suelen generar las imágenes de grasa y agua automáticamente con algoritmos más sofisticados. En este trabajo emplearemos las imágenes IP para realizar la segmentación, y la de FF para evaluar la calidad muscular del grupo de interés.

### **2.3 Introducción al Deep learning**

“La inteligencia artificial se puede definir un campo de la ciencia e ingeniería dedicado a la comprensión computacional de lo que comúnmente se denomina comportamiento inteligente y a la creación de artefactos que exhiben dicho comportamiento” [26] [27]. Durante mucho tiempo, se consideró que alcanzar una

inteligencia artificial con la capacidad de un humano, se obtendría programando una muy extensa lista de reglas lógicas. No obstante, los resultados de este enfoque terminaron fracasando debido a su gran complejidad, dando origen al *machine learning* o aprendizaje automático, el cual plantea un cambio en el paradigma de la programación.

Convencionalmente, el usuario programa ciertas reglas que aplican sobre una entrada y recibe una salida; en *machine learning*, el usuario brinda una entrada y una salida mientras que el algoritmo “aprende” la regla. En la Figura 2.5 se presenta un esquemático explicando este concepto.

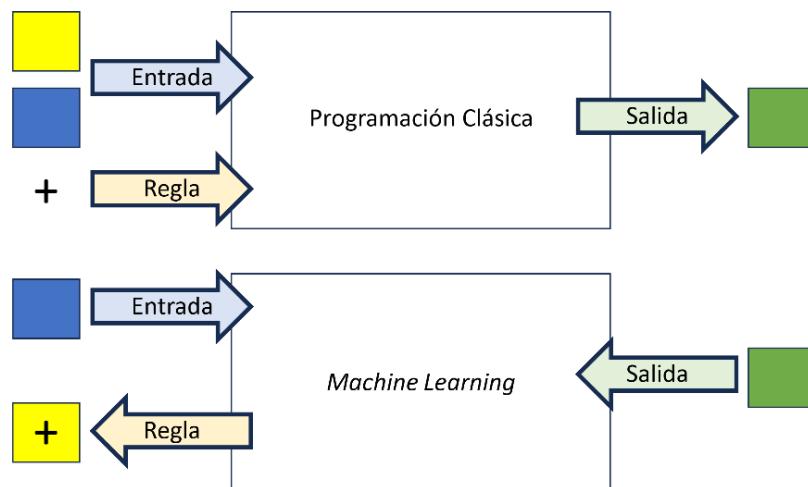


Figura 2.5. Esquemático de programación convencional y su contraste con el machine learning. En la imagen de arriba se muestra una entrada sencilla de dos colores y la orden de sumarlos, a la salida se obtiene el color resultante. A diferencia en la imagen de abajo, se ingresa el color azul como entrada y la salida verde deseada, tal que el algoritmo aprende a adicionar amarillo.

Dentro de *machine learning*, existe el sub-campo del *Deep learning*. Su nombre se debe concepto a “aprender” utilizando algoritmos con múltiples capas sucesivas, cuales transforman los datos de entrada y extraen información cada vez más relevante a medida que se profundiza en las capas del algoritmo [28].

### 2.3.1 Redes neuronales artificiales

Las redes neuronales artificiales (RNA) son la implementación precursora del *deep learning* y a partir de las cuales se desarrollaron los modelos (algoritmo de aprendizaje profundo ya entrenado) que se utilizan en este trabajo. Estos algoritmos se llaman así ya que han estado inspirados en el funcionamiento de las neuronas y su interconexión en el cerebro.

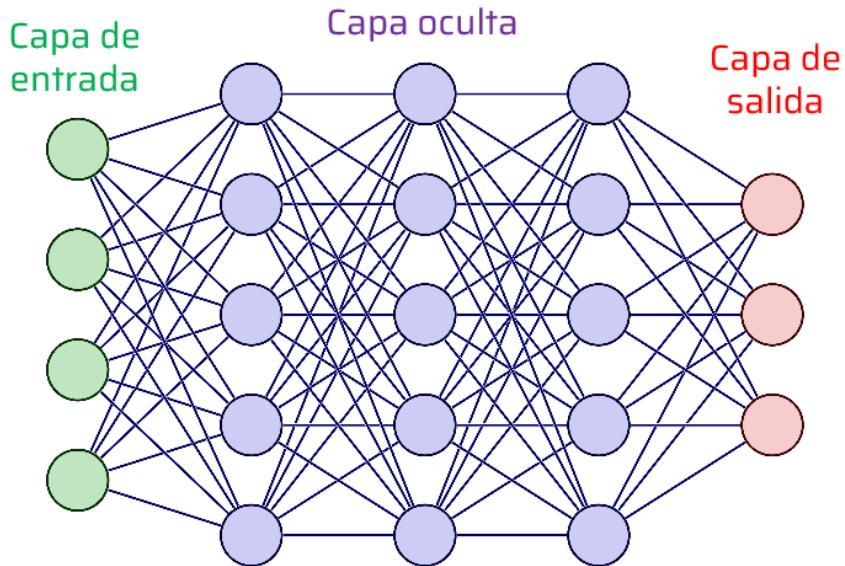
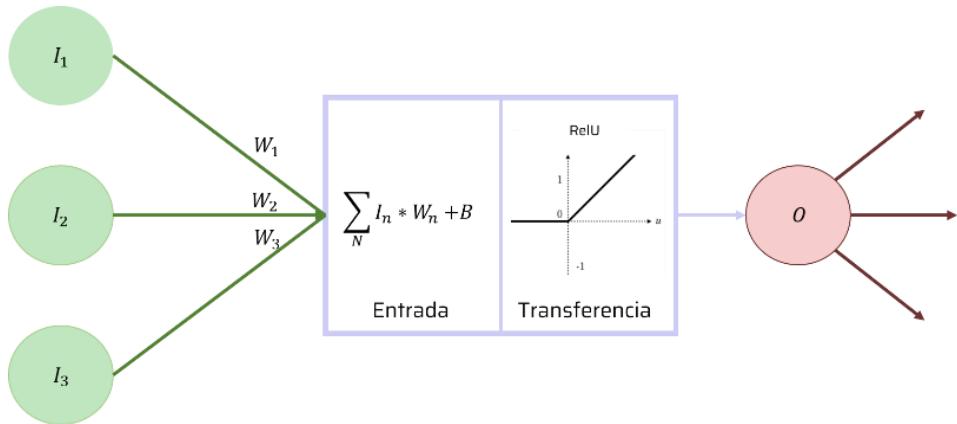


Figura 2.6. Arquitectura esencial de una red neuronal artificial. A la izquierda está la capa de entrada que recibe una única entrada por neurona, en el medio las capas intermedias y a la derecha la capa de salida. Imagen adaptada de [29].

Como se muestra en la Figura 2.6 se puede observar el concepto de capas sucesivas compuestas por neuronas interconectadas. A la primera capa y la última se las denomina capa de entrada y salida, mientras que a todas las demás, capas ocultas.

En una red neuronal artificial (RNA), la neurona procesa valores numéricos y tiene una o múltiples entradas ( $I$ ) (dependiendo en que capa esté ubicada) y una única salida ( $O$ ) que entrará en todas las neuronas de la siguiente capa. Es por eso que a las capas de las RNA originales se las describe como *fully connected* o completamente conectadas.

Estructuralmente, la neurona posee un bloque de entrada y uno de transferencia como muestra la Figura 2.7. El primero tiene la función de ponderar cada una de sus entradas con los pesos  $W$  que forman parte de la neurona, realizar una sumatoria y luego agregarle una constante o *bias* ( $B$ ). Tanto los pesos como el *bias* son parámetros que se actualizan en cada iteración del entrenamiento. El segundo bloque, aplica una función de activación (función no lineal) sobre la resultante de la operación anterior y conecta la salida con todas las neuronas de la capa siguiente.



*Figura 2.7 Esquemático de una neurona de una RNA. Cada neurona recibe entradas a las cuales las multiplica por el peso (W). Luego a la sumatoria de todas las multiplicaciones le agrega un constante (b) y le aplica una función de transferencia. La salida de la neurona se conectarán con todas las neuronas de la siguiente capa.*

### 2.3.2 Función de pérdida y optimizador

Cuando se entrena un modelo de *machine learning*, estos reciben una entrada y una salida objetivo o *ground truth* (GT) tal que luego de múltiples exposiciones, “aprendan” la relación que las vincula. De esta manera, al ingresar un dato desconocido, el modelo realiza la tarea para la cual fue previamente entrenado. Este entrenamiento implica comparar la salida del modelo contra la salida objetivo y modificar los parámetros para reducir su diferencia en la siguiente instancia. Esta comparación y actualización se realiza utilizando una función de pérdida y un optimizador.

La función de pérdida evalúa el desempeño del modelo en cada iteración del entrenamiento comparando la salida del modelo y el GT. La ecuación de la función de pérdida depende de la aplicación y su selección impactará en el entrenamiento de la red neuronal. La misma debe ser derivable ya que su derivada es utilizada para actualizar los parámetros de la red en cada iteración del entrenamiento en la dirección que se minimiza la función objetivo, con el objetivo final de encontrar la configuración óptima de parámetros.

Como todas las operaciones dentro de la neurona son continuas, se puede derivar la función de pérdida con respecto a cada uno de los pesos y constantes presentes en el algoritmo. De esta manera se crea un gradiente, cuya dimensión es igual al número de parámetros, que indica para qué lado crece y decrece la función de pérdida. Este proceso se denomina retro propagación y se utiliza en cada una de las iteraciones del

entrenamiento. A partir de este punto, si desplazamos los valores de los parámetros una determinada distancia o paso en la dirección adecuada, nos acercaríamos al mínimo local o global de la función. Esta distancia se la denomina *learning rate* (LR) y su magnitud debe ser la correcta para no atraparse ni pasar de alto en mínimos locales como se presenta en la Figura 2.8.

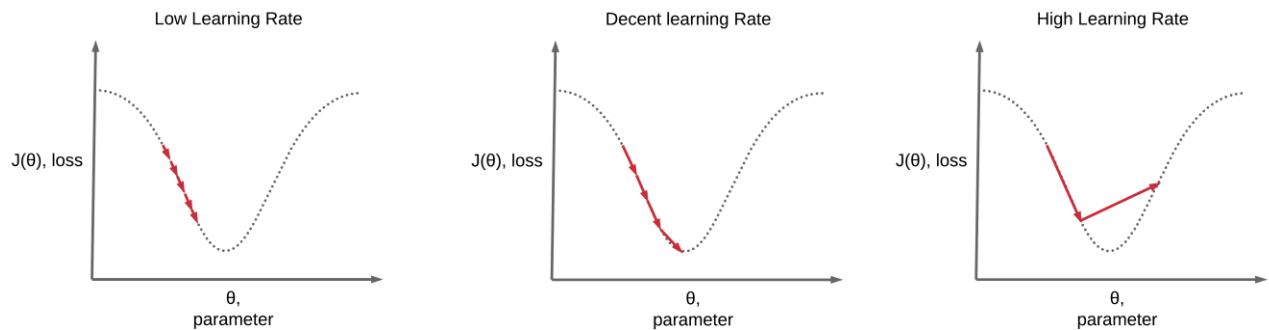


Figura 2.8. Desplazamiento de los valores de la función de pérdida a lo largo del gradiente para distintos learning rates [30].

Todo este proceso de actualizar los parámetros en la dirección opuesta al gradiente utilizando saltos de la magnitud del LR se lo denomina *gradient descent*. Hoy en día se utilizan optimizadores que modifican ligeramente este proceso para obtener mejores resultados o entrenamientos más rápidos.

### 2.3.3 Set de datos

Como se mencionó previamente, para entrenar un modelo de *machine learning* se necesitan datos de entrada y de salida. En el caso de segmentación o clasificación, los datos de salida suelen llamarse etiquetas o *labels* y son la información relevante que se quiere extraer de los datos de entrada.

Todos los datos que se utilizan componen el set de datos y generalmente se separa en tres partes. La primera, el set de entrenamiento, son los datos con los que se va a entrenar el modelo mediante el uso de la función de pérdida y optimizador. Se podría decir que son los datos que el modelo conoce. En cada iteración del entrenamiento, el modelo puede actualizar los parámetros del modelo para cada dato del set o para un lote de estos, denominado *batch*. Cuando se entrena con todos los datos del set de entrenamiento se cumple una época. Dependiendo de la cantidad de datos y el tamaño de la red, los modelos son entrenados durante múltiples épocas.

El segundo set, el de validación, es el conjunto de datos con el que se evalúa el desempeño del modelo mientras se lo entrena. Estos datos no son conocidos por el modelo y los resultados de la función de pérdida o métrica de evaluación son los que determinan cuando se debe finalizar el entrenamiento. Si bien el modelo no reconoce estos datos, suele haber un ligero sesgo ya que el usuario toma medidas durante el entrenamiento para mejorar el desempeño del modelo en este set (por ejemplo, cuando parar el entrenamiento).

Finalmente, se encuentra el set de testeo o evaluación, mucho más pequeño que los anteriores, que se utiliza para evaluar el rendimiento del modelo sin sesgo. En este trabajo, debido a la escasez de datos que poseemos, no se contará con un set de testeo para mejorar la fiabilidad del set de validación y la variedad en el set de entrenamiento.

Dentro del proceso de entrenar una red neuronal, es necesario mencionar el fenómeno de sobreajuste u *overfitting*. Este sucede cuando el modelo que se está entrenando presenta un gran desempeño en el set de entrenamiento mientras que se deteriora en el set de validación. Si se grafica el progreso de la función de perdida en el set de entrenamiento y validación se podrá observar un punto en el cual estas curvas divergen, dando a entender que está ocurriendo el *overfitting* y que el entrenamiento debe finalizar o se debe realizar una modificación en el proceso. Este fenómeno siempre ocurre en algún momento del entrenamiento y puede deberse a que el modelo cuenta con más parámetros de los que necesita, ya que le permite aprender características muy específicas del set de entrenamiento. En la Figura 2.9 se presenta un ejemplo claro de *overfitting* donde los valores de función de perdida para en validación divergen de los valores en entrenamiento.

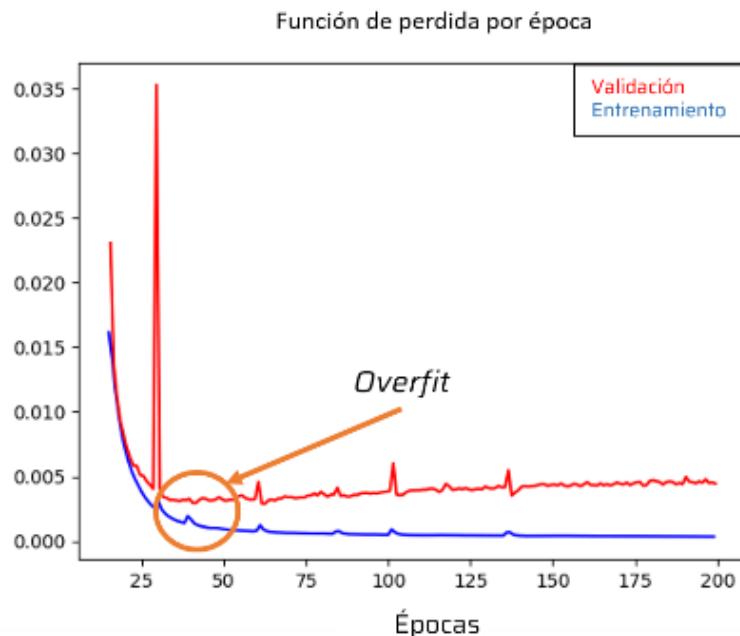


Figura 2.9. Ejemplo de overfit durante el entrenamiento de una RNA donde se indica con un círculo el momento en que la función de pérdida diverge entre el set de entrenamiento y el de validación.

## 2.4 Redes neuronales convolucionales

Si intentásemos utilizar una red neuronal artificial para procesar imágenes, cada píxel de la imagen correspondería a una neurona de la capa de entrada. Si bien este enfoque parece razonable, podemos observar que a medida que aumenta la resolución de la imagen, aumenta la cantidad de neuronas de entrada y junto con ellas, la cantidad de parámetros. Imaginemos que contamos con una imagen *full HD*, con resolución de 1920x1080 píxeles, equivalente a 2 millones de neuronas de entrada, la siguiente capa de la arquitectura tendría 2 millones de pesos en cada una de las neuronas, de esta manera se puede observar cómo crece rápidamente el número de parámetros volviendo inviable su implementación.

Por esta razón, para el procesamiento de imágenes se emplean redes neuronales convolucionales (RNC), propuestas originalmente por Yann LeCun en la década del 80 [31] pero adoptadas masivamente recién en la última década. A diferencia de las redes anteriores, la neurona está compuesta por un filtro de convolución sobre la imagen de entrada, una constante y una función de activación como se presenta en la Figura 2.10.

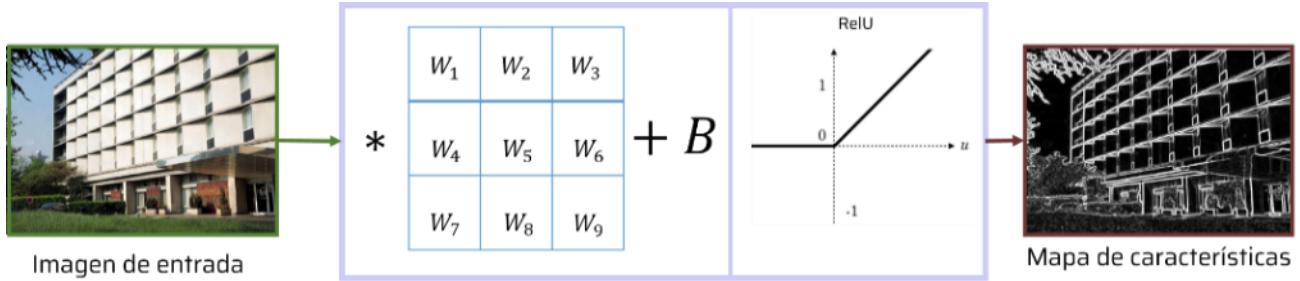


Figura 2.10. Unidad básica de red neuronal convolucional. La imagen de entrada es convolucionada por un filtro de dimensión NxN cuyos valores son los pesos del modelo a la cual se le agrega un bias y luego pasa por la función de activación para obtener distintos mapas de características.

Por lo tanto, cada una de las capas convolucionales estará compuesta por múltiples neuronas que procesarán la imagen de entrada y entregarán una imagen de información relevante a la salida. Las imágenes extraídas por cada una de las neuronas se las denomina mapas de características o *feature maps* y los parámetros de estos modelos son los pesos de cada filtro (W) y las constantes (B).

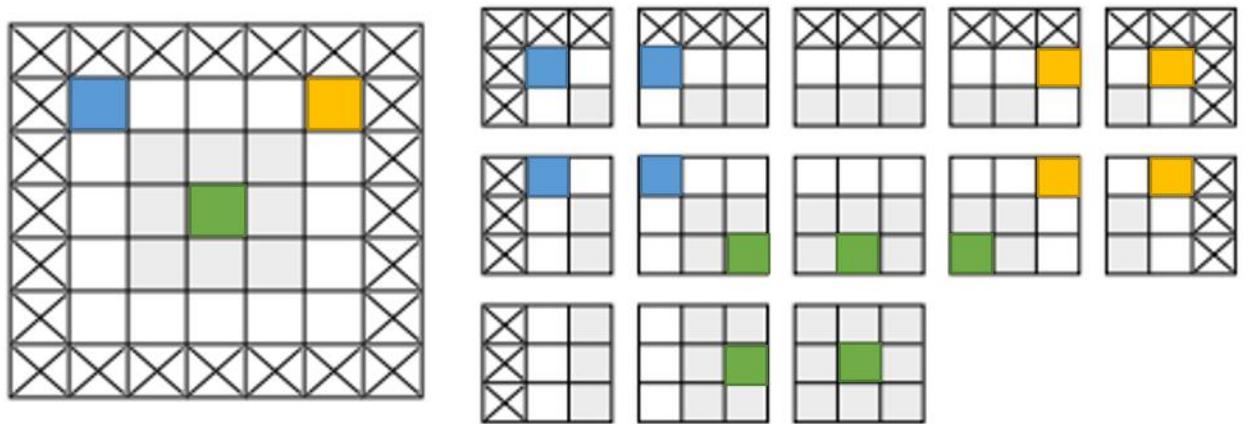
#### 2.4.1 Convolución

La convolución es la operación matemática que realizan los filtros sobre la imagen al procesarla. Este proceso consiste en deslizar el filtro de tamaño NxN sobre todos los fragmentos NxN posibles de la imagen, y para cada instancia realizar la sumatoria de la multiplicación punto por punto entre los coeficientes del filtro y el valor de los vértices del fragmento. Este proceso se puede observar en la Figura 2.11 donde se ejemplifica la operación punto por punto de un fragmento con un filtro.



Figura 2.11 Ejemplo de convolución entre un filtro 3x3 y una imagen 4x4. Se observa la multiplicación punto por punto entre el filtro y el fragmento centrado en el píxel de valor 100 [32].

El tamaño de la imagen filtrada depende del tamaño de la imagen de entrada, el tamaño del filtro y dos factores adicionales, el paso o *stride* y el zero *padding*. El paso hace referencia a cuantos pixeles se desplaza el filtro luego de cada convolución, de esta manera si el paso es mayor, tendremos menos cantidad de operaciones. En cuanto al *zero padding*, es una técnica que permite tener una imagen de salida del mismo tamaño que la imagen de entrada. Consiste en agregar líneas de Os en los bordes de la imagen para que, al convolucionar el filtro, no se pierda información en dicha región.



*Figura 2.12 Ejemplo de padding para una imagen 5x5 que se convoluciona con un filtro 3x3. Las cruces representan los Os agregados mientras que los pixeles de colores permiten visualizar mejor los fragmentos 3x3. Figura adaptada de [28].*

En la figura 2.12 se puede observar un ejemplo de *padding* para una imagen 5x5 que es convolucionada por un filtro 3x3 con paso de 1. Cada recorte 3x3 devuelve un valor numérico que al concatenarlos formarían una imagen 5x5. Para resumir, el tamaño cada una de las dimensiones de una imagen filtrada seguiría la siguiente formula:

$$O = \frac{I-K+2P}{S} + 1 \quad (4)$$

donde  $O$  es el tamaño de la salida en una dimensión,  $I$  es el tamaño de entrada,  $K$  el tamaño del filtro,  $P$  el tamaño del *Padding* y  $S$  el paso.

#### 2.4.2 Max Pooling

En las RNC se suele utilizar la operación *Max Pooling* la cual permite seleccionar la información más relevante de la imagen de entrada y luego condensarla, descartando información innecesaria. Este proceso consiste en aplicar un filtro NxN que extraiga el valor máximo de cada fragmento NxN de la imagen. Generalmente, se utilizan filtros de

tamaño  $2 \times 2$ , que reducen el tamaño de la imagen de entrada a la mitad en ambas dimensiones. En la Figura 2.13 se muestra un ejemplo de *Max Pooling* donde se puede observar cómo se condensa la información de la imagen de entrada.

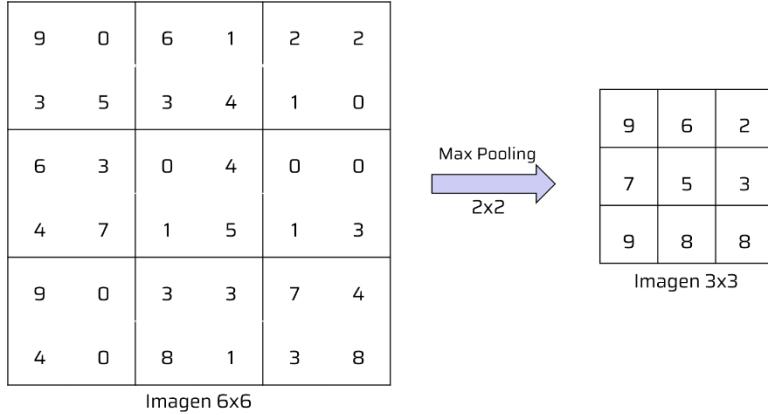


Figura 2.13 Implementación de *Max pooling* en una imagen  $6 \times 6$  en la cual se extrae el valor máximo de cada fragmento  $2 \times 2$ . El tamaño de las dimensiones de la imagen de salida resulta la mitad de las de entrada.

#### 2.4.3 Redes neuronales convolucionales de clasificación

La primera red neuronal convolucional fue implementada en 1998 y se le otorgó el nombre de *LeNet* y su función era clasificar imágenes de  $32 \times 32$  píxeles de letras y números escritos a mano [31]. Si bien, esta arquitectura no es relevante para este trabajo, sirve para explicar la utilidad de cada uno de los componentes.

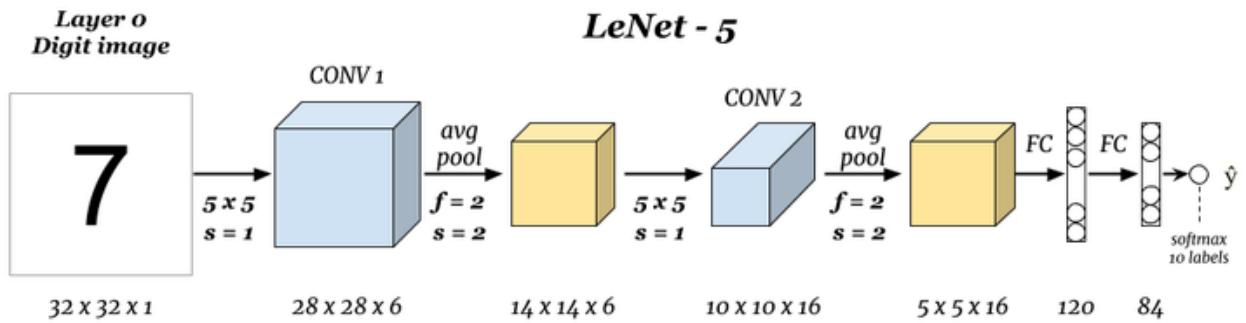


Figura 2.14 Arquitectura *LeNet* compuesta por capas convolucionales y de pooling de valor medio intercaladas seguidas por capas fully connected de una RNA [33].

Esta arquitectura consta de dos capas convolucionales, dos capas de *pooling* y dos capas ocultas completamente conectadas. La primera capa convolucional contiene 6

filtros de tamaño 5x5 con paso de 1 y sin *padding* que se convolucionan con una imagen de entrada de 32x32 con un único canal. Debido a que cada convolución genera una imagen de salida, la salida de la primera capa consiste en 6 mapas de características de 28x28. Luego de esta capa se aplica una capa de *pooling de valor medio* con filtros 2x2 que reduce la resolución de las 6 imágenes a 14x14. Este proceso se repite con una segunda capa convolucional y *pooling* para luego conectarse con una red neuronal artificial.

La lógica de esta arquitectura consiste en utilizar capas convolucionales para extraer características relevantes de la imagen y reducir su resolución múltiples veces. De esta manera, al final de las capas contamos con imágenes muy pequeñas, pero con información muy relevante de la imagen original permitiendo realizar una clasificación con una RNA convencional. Este tipo de redes se volvieron muy populares con las arquitecturas *AlexNet* y *Resnet* ya que tenían un rendimiento ampliamente superior en comparación con las RNA tradicionales.

## 2.5 Arquitectura U-Net

En este trabajo utilizaremos una red neuronal convolucional de segmentación de imágenes. La segmentación consiste en generar etiquetas que indiquen regiones de interés dentro de la imagen de entrada. A diferencia de las redes neuronales convolucionales de clasificación de la sección anterior, las redes de segmentación no tienen capas completamente conectadas debido a que la salida del modelo es una imagen del mismo tamaño que la de entrada.

Dentro de este tipo de redes, la arquitectura *U-Net*, presentada en el MICCAI 2015 [16], se ha destacado en esta tarea mostrando resultados muy superiores a los algoritmos del momento. Hoy en día este algoritmo y sus modificaciones son el estándar en la segmentación de imágenes médicas.

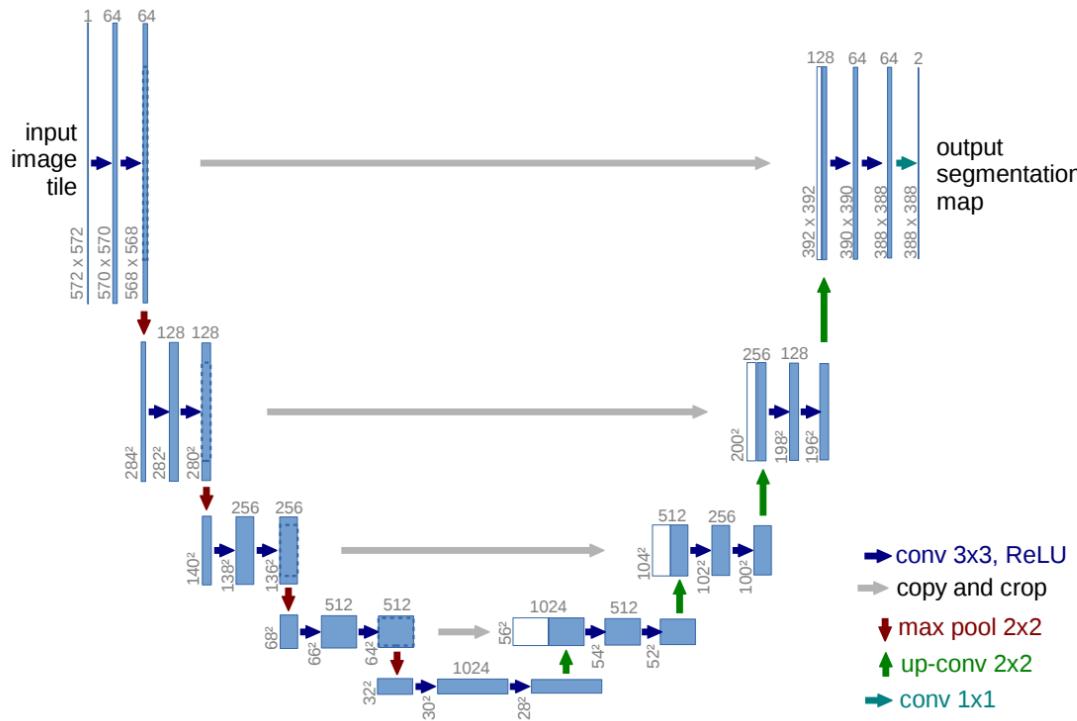


Figura 2.15 Arquitectura U-Net. Su nombre se debe a la forma del esquemático donde la parte decreciente representa la compresión de la imagen y la ascendente, la descompresión [16].

En la Figura 2.15 se presenta el esquemático original que caracteriza el funcionamiento de la *U-Net* donde se observa la forma de “U” que le da origen a su nombre. Como entrada del modelo, se cuenta con una imagen blanco y negro y a la salida, una segmentación de una o múltiples clases.

Esta arquitectura tiene dos caminos bien distinguidos, uno de compresión o *encoder* y otro de descompresión o *decoder*. El primero, la parte descendente de la red, posee capas de bloques de dos capas convolucionales consecutivas y una de *Max Pooling*. Al igual que en las RNC anteriores, las capas convolucionales extraen información de la imagen de entrada creando mapas de características mientras que la capa de *Max pooling* divide la resolución de estos a la mitad. De esta manera al profundizar en la red se obtiene información cada vez más relevante y compacta de la imagen original.

El segundo camino, la descompresión, utiliza bloques compuesto por dos capas convolucionales y capas de *upsample*. Estas segundas capas, funcionan de manera inversa al *Max Pooling*, por cada pixel de una imagen de entrada, se forman 4 pixeles

iguales tal que la resolución se duplica en ambas dimensiones (proceso presentado en Figura 2.16). De esta manera, a medida que se asciende en la red se recupera el tamaño de la imagen original.

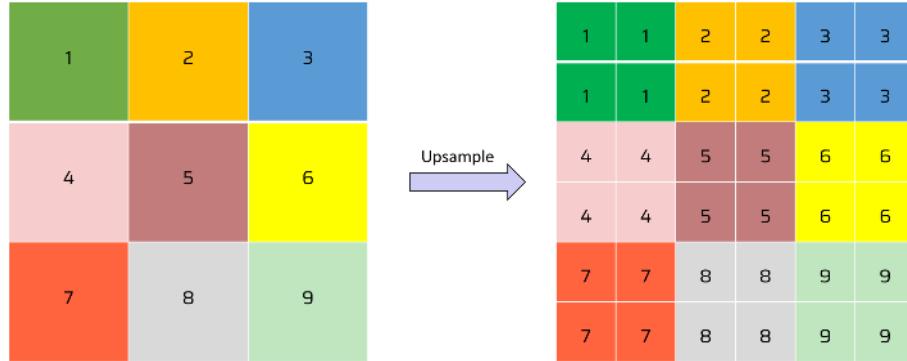


Figura 2.16 Ejemplo de Upsample. Cada recuadro de la imagen de la izquierda representa un píxel distinto el cual al aplicarle la operación de upsample se duplica en ambas dimensiones.

La característica principal de esta arquitectura se observa en las líneas grises presentes en la figura 2.15, llamadas *skip connections*. Estas representan un proceso de concatenación que toma los mapas de características extraídos en cada uno de los bloques de la codificación y los concatena con la entrada de los bloques de decodificación. De esta manera, a medida que se “asciende” en el algoritmo, se recupera información importante de la imagen que, de otra manera, se perdería en la codificación.

En este trabajo entrenaremos modelos de segmentación de imágenes bidimensionales y tridimensionales cuyas arquitecturas tienen leves modificaciones respecto de la arquitectura *U-net* original. Para el entrenamiento, los datos de entrada de los modelos serán imágenes Dixon *In-Phase* de la zona lumbar y la salida objetivo serán segmentaciones manuales de los músculos de interés. Una vez entrenados, los modelos se utilizarán para evaluar la calidad muscular en imágenes Dixon *In-Phase* de sujetos físicamente activos y sedentarios para las cuales el modelo extraerá etiquetas de 6 u 8 músculos dependiendo si se trata de un modelo bidimensional o tridimensional.

## **3 Redes neuronales convolucionales para la segmentación automática de los músculos lumbares**

En este capítulo se detallarán todas las técnicas y recursos utilizados, los pasos ejecutados y las evaluaciones realizadas para poder construir modelos de segmentación automática acordes con los objetivos presentados. Este capítulo abarca desde la adquisición de datos hasta la ejecución de los modelos entrenados en imágenes de prueba.

### **3.1 Datos**

Este trabajo cuenta con imágenes de resonancia magnética e información antropométrica de 118 sujetos que formaron parte de un estudio para analizar el impacto de la actividad física en la salud muscular en el Reino Unido. Ochenta y ocho sujetos corresponden a personas sedentarias que estaban por empezar un programa de ejercicio físico del NHS (National Health Service) (Reino Unido) y 30 a un grupo de ciclistas aficionados que realizan actividad física regularmente.

#### **3.1.1 Adquisición de datos**

La adquisición de las imágenes de resonancia magnética se realizó en Londres (Reino Unido) en un resonador magnético cerrado *Siemens Magnetom Vida* (Erlangen, Alemania) de 3T utilizando una bobina de cuerpo entero. El protocolo de adquisición consistió en una secuencia Dixon cubriendo la zona lumbar y sacro ilíaca para poder visualizar el contraste entre el tejido muscular y el tejido graso. El espesor de los cortes fue de 1.5 mm, el espaciado entre estos de 1.95 mm, tiempo de eco (TE) 45 ms y se empleó un campo de visión (FOV) que cubre el cuerpo del sujeto desde L1, donde comienza el psoas, hasta el trocánter menor donde finaliza los músculos iliaco y psoas. El tamaño de voxel fue de 0.47 x 0.47 x 1.95 mm<sup>3</sup>.

#### **3.1.2 Construcción del set de datos**

Como se mencionó en el capítulo anterior, para entrenar un modelo del tipo U-Net se necesitan datos de entrada compuestos por las imágenes IP de RMN adquiridas con secuencia Dixon y las etiquetas o *labels* correspondientes a la salida deseada del modelo (GT). Las etiquetas estarán compuestas de segmentaciones manuales de los músculos Psoas (P), Cuadrados lumbares (CL), Iliacos (I), Erector Spinae (ES) y Multífidos (M). Estos últimos dos los agruparemos como un conjunto (ES+M) debido a que son músculos adyacentes con funciones anatómicas similares y que se encuentran rodeados por la fascia toracolumbar [2].

Las imágenes y etiquetas correspondientes a los modelos 2D se ubican a la altura de la cuarta vértebra lumbar (L4) debido a que el área de los músculos es mayor y suele

ser una zona relevante para evaluar la infiltración de tejido graso [9] [23]. En el caso de las imágenes y etiquetas del modelo 3D, abarcan la región entre la primera vértebra lumbar hasta el trocánter menor del fémur.

### 3.1.2.1 Segmentación manual

Las etiquetas fueron creadas a partir de las imágenes presentadas en el punto 3.1 respetando la misma proporción de sujetos activos y sedentarios, y de hombres y mujeres. Para ello se utilizó el programa de visualización y procesamiento de imágenes Simpleware ScanIP, empleando las herramientas de generación de máscaras y segmentación manual. El criterio de segmentación consistió en abarcar toda la zona de cada uno de los músculos incluyendo la grasa intramuscular (diferente a la intermuscular que se encuentra entre las fascias de los músculos) para que el modelo sea capaz de identificar la fracción de grasa del músculo bajo análisis. Por otro lado, las segmentaciones fueron realizadas en las imágenes *IP* de la secuencia Dixon debido a su similitud con las imágenes adquiridas con la secuencia *T1-weighted*.

En el caso de las etiquetas 2D, se segmentaron manualmente en un plano axial en la altura L4 los pares de músculos P, CL y los conjuntos ES+M. A partir de estas segmentaciones se pudieron crear 58 conjuntos de imágenes y etiquetas fundamentales para el entrenamiento y evaluación del modelo bidimensional. En la Figura 3.1 se presenta una segmentación manual en altura L4 de los músculos involucrados.

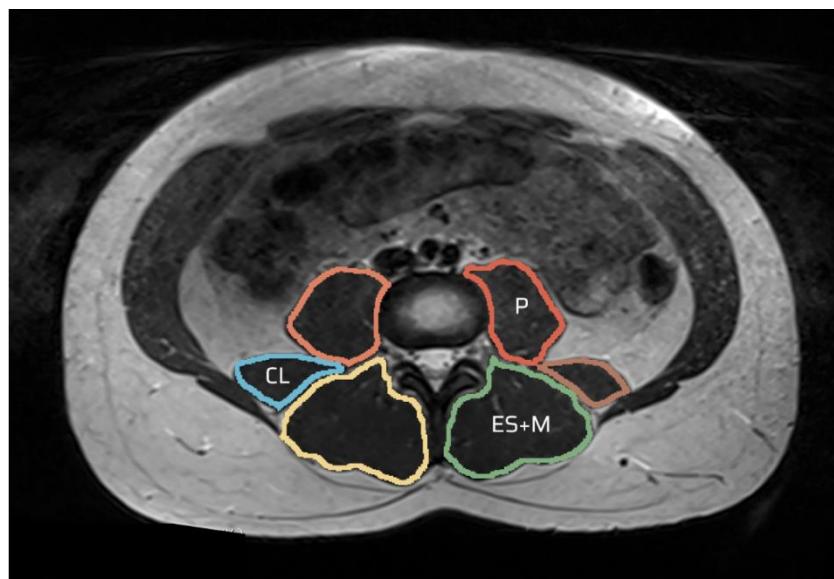
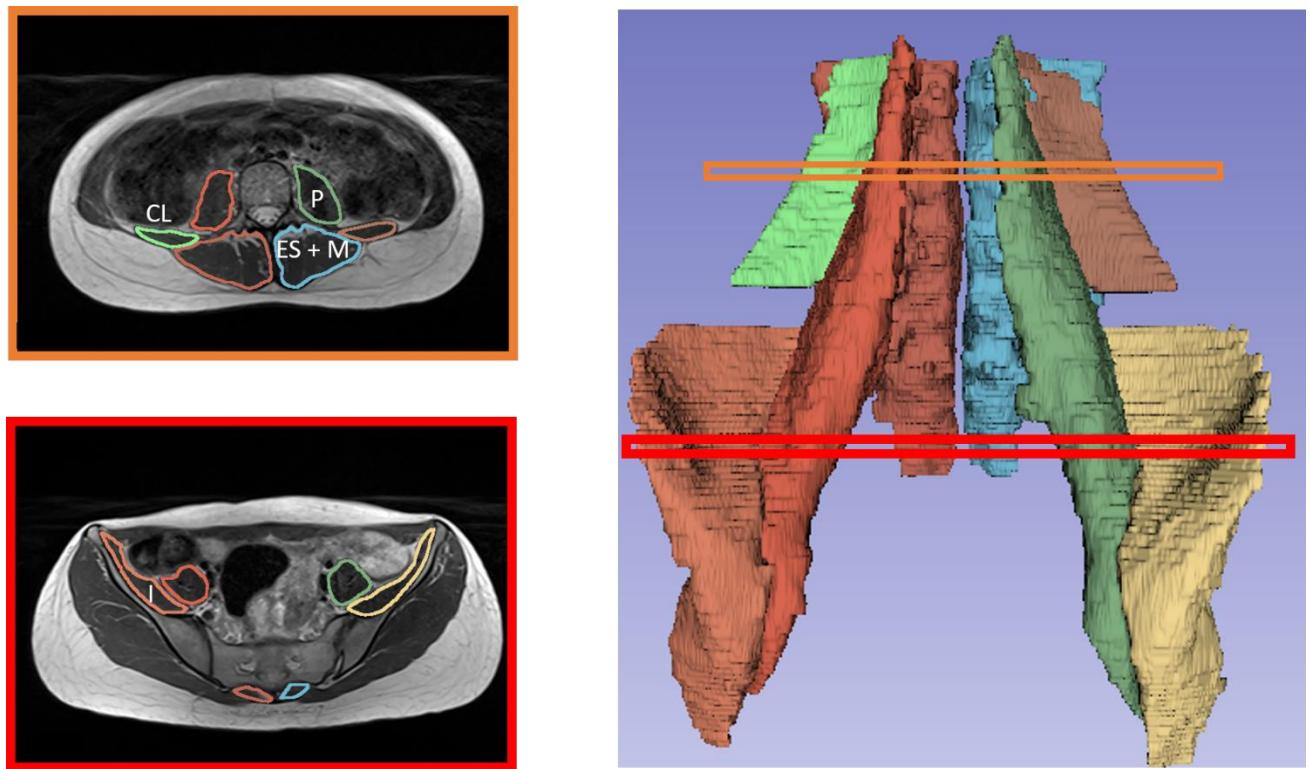


Figura 3.1. Segmentación manual de músculos de zona lumbar en un corte axial a la altura de L4.

Para el caso de los conjuntos 3D, el trabajo fue más extenso ya que cada imagen consistía en 150 cortes en promedio y 4 pares de músculos de interés, los 3 anteriores más los iliacos. A diferencia del set de datos 2D, la morfología muscular es mucho más compleja y varia drásticamente en los puntos de inserción. Dentro de los desafíos encontrados se encuentra la segmentación del músculo psoas e ilíaco a la altura sacrolumbar, donde estos músculos forman un conjunto difícil de discriminar y sus fascias pierden resolución dependiendo de la intensidad de imagen. Se pudieron obtener conjuntos de imágenes y etiquetas provenientes de 12 mujeres (8 sedentarias y 4 ciclistas) y 14 hombres (9 sedentarios y 5 ciclistas). En la Figura 3.2 se puede observar una segmentación manual 3D de los músculos de zona lumbar de interés.



*Figura 3.2. Izquierda: etiquetas manualmente segmentadas de los músculos de la zona lumbar en dos cortes axiales a la altura de L2 y la cresta ilíaca. Derecha: modelo 3D generado a partir de la segmentación manual 3D, donde se muestran las alturas de los dos cortes visualizados a la izquierda.*

### 3.1.2.2 Aumentado de datos

Usualmente los modelos de segmentación o clasificación de imágenes suelen entrenarse utilizando repositorios públicos con una gran cantidad de datos. Esto mejora la capacidad de generalizar de los modelos cuando se los utiliza con datos o imágenes nuevas. Dado que no se encuentran imágenes y segmentaciones de zona lumbar adquiridas con la secuencia Dixon en repositorios públicos, y que realizar segmentaciones

manuales lleva mucho tiempo, se implementó un aumentado de las imágenes y sus respectivas segmentaciones manuales.

El aumentado de datos consiste en generar imágenes nuevas a partir de modificaciones y deformaciones espaciales (o en intensidad) de nuestro set de datos original. Para ello se utilizaron herramientas de la biblioteca de código libre para procesamiento de imágenes SimpleITK en Python. Se utilizaron dos estrategias distintas para el aumentado de las imágenes y sus segmentaciones manuales, una mediante deformaciones lineales y la otra mediante transformaciones no lineales. A continuación, se especifica paso a paso el aumentado realizado.

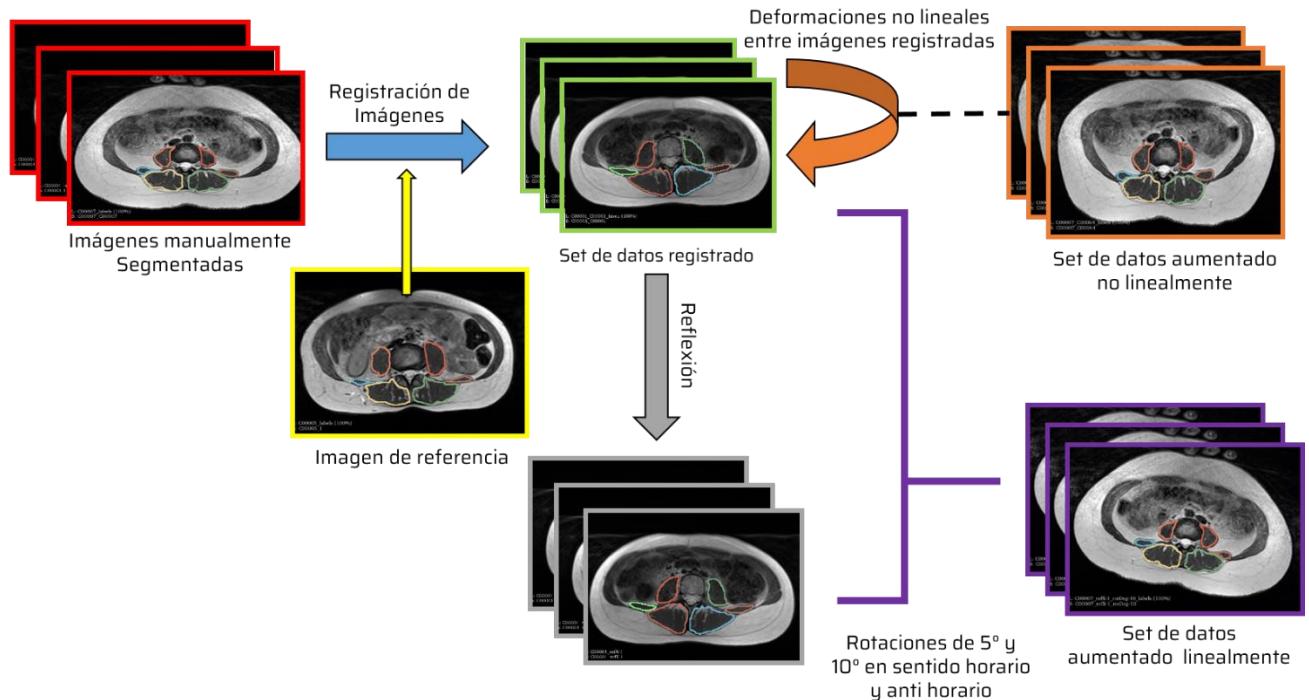
Previo al aumentado, se acondicionó el set de datos crudo para lograr una homogeneidad entre las imágenes en términos de tamaño de las imágenes y de los véxeles. Para ello se recurrió a aplicar un registro rígido de imágenes, el cual consiste en aplicar transformaciones de escalado, traslación y rotación para realinear todas las imágenes del set de datos a una elegida como referencia. La imagen de referencia del set de datos 2D fue elegida mediante una inspección visual en la que se buscó un sujeto que tuviese un tamaño normal y su imagen estuviese centrada y con los músculos bien definidos. Para el caso del set de datos 3D, se escogió una imagen que cumpliese con este criterio pero que también tuviese la cantidad de cortes más repetida en todo el conjunto. De esta manera se homogenizó la cantidad de cortes que componía cada una en 145 por imagen. Adicionalmente, se redujo el tamaño de las imágenes a un cuarto transformando la resolución inicial de 800 x 640 x 145 a 400 x 320 x 145. Esta medida fue fundamental debido a que el tamaño de las imágenes impacta directamente en la demanda de memoria RAM necesaria para entrenar las redes neuronales de este trabajo.

Una vez registradas las imágenes, se realizó el primer aumentado, ejecutando transformaciones lineales compuestas de rotaciones y reflexiones. Debido a que los músculos de interés son simétricos, la reflexión permite generar imágenes virtuales muy cercanas a la anatomía normal. Por otro lado, las rotaciones simulan malos posicionamientos del paciente cuando se adquieren las imágenes. De esta manera se generaron imágenes virtuales con rotaciones entre -10° y 10° con pasos de 5°, repitiendo las mismas operaciones para las imágenes reflejadas. De esta manera, por cada imagen manualmente segmentada se generan 10 virtuales.

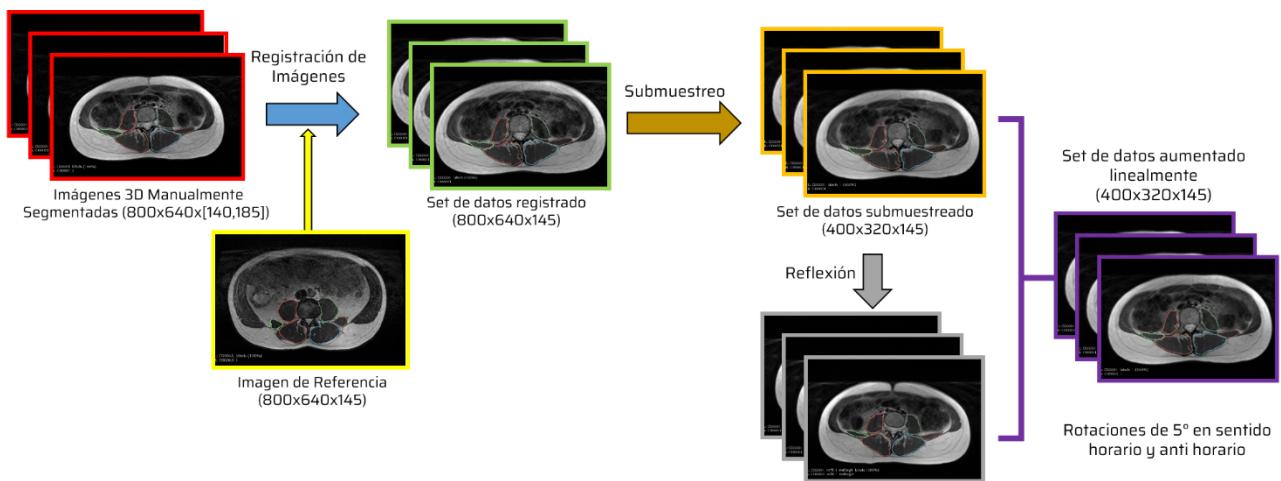
Por último, se realizaron los aumentados por transformaciones no lineales utilizando la librería Simple Elastix de python. En este tipo de aumentado, cada imagen se registró de forma no lineal con cada una de las imágenes restantes del set de datos homogeneizado. En este proceso, la imagen a ser aumentada se deforma mediante

transformaciones no lineales *BSpline* para asemejarse a una imagen de referencia. Los parámetros de deformación están presentes en un archivo de texto que se utiliza en conjunto con las funciones de la librería y se optimizaron empíricamente con el fin de obtener deformaciones anatómicamente similares a la realidad. De esta manera se obtuvo un algoritmo que utiliza tres niveles de resolución, la correlación cruzada normalizada como métrica de similitud, *adaptive stochastic gradient descent* como optimizador, 2000 muestras espaciales por iteración y 1000 iteraciones. Este aumentado se realizó utilizando cada una de las imágenes del set de entrenamiento como imagen objetivo y siendo deformada por todas las demás. De esta manera, si se cuenta con una cantidad  $N$  de imágenes se pueden generar  $N * (N-1)$  imágenes aumentadas. Para finalizar, las transformaciones (tanto lineales como no lineales) obtenidas se aplican a las imágenes con las etiquetas para transferir las segmentaciones manuales al espacio de las imágenes aumentadas.

Para el set de datos 2D, se realizó el aumentado lineal y no lineal como muestra la Figura 3.3 mientras que el set de datos 3D fue aumentado únicamente de manera lineal y con solo un ángulo de rotación como indica la Figura 3.4.



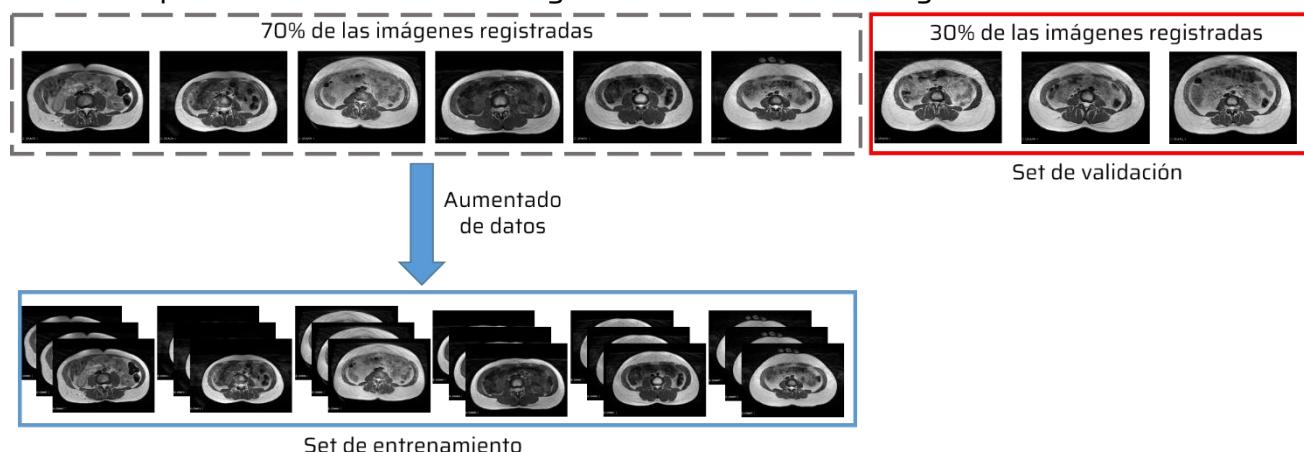
*Figura 3.3 Registración y aumentado para set de datos 2D. En primera instancia el set de datos es registrado utilizando una imagen de referencia tal que todas tengan el mismo tamaño de imagen y voxel. Luego el set registrado es aumentado linealmente utilizando reflexiones y rotaciones, y aumentado no linealmente utilizando como referencia las mismas imágenes del set.*



*Figura 3.4 Registración y aumentado lineal para set de datos 3D. Las imágenes de distinto tamaño son registradas con respecto a una de referencia, homogeneizando el tamaño de todas. Luego el set de datos es submuestreado para reducir el tamaño a un cuarto y finalmente es aumentado linealmente con reflexiones y rotaciones de 5° en sentido horario y anti horario.*

### **3.1.3 División del set de datos en entrenamiento y validación**

En cuanto a la división del set de datos entre set de entrenamiento y set de validación, todos los modelos fueron entrenados con una proporción 70/30% a nivel paciente. Esta aclaración se debe a que el aumentado de datos se implementó únicamente en el set de entrenamiento, de manera que el set de validación estuviese compuesto solo de imágenes reales (proceso representado en Figura 3.5). De esta manera el set de datos 2D se compone de 2000 imágenes de entrenamiento y 18 de validación mientras que el set 3D tiene 108 imágenes de entrenamiento y 8 de validación.



*Figura 3.5. División del set de datos entre entrenamiento y validación. A partir de los datos registrados, se separó un 30% como set de validación mientras que con el otro 70% se realizó un aumentado para construir el set de entrenamiento.*

## 3.2 Componentes esenciales del entrenamiento

Empleando los sets de datos 2D y 3D, se entrenaron sus respectivas RNC empleando las mismas métricas de evaluación y configuraciones de entrenamiento para ambos casos. Las configuraciones empleadas son desarrolladas en esta sección.

### 3.2.1 Métricas de evaluación

Como métrica principal para evaluar el desempeño de los modelos, se optó por el coeficiente Dice, que se considera el standard a la hora de evaluar el desempeño de algoritmos de segmentación de imágenes [34]. Este cuantifica el solapamiento entre dos muestras binarias, en este caso entre la segmentación automática y la manual. La misma está definida como:

$$DSC = \frac{2Tp}{2Tp + Fp + Fn} \quad (5)$$

donde  $T_p$  hace referencia a los verdaderos positivos,  $F_p$  a los falsos positivos y  $F_n$  a los falsos negativos. Como se puede observar el rango de valores es de 0 a 1 y el puntaje penaliza los errores por sobre y subestimaciones.

Como métricas secundarias, se utilizó la sensibilidad y la precisión para evaluar aspectos característicos de los modelos:

$$Sens = \frac{Tp}{Tp + Fn} \quad (6)$$

$$Prec = \frac{Tp}{Tp + Fp} \quad (7)$$

No se consideró la especificidad como una métrica debido a que el tamaño de las imágenes es muy superior al tamaño de las segmentaciones y consecuentemente los resultados son muy similares a 1.

### 3.2.2 Función de pérdida y optimizador

Como función de perdida se seleccionó la entropía cruzada binaria (*BCE, por sus siglas en inglés*) que evalúa las diferencias entre dos distribuciones de probabilidades y es frecuentemente usada para entrenar modelos de segmentación de imágenes [35] [36]. En estos casos, la etiqueta de la segmentación manual objetivo (GT), tiene una

distribución binaria tal que es 0 o 1 en cada voxel. En cambio, la salida del modelo consiste en un mapa de probabilidades, donde se obtiene la probabilidad (entre 0 y 1) de que cierto voxel pertenezca a la etiqueta a segmentar. La BCE penaliza las estimaciones erróneas del modelo como indica la siguiente ecuación:

$$L(X, Y) = -[Y * \log(X) + (1 - Y) * \log(1 - X)] \quad (8)$$

siendo  $X$  la estimación del modelo en un determinado píxel e  $Y$  el valor binario de la etiqueta en esa misma coordenada. Para minimizar la función de perdida, se utilizó el optimizador ADAM con un *learning rate* de 0.0001, que es uno de los algoritmos de optimización más empleados en *deep learning* [37] [38].

### **3.3 Red neuronal convolucional 2D**

El entrenamiento del modelo 2D de segmentación de los músculos de la zona lumbar se realizó de manera local en una computadora personal con una placa de video *Nvidia RTX 2060* compatible con CUDA de 6 Gb de memoria RAM y utilizando la librería Pytorch de Python. Primero, se entrenó una RNC para la segmentación automática del grupo ES+M, y luego para los tres músculos estudiados a la altura de L4.

#### **3.3.1 Mono clase**

Teniendo en cuenta lo mencionado previamente, primero se desarrolló un modelo simple capaz de segmentar un único músculo, más específicamente el grupo ES+M izquierdo. Esta selección se debe a que el tamaño de este es muy superior a los otros tipos de músculo a la altura L4.

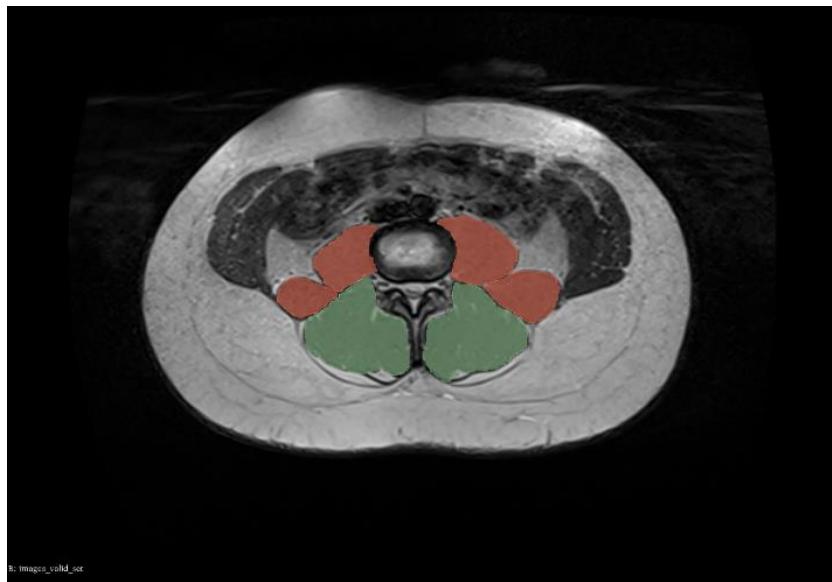


Figura 3.6. Corte axial en altura L4 En verde se muestran las etiquetas correspondientes a los conjuntos ES+M mientras que en rojo los CL y P. Se observa la diferencia de tamaño a favor de los ES+M.

Como se puede ver en la Figura 3.6, el tamaño del conjunto ES+M es considerablemente mayor a los otros músculos involucrados y está rodeado de tejido graso subcutáneo, facilitando la tarea de aprendizaje de la RNC para delimitar el contorno del músculo.

Una vez seleccionado el tejido a segmentar, se utilizó la configuración original de la arquitectura U-Net, consistente de 4 bloques codificadores y decodificadores, 64 filtros bidimensionales de 3x3 con *zero padding* de 1 y *max pooling* de 2x2.

Para entrenar la red con una optimización de la demanda de memoria, se utilizó un tamaño de batch reducido de sólo 3 imágenes, y se empleó la herramienta *Automatic Mixed Precision (AMP)* que detecta automáticamente los parámetros del entrenamiento que pueden ser implementados en punto flotante de 16 bits (en vez de 32) sin perjudicar el entrenamiento. Estas estrategias resultaron en una reducción en el uso de memoria, necesitando menos de 6 GB de RAM, y una aceleración en los tiempos de entrenamiento [39]. A partir de estas configuraciones, los resultados de este primer modelose pueden observar en la Figura 3.6.

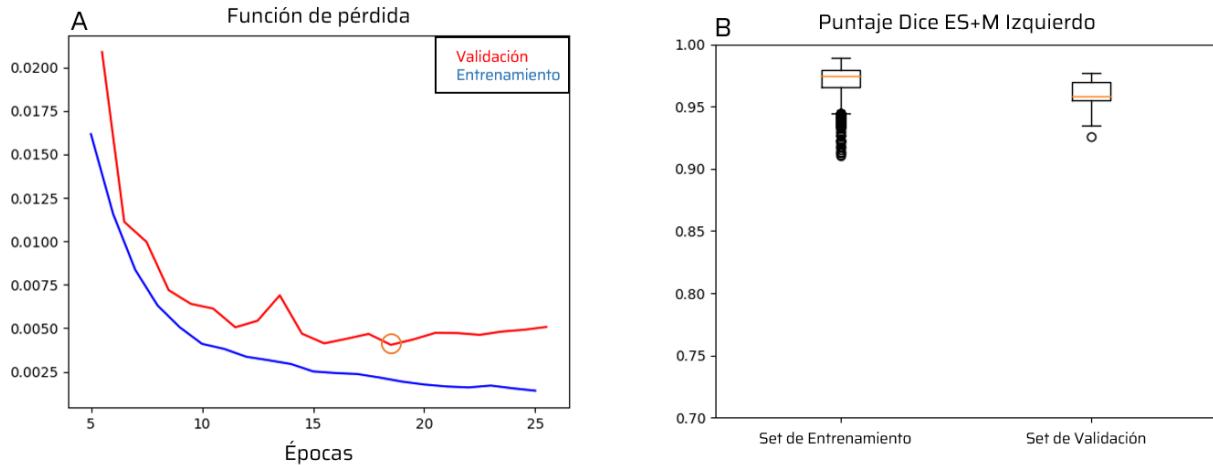


Figura 3.7. En la figura A se presentan los valores de la BCE en cada época del entrenamiento de la RNC para la segmentación del grupo ES+M. La curva roja representa los resultados en el set de validación mientras que la azul, en el entrenamiento. El círculo naranja indica el mejor resultado, obtenido en la época 18. En la figura B se presentan boxplots de los puntajes Dice para el modelo monoclas a la hora de segmentar el conjunto ES+M Izquierdo en el set de validación y entrenamiento.

Como se puede ver en el Figura 3.7-A, y tomando en cuenta los valores de la función de perdida como métrica, el modelo óptimo se encuentra en la época número 18. Así mismo, se calculó el valor de puntaje Dice para este mismo modelo en el set de validación y entrenamiento los cuales se presentan en la Figura 3.7 B

Como se puede observar, el puntaje Dice para el set de validación es muy similar con respecto al de entrenamiento y son considerablemente altos. Estos valores de Dice se traducen en segmentaciones confiables como las que se muestra en la Figura 3.8.

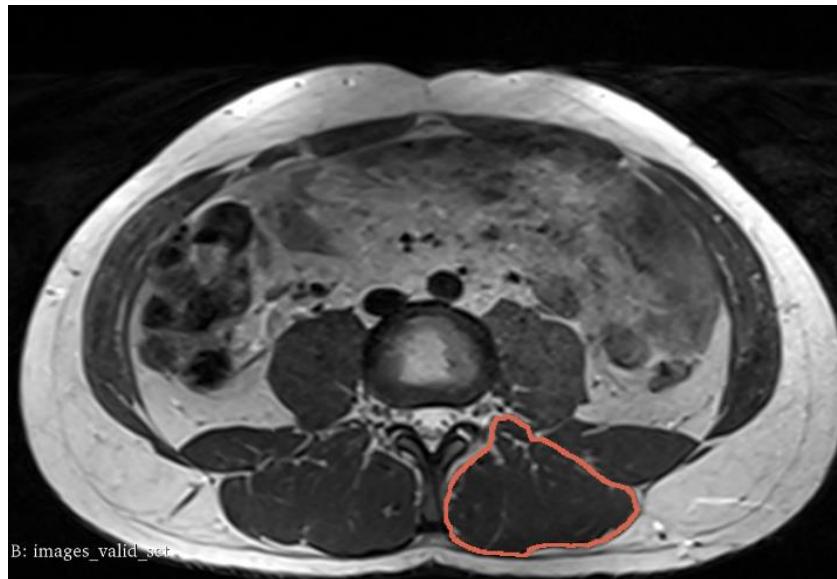


Figura 3.8. Segmentación automática del grupo ES+M izquierdo acertada en una imagen del set de validación.

Aun así, las diferencias contra las etiquetas automáticas y manuales no son nulas y el modelo es vulnerable a la sobre y subsegmentación. En la Figura 3.9, se muestra un ejemplo de sobresegmentación y otro de subsegmentación.

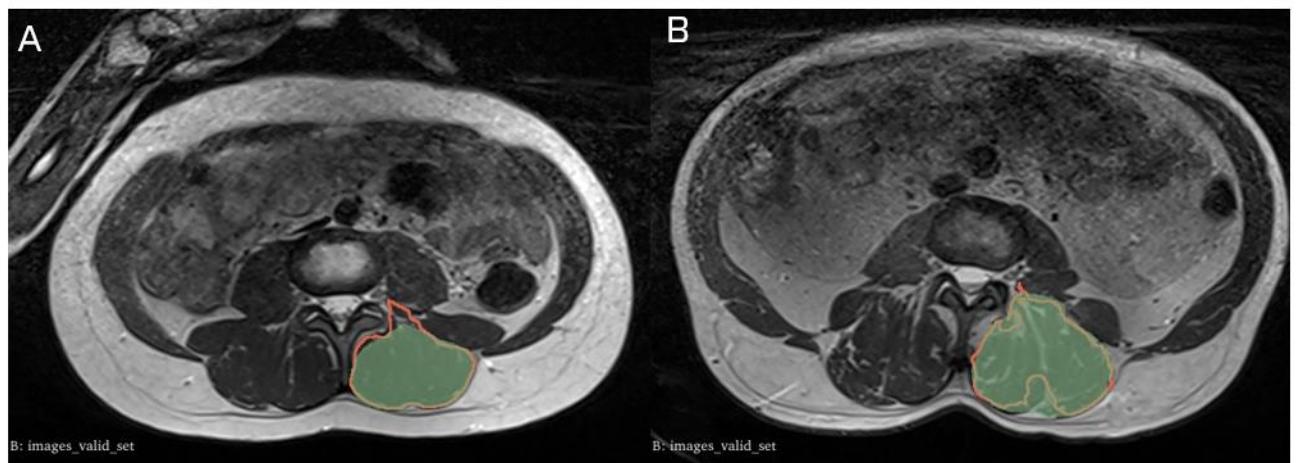


Figura 3.9 La sombra verde representa la segmentación manual mientras que el delineado rojo, la segmentación automática. En la figura A se observa una sobresegmentación por parte del modelo mientras que en la figura B una subsegmentación.

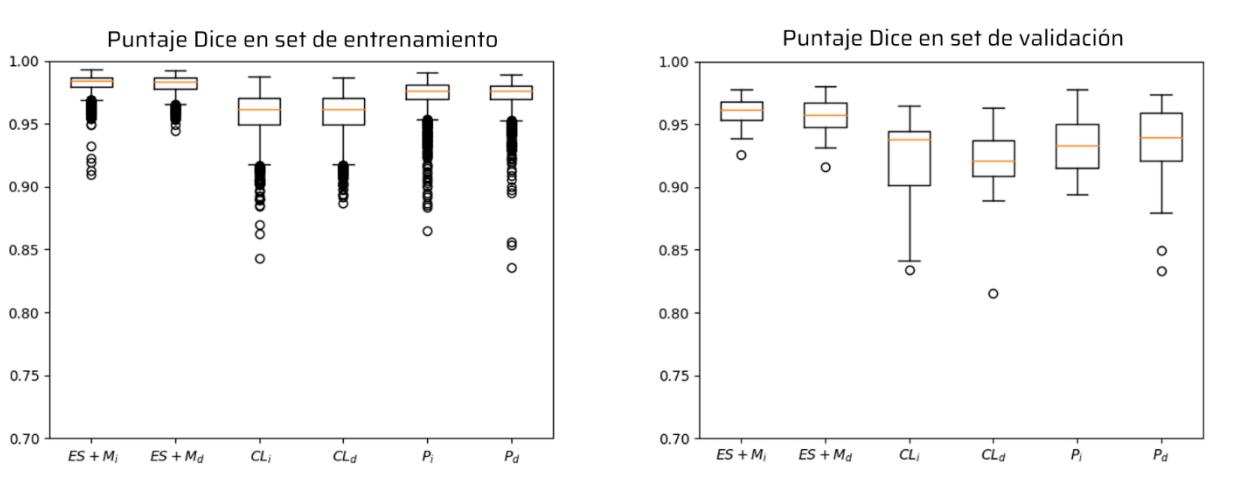
En conclusión, la arquitectura original U-Net y los hiperparámetros que se utilizaron, son suficientes para crear un modelo que entregue excelentes resultados a la hora de segmentar el grupo ES+M en imágenes Dixon de RMN.

### 3.3.2 Multiclasificación

Empleando la misma configuración que para el modelo monoclasificación, se entrenó un modelo multiclasificación con los mismos hiperparámetros y usando el mismo set de datos. En este caso el modelo segmenta los músculos P, CL y ES+M, con etiquetas distintas para el lado derecho e izquierdo.

Es importante mencionar que en este modelo y todos los que le siguen (debido a que se trata de modelos multiclasificación) se entrenaron utilizando el método de codificación *One-Hot*. Este permite subdividir la imagen objetivo en la cantidad de clases que estén representadas en la misma. De esta manera, se puede evaluar la función de perdida entre cada una de las etiquetas objetivo y las múltiples salidas del modelo.

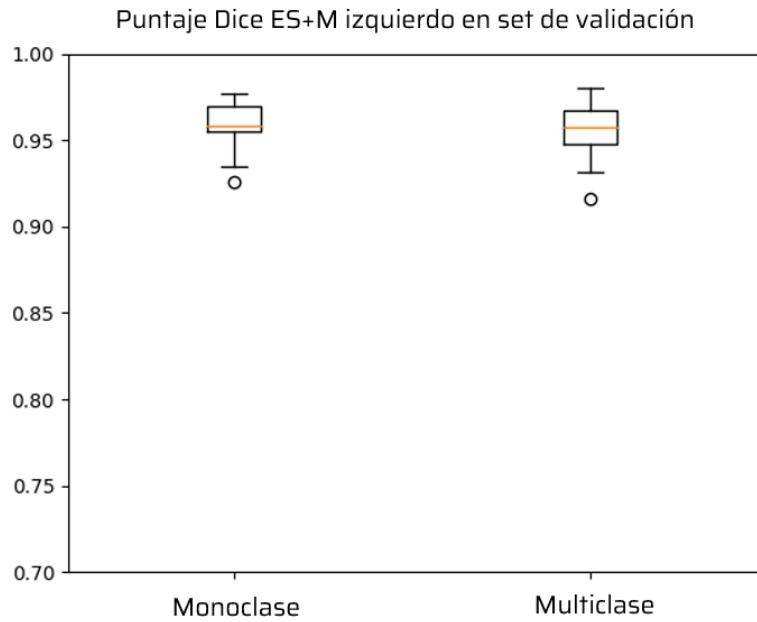
Se entrenó durante 35 épocas, pero se obtuvo un rendimiento máximo en la número 26. Los resultados en cuanto a puntaje Dice en el set de entrenamiento y validación se presentan en la Figura 3.10.



*Figura 3.10. Boxplots de los puntajes Dice para cada músculo del modelo multiclasificación en el set de entrenamiento y validación. El subíndice i representa los músculos de la izquierda mientras que el d, los de la derecha.*

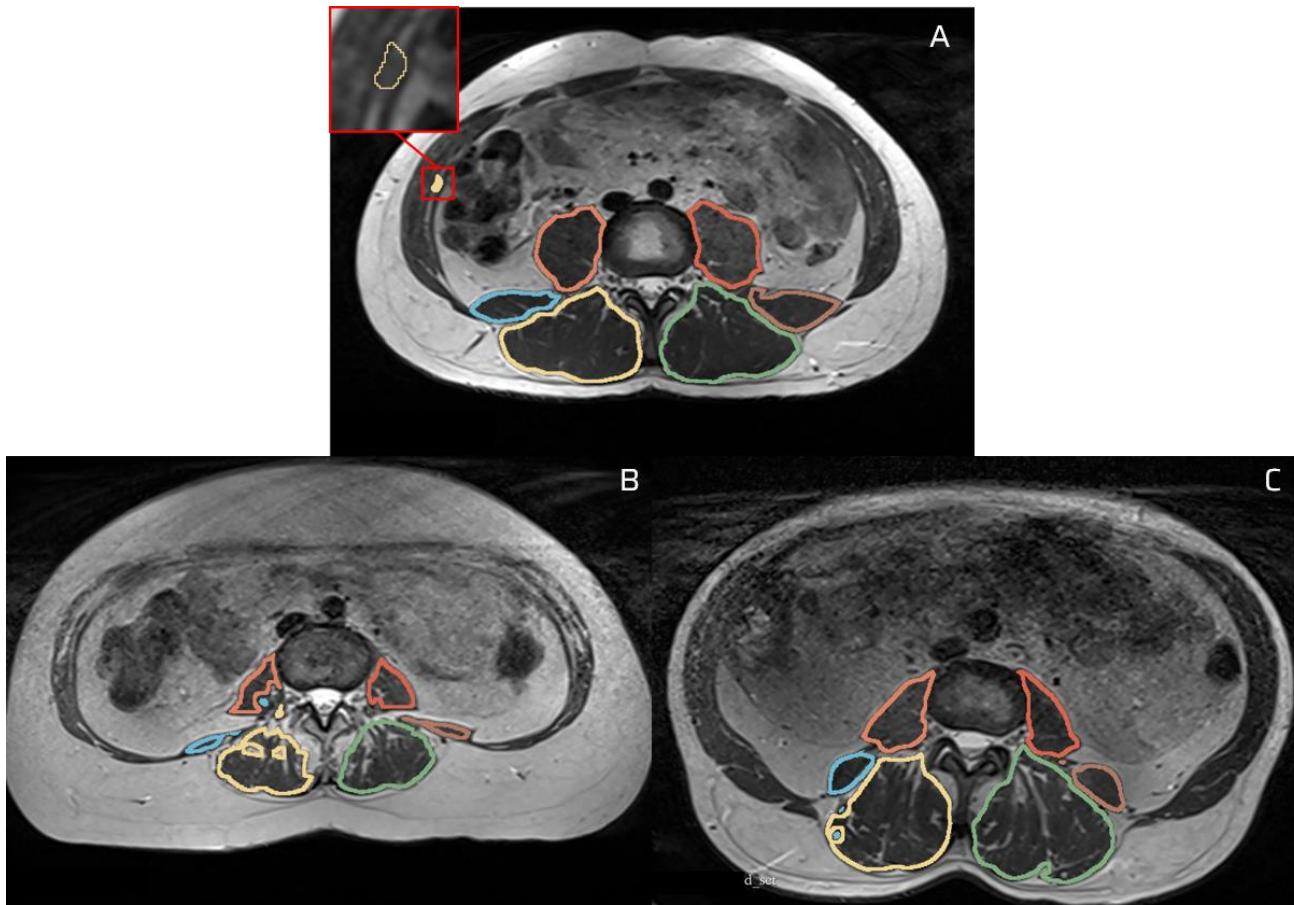
Como se puede observar, las medias de los resultados en el set de validación son lógicamente menores comparado con las de entrenamiento y también se puede ver que los músculos más pequeños como los cuadrados lumbares presentaron peores resultados que el resto. No obstante, los resultados para el set de validación son muy

buenos con todas las medianas mayores a 0.9. Al compararlo con el modelo de una única clase encontramos como se ve en la figura 3.11, que no hay un deterioro del desempeño del modelo multiclas.



*Figura 3.11. Comparativa entre modelo Monoclas y Multiclas para el mismo músculo. No se observan diferencias significativas entre el desempeño de los modelos.*

Cabe destacar que el modelo produce algunas segmentaciones con resultados relativamente pobres que se representan como los valores atípicos en los *boxplot*. En la Figura 3.12 se muestran las 3 segmentaciones con peor performance que corresponden a dichos valores atípicos. Si bien los defectos presentados a continuación pueden parecer menores, repercuten considerablemente cuando se intenta medir el tamaño y la fracción de tejido graso.



*Figura 3.12. Defectos en las segmentaciones realizadas por el modelo en ciertas imágenes del set de validación. En la figura A se presenta una sobresegmentación en el extremo superior izquierdo de la imagen. En la b, una subsegmentación en el Psoas y el conjunto ES+M. En la C, una superposición de segmentaciones entre el ES+M y el CL.*

### 3.3.2.1 Optimización del modelo 2D

En esta sección se analizan distintas estrategias para optimizar el modelo entrenado en la sección anterior, con el objetivo de evitar los defectos observados en los valores atípicos y optimizar el entrenamiento para obtener similares resultados con menos recursos computacionales. Las optimizaciones realizadas en el modelo 2D podrán ser transferidas al 3D.

#### 3.3.2.1.1 Función de pérdida Normalizada

Como primera medida se intentó reducir la cantidad de defectos en las segmentaciones mediante el uso de una función de perdida normalizada. Esta técnica consiste en que la función de perdida penalice de manera uniforme cada una de las clases que el modelo intenta segmentar. En este caso, se debe ponderar el tamaño de cada una de las etiquetas con respecto al tamaño de la imagen y con respecto al excedente de la imagen.

$$W_A = \frac{T}{A_m} \quad (9)$$

$$W_R = \frac{T-A_m}{A_m} \quad (10)$$

Al primero (9) lo llamaremos pesaje Absoluto  $W_A$  en el cual se compara el tamaño total de la imagen contra el área del músculo  $A_m$ . El segundo (10)  $W_R$ , relativo, el cual compara el área de todo lo que rodea a la etiqueta contra el área del músculo. Ambos pesajes se insertan en la función de perdida (11) como un término  $W$  que pondera el BCE de cada etiqueta:

$$L(X, Y)_M = -W_m[Y * \log(X) + (1 - Y) * \log(1 - X)] \quad (11)$$

cabe destacar que por cada músculo existe una función de pérdida  $L_m$  la cual contiene el término  $W_m$ . Todas estas son luego promediadas para obtener una función de perdida general. Es en esta instancia en la que el término  $W_m$  cumple su función de normalizar el valor total. De esta manera se entrenaron modelos normalizados y se evaluó su desempeño en términos de puntaje Dice.

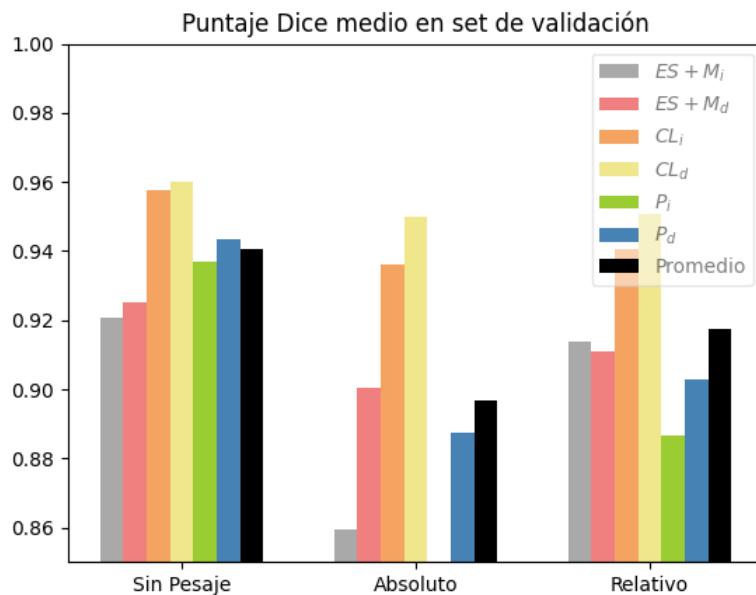
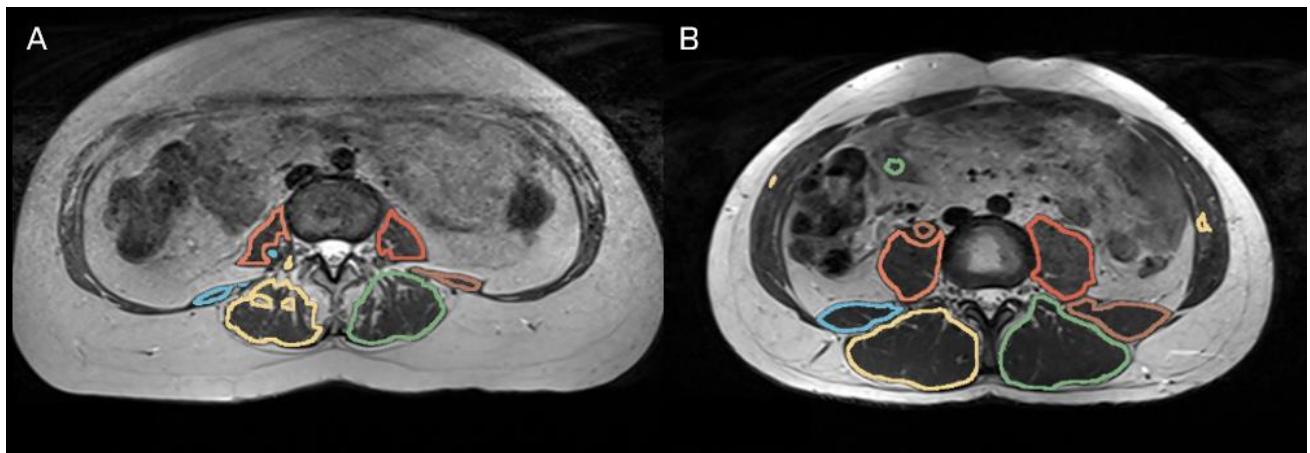


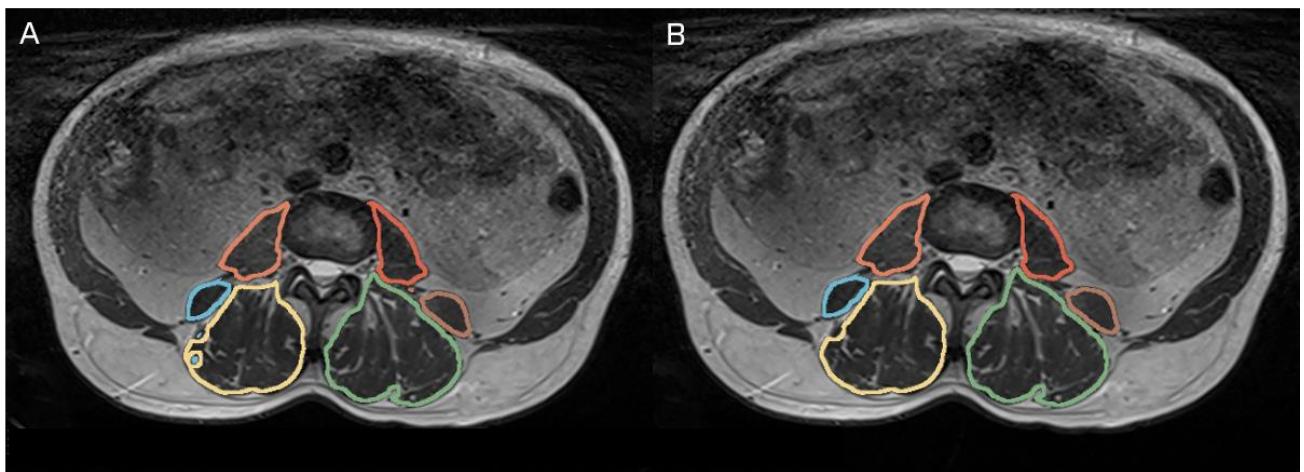
Figura 3.13. Diferencias de desempeño en set de validación según técnica de pesaje. Las columnas de colores representan el valor medio de cada músculo mientras que la columna negra corresponde al promedio de todas las columnas.



*Figura 3.14. Ejemplo de defectos encontrados al utilizar una función de pérdida normalizada. En la imagen A se pueden ver subsegmentaciones en el lado izquierdo de la imagen mientras que en la imagen B se presentan solapamientos de segmentaciones y sobresegmentaciones.*

Se puede observar en la Figura 3.13 que los resultados en términos de puntaje Dice fueron peores para ambas técnicas de pesaje con respecto a los modelos no normalizados con una media promedio de 0.941 para el modelo entrenado sin pesaje, 0.917 para el entrenado con pesaje absoluto y 0.927 para el relativo. Por otro lado, en la Figura 3.14 se observa que se generaron aún más defectos de sobresegmentación y solapamiento.

Aunque el experimento resultó redundante ya que no se corrigieron los defectos en la imagen, se determinó que realizar un post procesamiento de las etiquetas es el enfoque más sencillo y acertado. Este procesamiento consta de dos partes, la primera es llenar todas las áreas encerradas por una determinada etiqueta. Si observamos la Figura 3.14-A podemos ver que en el ES+M derecho (a la izquierda de la imagen) se encuentran islas dentro de la etiqueta que siguen correspondiendo al músculo, al llenarlas mejoramos la calidad de la segmentación. La segunda parte consiste en eliminar todos los fragmentos más pequeños de cada etiqueta. En la Figura 3.14-B encontramos múltiples sobresegmentaciones de múltiples etiquetas, cada una de estas son más pequeñas que la segmentación correcta, de esta manera al eliminar las pequeñas, eliminamos sobresegmentaciones. En la Figura 3.15 se observa como al implementar el procesamiento mejoramos la calidad de la segmentación.



*Figura 3.15. Diferencias de desempeño para el modelo original sin pesaje al utilizar post procesamiento. La figura A muestra la segmentación automática realizada por el modelo con islas dentro del ES+M, superposiciones con el CL y una pequeña sobre segmentación a la derecha de la imagen. La figura B muestra como el post procesado elimina la superposición de etiquetas, rellena islas y elimina sobre segmentaciones.*

### **3.3.2.1.2 Entrenamiento con distintos sets de datos**

Como segunda medida de optimización y mejora, se entrenaron modelos de las mismas características que los anteriores, pero en vez de utilizar todo el set de datos aumentado, se utilizaron las distintas componentes que lo forman de forma independiente. De esta manera se logró evaluar el desempeño de los modelos y determinar qué estrategia de aumentado de datos es la más adecuada.

Para evaluar los diferentes tipos de aumentado, se compararon los puntajes Dice de cada uno de los músculos para múltiples modelos entrenados con la misma arquitectura, técnica de pesaje y la misma división del set de datos de las imágenes originales (70/30).

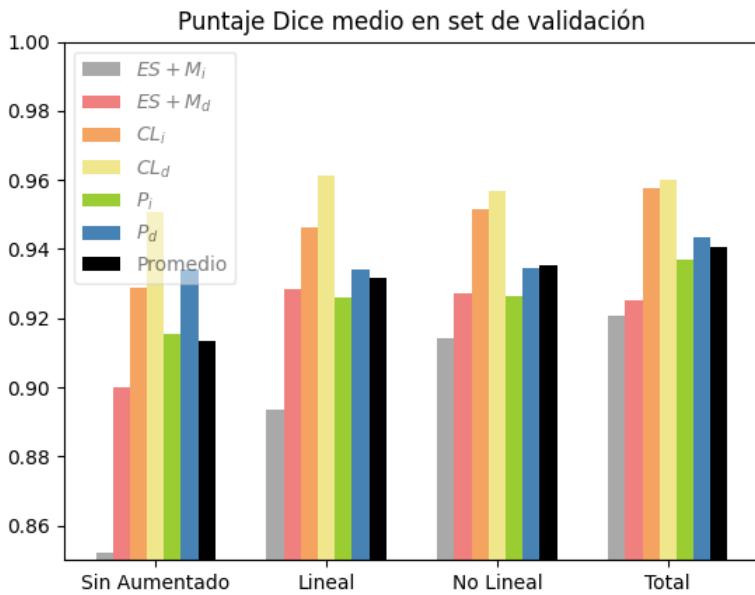


Figura 3.16. Diferencia de desempeño según tipo de Aumento. Las columnas de colores representan el valor medio de cada músculo mientras que la columna negra corresponde al promedio de todas las columnas.

Se puede observar en el Figura 3.16 que la diferencia entre los modelos entrenados con todo el aumento y el aumento no lineal son ligeramente distintas con medias promedio de 0.941 y 0.935 respectivamente. En cuanto al lineal, el valor promedio de sus medias fue de 0.932, mostrando una pequeña baja en el rendimiento del modelo comparado con los casos anteriores. Es importante considerar que el set de datos de aumento lineal consiste en un quinto del tamaño del aumento total. Vale destacar que la diferencia entre cualquier tipo de aumento y los datos no aumentados es sustancial, ya que la media para este último es de 0.91.

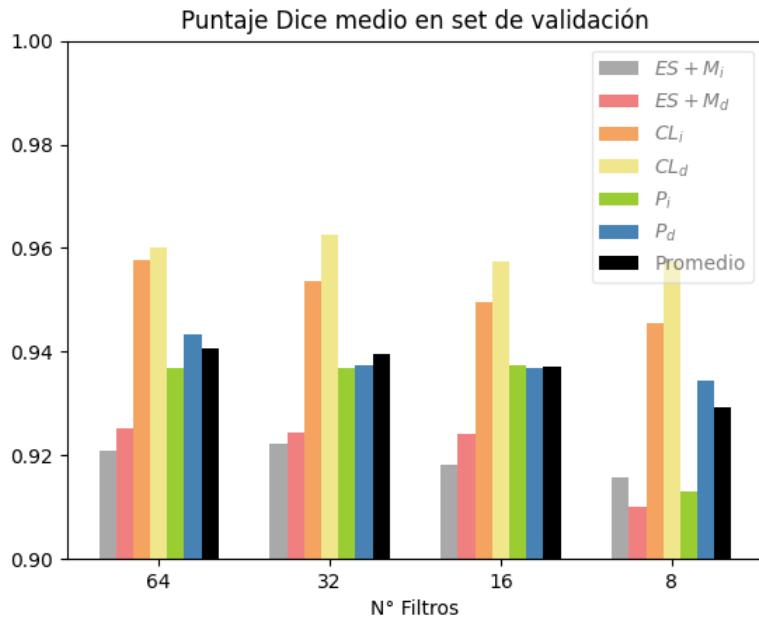
En conclusión, el aumento no lineal tuvo un mayor impacto positivo en el rendimiento del modelo. Sin embargo, en caso de contar con escasos recursos computacionales, resulta beneficioso utilizar el set de datos con aumento lineal.

### 3.3.2.1.3 Variación en la cantidad de filtros en la arquitectura U-Net

Como última técnica de optimización, se entrenaron modelos utilizando el mismo set de entrenamiento y sin métodos de pesaje, pero variando la cantidad de filtros iniciales de la arquitectura U-Net. Se partió de 64 filtros 3x3 al igual que en la arquitectura original [16] y se fue dividiendo por dos la cantidad de filtros hasta llegar a 8.

Tomando en cuenta que la cantidad de filtros iniciales influye directamente sobre la cantidad de memoria RAM demandada a la hora de entrenar, podemos comprender

porque es beneficioso reducir la cantidad de filtros iniciales. De esta manera, utilizando configuraciones más eficientes nos permitimos incrementar el número de imágenes del *batch* y acelerar los tiempos de entrenamiento.

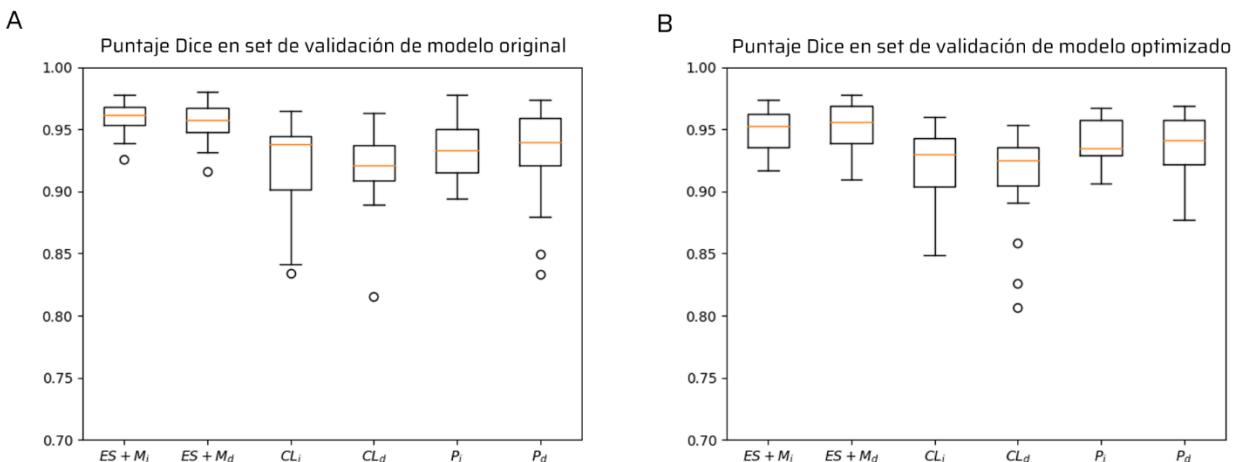


*Figura 3.17. Comparación del desempeño de distintas RNC U-Net con distinta cantidad de filtros 3x3 iniciales. Las columnas de colores representan el valor medio de cada músculo mientras que la columna negra corresponde al promedio de todas las columnas.*

Observando la Figura 3.17 se puede concluir que, para el entrenamiento de un modelo 2D, la cantidad inicial de filtros más eficiente es de 16 ya que tiene un desempeño promedio de 0.937 casi tan bueno como la configuración original con un puntaje Dice promedio de 0.94. Utilizar 8 filtros iniciales ya no resulta recomendable, porque impacta negativamente en el desempeño del modelo para segmentar el músculo ES+M derecho y el P izquierdo con un puntaje promedio de 0.93.

### 3.3.2.2 Modelo 2D optimizado

Luego de realizar esta serie de experimentos donde se evaluaron técnicas de optimización, se decidió entrenar un modelo a partir de las conclusiones obtenidas. Este modelo se entrenó con una arquitectura de 16 filtros sin pesaje y utilizando el set de datos aumentado linealmente.



*Figura 3.18 Boxplots de puntaje Dice en set de validación para modelo original y modelo optimizado.*

El modelo fue entrenado durante 200 épocas obteniendo un desempeño máximo en la 163. Los puntajes Dice del modelo optimizado se presentan en la Figura 3.18-B y resultan ser similares o tan buenos como los resultados del modelo original (Figura 3.18-A). Teniendo en cuenta que las demandas de memoria RAM que requiere el entrenamiento del modelo optimizado y que este proceso utiliza muchas menos imágenes es preferible utilizar estas configuraciones en términos de eficiencia.

### **3.3.2.3 Modelo 2D de máximo desempeño**

Una vez entrenado el modelo original y el optimizado, se entrenó un último modelo que tuviese el mejor desempeño posible. Para ello se utilizó una placa de video NVIDIA GeForce RTX 3090 con 24 GB de memoria RAM para entrenar un modelo con la arquitectura original de 64 filtros, utilizando el set de datos con aumento total, ninguna técnica de pesaje y, a diferencia de los modelos anteriores, empleando un batch de 8 imágenes. El criterio para frenar el entrenamiento fue considerando el puntaje Dice y no el resultado de la función de pérdida. El modelo entrenado posee 13 millones de parámetros entrenables y a continuación se expone el desempeño del mismo y las estadísticas del entrenamiento de manera detallada.

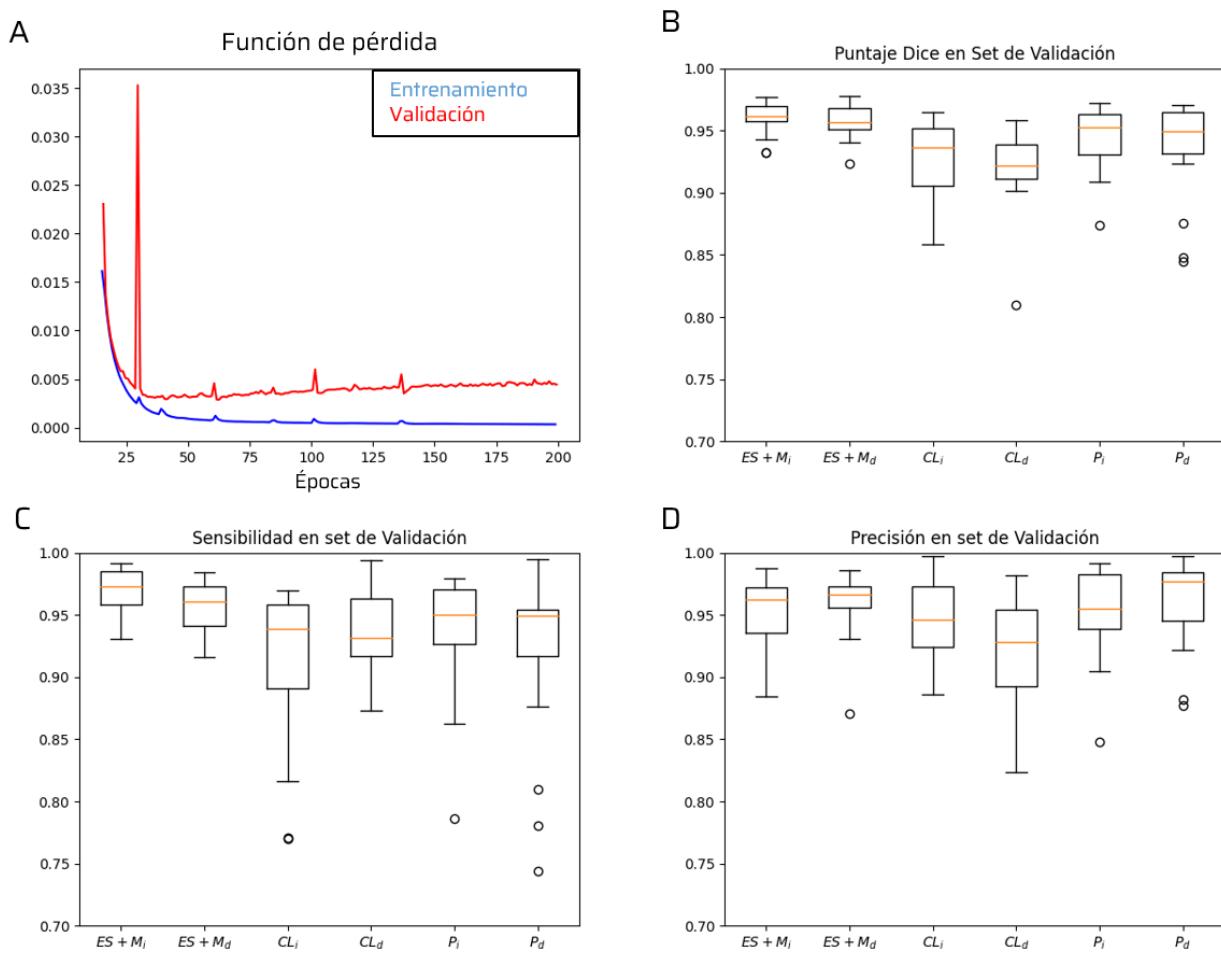


Figura 3.19. Curva de valor de función de perdida por época y Boxplots de desempeño en set de validación en términos de puntaje Dice, Sensibilidad y precisión para modelo 2D de máximo desempeño.

Si bien no se utilizó el valor de la función de pérdida como indicador de entrenamiento, se puede observar en la Figura 3.19-A que entre las épocas 25 y 50, las curvas comienzan a diverger indicando un overfitting del modelo. No obstante, el modelo sigue entrenando y en la época 178 se obtiene un modelo cuyo desempeño promedio entre los músculos en términos de puntaje Dice es máximo.

Como se puede ver en el Figura 3.19-B, los músculos con peor desempeño fueron los más pequeños que nuevamente y debido a las características del puntaje Dice, son más vulnerables a pequeñas imprecisiones. Las medianas (IQR) de los pares de músculo fueron 0.951 (0.035) para los P, 0.929 (0.043) para los CL y 0.962 (0.018) para los conjuntos ES+M.

Para analizar más en detalle la performance de la segmentación automática, se presentan en la Figura 3.19-C y 3.19-D boxplots con los valores de sensibilidad y precisión para cada músculo. Se puede observar que las segmentaciones de peor performance se ven mayormente afectada por una pérdida en la sensibilidad. Finalmente, en la tabla 3.1 se pueden observar los puntajes Dice para cada una de las imágenes del set de validación.

*Tabla 3.1. Puntajes Dice en cada imagen de set de validación para modelo 2D*

Imagen	P <sub>i</sub>	CL <sub>i</sub>	ES+M <sub>i</sub>	P <sub>d</sub>	CL <sub>d</sub>	ES+M <sub>d</sub>
1	0.970	0.869	0.949	0.969	0.924	0.974
2	0.939	0.910	0.962	0.947	0.919	0.970
3	0.933	0.907	0.969	0.966	0.908	0.941
4	0.874	0.889	0.958	0.848	0.909	0.966
5	0.965	0.935	0.971	0.876	0.918	0.967
6	0.972	0.938	0.974	0.941	0.951	0.955
7	0.930	0.905	0.932	0.967	0.937	0.950
8	0.952	0.895	0.932	0.956	0.906	0.940
9	0.966	0.965	0.943	0.948	0.945	0.969
10	0.943	0.956	0.957	0.941	0.934	0.949
11	0.959	0.951	0.970	0.928	0.918	0.969
12	0.964	0.950	0.976	0.961	0.943	0.957
13	0.958	0.913	0.965	0.950	0.933	0.954
14	0.958	0.951	0.962	0.844	0.809	0.924
15	0.915	0.953	0.964	0.961	0.901	0.952
16	0.953	0.956	0.977	0.968	0.919	0.968
17	0.909	0.858	0.960	0.970	0.940	0.977
18	0.920	0.951	0.962	0.923	0.959	0.956
Mediana	0.953	0.937	0.962	0.949	0.922	0.957
(iqr)	(.032)	(.046)	(.013)	(.034)	(.028)	(.018)
Media	0.943	0.925	0.960	0.937	0.921	0.958
(std)	(.026)	(.033)	(.013)	(.040)	(.032)	(.014)

### 3.4 Red neuronal convolucional 3D

En esta sección se extiende el trabajo de la sección anterior para lograr la segmentación automática de los músculos de la zona lumbar a partir de la imagen Dixon 3D, lo cual permitirá hacer un análisis volumétrico del tamaño del músculo y su infiltración grasa. En este caso, se segmentan los 4 pares de músculos involucrados en la zona lumbar (a los tres pares de músculos segmentados para el caso 2D, se agregan los músculos Iliacos). A continuación, se detallará el entrenamiento y el desempeño del modelo en cuestión.

Para poder generar etiquetas 3D se entrenó y evaluó un modelo de segmentación basado en una arquitectura U-Net 3D. Este modelo recibe en la entrada todos los véxeles de la imagen Dixon IP 3D y brinda a la salida las etiquetas en una imagen 3D del mismo tamaño. A diferencia de la arquitectura original, los filtros de las capas de convolución y *max pooling* se expanden en la tercera dimensión pasando de 3x3 y 2x2 a 3x3x3 y 2x2x2 respectivamente. Por esta razón, el entrenamiento de un modelo de estas características implica una demanda considerablemente alta de recursos computacionales, requiriendo la adopción de medidas particulares para poder entrenarlo sin la necesidad de recurrir a una implementación con múltiples GPUs. Primero, como se detalla en el capítulo 3.1.2.2 las imágenes fueron escaladas a un cuarto de su tamaño original. Segundo, el tamaño de batch de entrenamiento consistió en una única imagen. Finalmente, y partiendo de la información obtenida sobre el desempeño de los modelos 2D utilizando diferente cantidad de filtros iniciales, se optó por utilizar 16 en vez de 64.

Todas estas decisiones se tomaron realizando pruebas empíricas y teniendo en cuenta un estudio de caso realizado por *Dell Technologies* en el cual se evalúa las diferentes demandas de memoria RAM para distintos tamaños de imagen de entrada, tamaño del batch y cantidad de filtros (resultados presentados en Figura 3.20). [26]

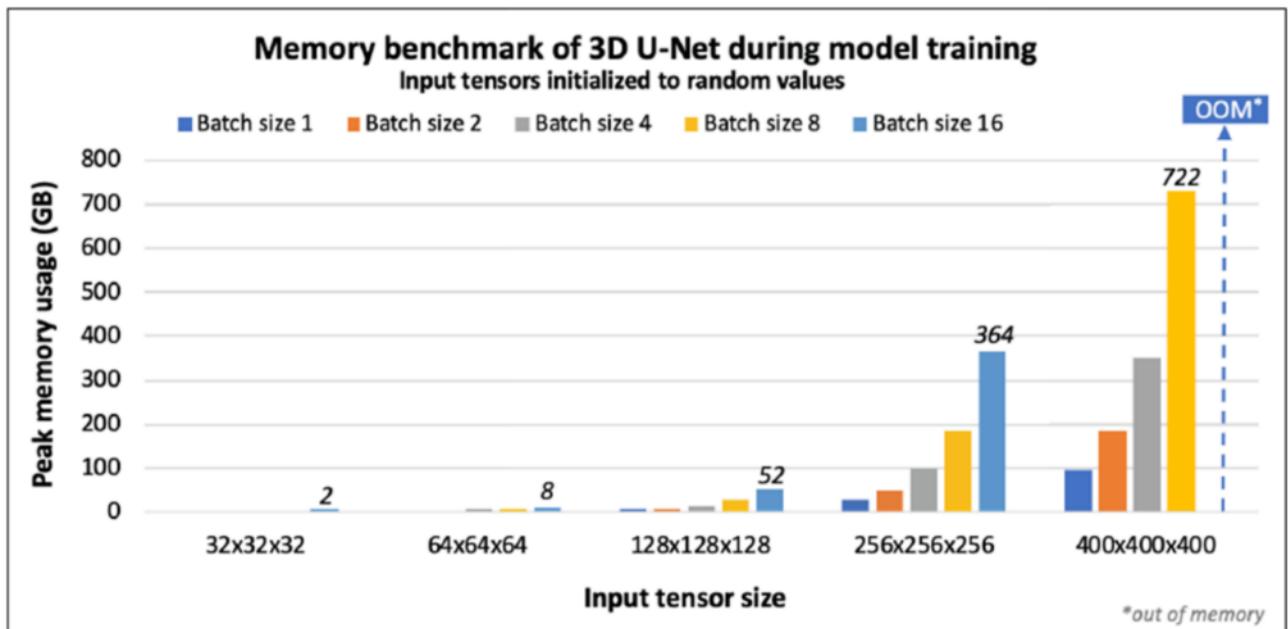


Figura 3.20. Consumo teórico para modelo de tipo U-Net 3D. Las columnas de colores corresponden a distintos tamaños de batch siendo azul: 1, naranja: 2, gris: 4, amarillo: 8 y celeste: 16. Se puede observar como el tamaño de la imagen y batch incrementa la demanda de recursos computacionales [26].

Según este estudio, el consumo de memoria debido a la creación de mapas de características sigue la siguiente formula:

$$Bytes = \left[ H * W * D * \left( C_{in} + F * \left( \frac{10297}{572} \right) + C_{out} \right) * 16 * B \right] - 1 \quad (12)$$

Para los cuales  $H$ ,  $W$  y  $D$  es el tamaño tridimensional de la imagen,  $C_{in}$  y  $C_{out}$  son la cantidad de canales de entrada y salida,  $B$  es el tamaño del *batch* y  $F$  la cantidad inicial de filtros. Si bien, no se realizaron cálculos siguiendo la formula, permitió decidir que variables ajustar y mediante pruebas experimentales se logró entrenar un modelo con un buen desempeño en el hardware disponible (NVIDIA RTX 3090) con 24 GB de RAM.

A partir de estas configuraciones, se entrenó un modelo con 2.5 millones de parámetros entrenables durante 200 épocas donde el mejor desempeño en términos de valor de la función de perdida estuvo en la instancia 173.

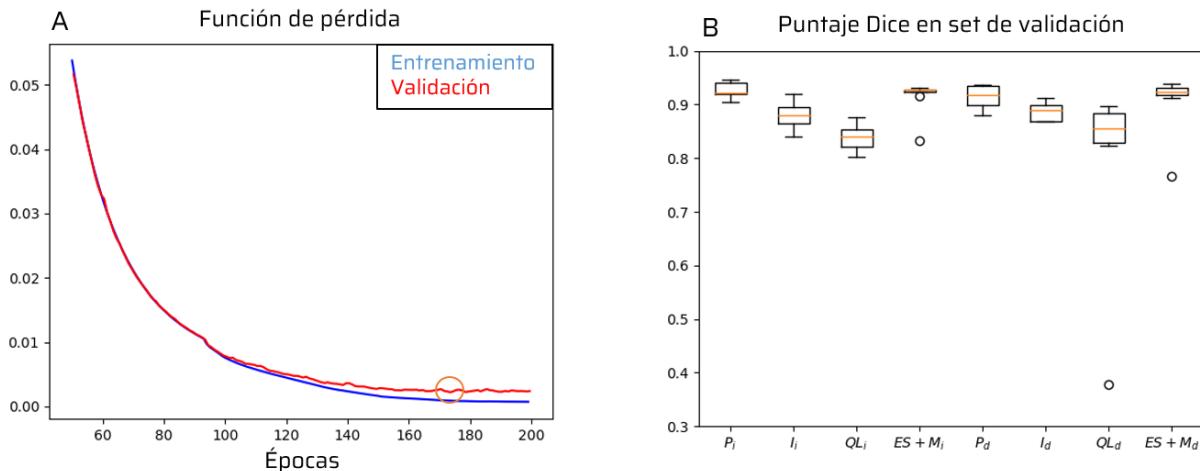


Figura 3.21 En la figura A se presentan los valores de la BCE en cada época del entrenamiento. La curva roja representa los resultados en el set de validación mientras que la azul en el de entrenamiento. El circulo naranja indica el mejor resultado, obtenido en la época 173. En la figura B se presentan boxplots de los puntajes Dice para el modelo 3D a la hora de segmentar las imágenes del set de validación.

Si bien los puntajes Dice de este modelo fueron inferiores a los modelos de segmentación 2D, la tarea de segmentación de todo el volumen muscular tiene mayor complejidad ya que las regiones involucradas son más heterogéneas. En la Figura 3.21-B, se observa los puntajes Dice obtenidos en el set de validación.

Al igual que en los modelos anteriores, los puntajes para los músculos más pequeños resultan ser los peores debido a que son más susceptibles a los errores de segmentación. No obstante, las medianas (IQR) de los puntajes Dice para los pares de

músculos son considerablemente altos, siendo de 0.921 (0.024) para los P, 0.881 (0.031) para los I, 0.847 (0.044) para los CL y 0.926 (0.010) para los ES+M. Por otro lado, las segmentaciones del modelo son consistentes y considerablemente buenas en todo el set de validación. En la Figura 3.22, se muestra una segmentación automática en una imagen del set de datos precisa.

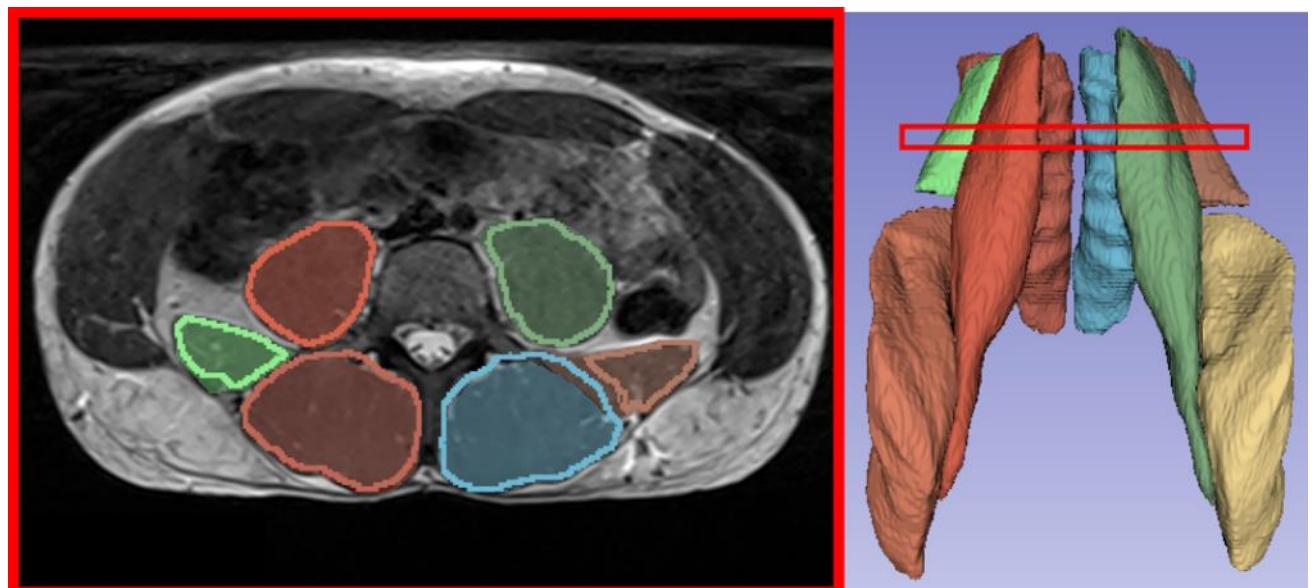


Figura 3.22 Segmentación automática de imagen de set de validación precisa. En el corte axial, el contorno corresponde a la predicción del modelo mientras que la sombra corresponde a la segmentación manual.

Si bien dentro de las 8 imágenes que componen el set de validación los resultados fueron similares a la imagen anterior, también se obtuvo una segmentación particularmente mala para el músculo CL derecho. Esta está representada en la Figura 3.21 como el *valor atípico* con un Dice de aproximadamente 0.4, y puede visualizarse en la Figura 3.23.

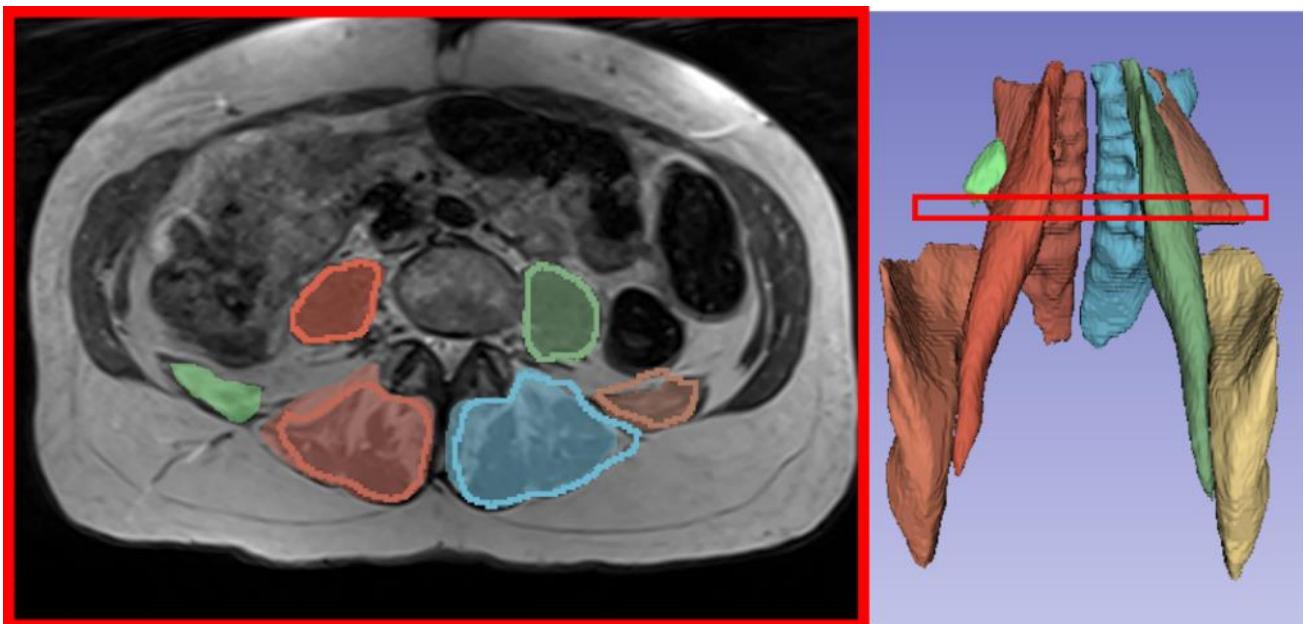


Imagen 3.23. Segmentación automática de imagen de set de validación imprecisa. En el corte axial, el contorno corresponde a la predicción del modelo mientras que la sombra corresponde a la segmentación manual. Se puede observar que en este corte el CL derecho (a la izquierda de la imagen) no ha sido detectado por el modelo. También se lo puede observar en el modelo 3D, dónde el CL derecho (en verde claro) se encuentra subsegmentado.

Respecto de esta segmentación poco precisa, es necesario destacar que la imagen corresponde a un sujeto con un alto grado de infiltración graso, y en el set de entrenamiento se encontraban pocas imágenes de estas características. Debido al reducido tamaño del set de validación, esta imagen en particular tiene un impacto grande en la evaluación general del modelo. En la Tabla 3.2, se pueden ver los resultados de cada uno de los músculos en todo el set de validación, esta imagen en particular es la numero 4.

*Tabla 3.2. Puntajes Dice en cada imagen de set de validación para modelo 3D*

Imagen	P <sub>i</sub>	I <sub>i</sub>	CL <sub>i</sub>	ES+M <sub>i</sub>	P <sub>d</sub>	I <sub>d</sub>	CL <sub>d</sub>	ES+M <sub>d</sub>
1	0.942	0.892	0.802	0.915	0.919	0.869	0.856	0.919
2	0.921	0.859	0.825	0.926	0.916	0.869	0.855	0.922
3	0.923	0.92	0.844	0.93	0.937	0.913	0.88	0.936
4	0.92	0.84	0.806	0.832	0.881	0.904	0.377	0.766
5	0.905	0.88	0.849	0.927	0.891	0.898	0.83	0.938
6	0.941	0.866	0.836	0.927	0.934	0.868	0.896	0.925
7	0.947	0.903	0.876	0.926	0.937	0.898	0.891	0.911
8	0.92	0.878	0.865	0.927	0.902	0.882	0.822	0.931
Mediana	0.921	0.878	0.844	0.927	0.916	0.898	0.855	0.925
(iqr)	(.021)	(.031)	(.033)	(.004)	(.036)	(.031)	(.055)	(.015)
Promedio	0.925	0.878	0.843	0.914	0.914	0.890	0.793	0.904
(std)	(.014)	(.025)	(.026)	(.033)	(.022)	(.018)	(.173)	(.057)

## **4. Evaluación de la calidad muscular en sujetos sanos sedentarios y activos**

Contando con las herramientas de segmentación automática de los músculos de la zona lumbar, en este capítulo se evalúa la calidad muscular en términos de tamaño y fracción de tejido graso en datos correspondientes a 118 sujetos. Se compara la precisión de las métricas basadas en cortes 2D respecto de las más precisas 3D, así como las distintas variables que afectan estas métricas.

### **4.1 Procesamiento de imágenes**

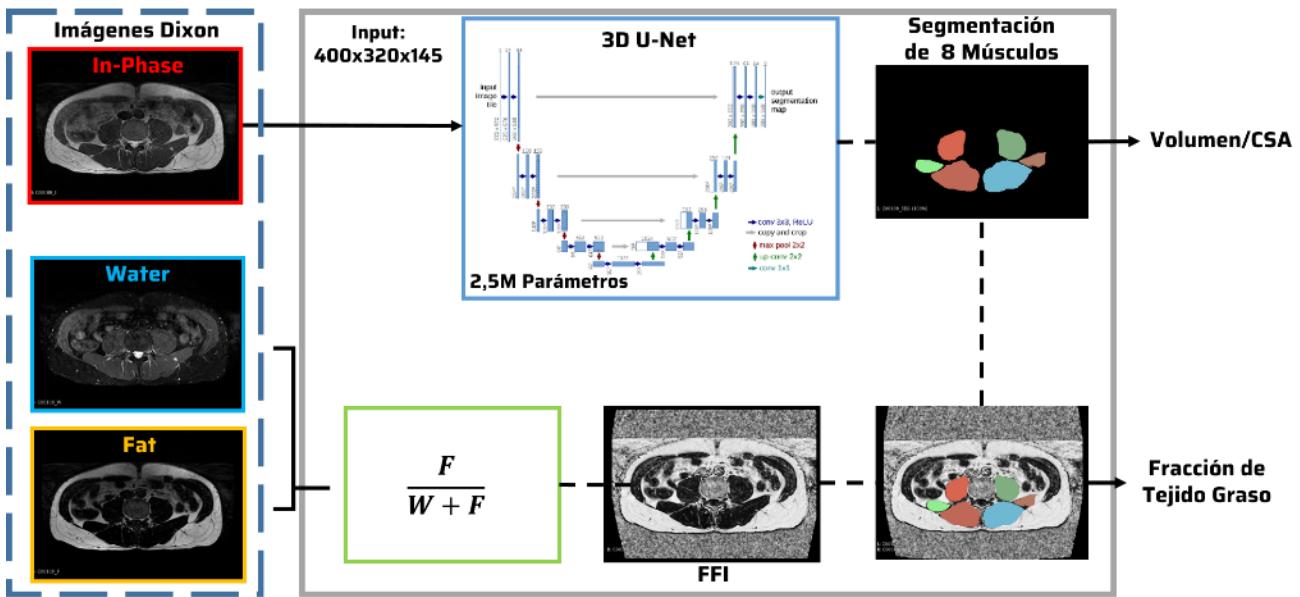
Se utilizaron los modelos de segmentación presentados en los capítulos 3.3.2.3 y 3.4 para segmentar las imágenes Dixon IP de 118 sujetos distintos, obteniendo etiquetas 2D y 3D para cada sujeto respectivamente. Para obtener métricas de tamaño muscular, se obtuvieron los volúmenes y áreas de sección transversal de los músculos de interés de todos los sujetos empleando las ecuaciones 13 y 14 respectivamente.

$$V_M = T \sum_{i=1}^L L_{Mi} \quad (13)$$

$$A_M = T \sum_{i=1}^L L_{Mi} \quad (14)$$

donde  $V_M$  y  $A_M$  corresponden al volumen y el área a la altura de L4 del músculo M respectivamente, T corresponde al tamaño de voxel (o píxel para el caso 2D),  $L_{Mi}$  corresponde al voxel  $i$  de la etiqueta  $L_M$  para el músculo  $M$ .

Para obtener métricas la infiltración grasa en cada músculo, primero se crearon las imágenes FF utilizando las imágenes Dixon de agua y grasa de cada sujeto como mencionamos en el capítulo 2.2. Luego se aplicaron las etiquetas correspondientes a cada músculo sobre la FF, quedándose así con los valores de FF para cada voxel pertenecientes a un músculo dado. Por último, se promediaron dichos valores y se obtuvieron las fracciones de tejido graso de todo el músculo (para las imágenes 3D) y de un corte a la altura de L4 (para el caso 2D) para todos los sujetos. En la Figura 4.1 se puede observar un esquemático de todo el proceso.



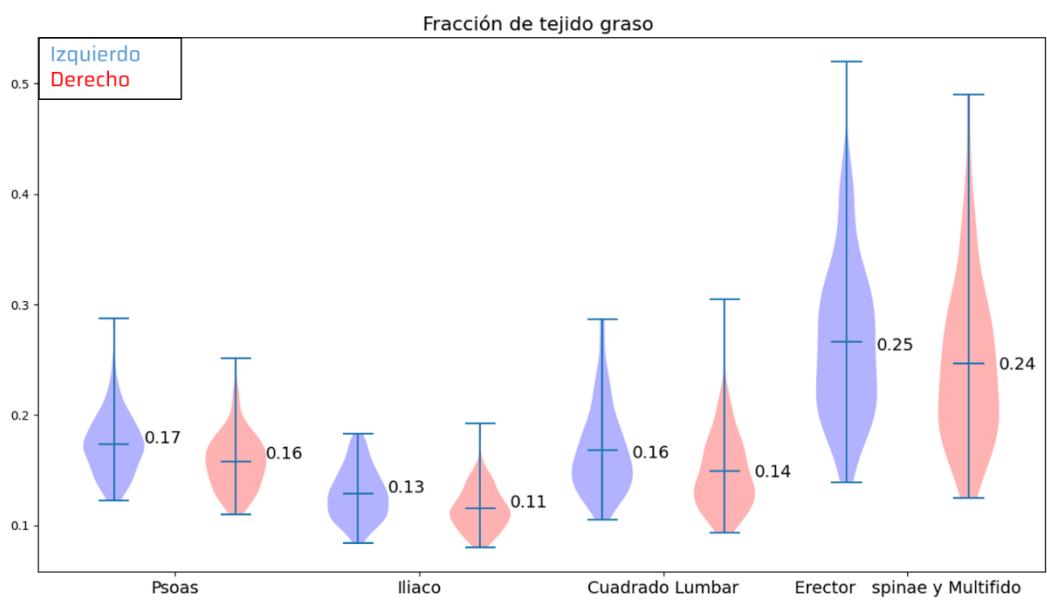
*Figura 4.1 Esquemático de implementación de modelos de segmentación automática. A la izquierda se muestran las imágenes de entrada, las cuales consisten de imágenes Dixon IP, W y F. La imagen IP es segmentada utilizando las RNC U-Net y se extrae una etiqueta de los 8 músculos en cuestión. Las imágenes W y F se utilizan para construir la imagen de FF. A partir de la segmentación y la imagen FF se obtiene el tamaño y la fracción de tejido graso de los músculos.*

## 4.2 Resultados Cuantitativos

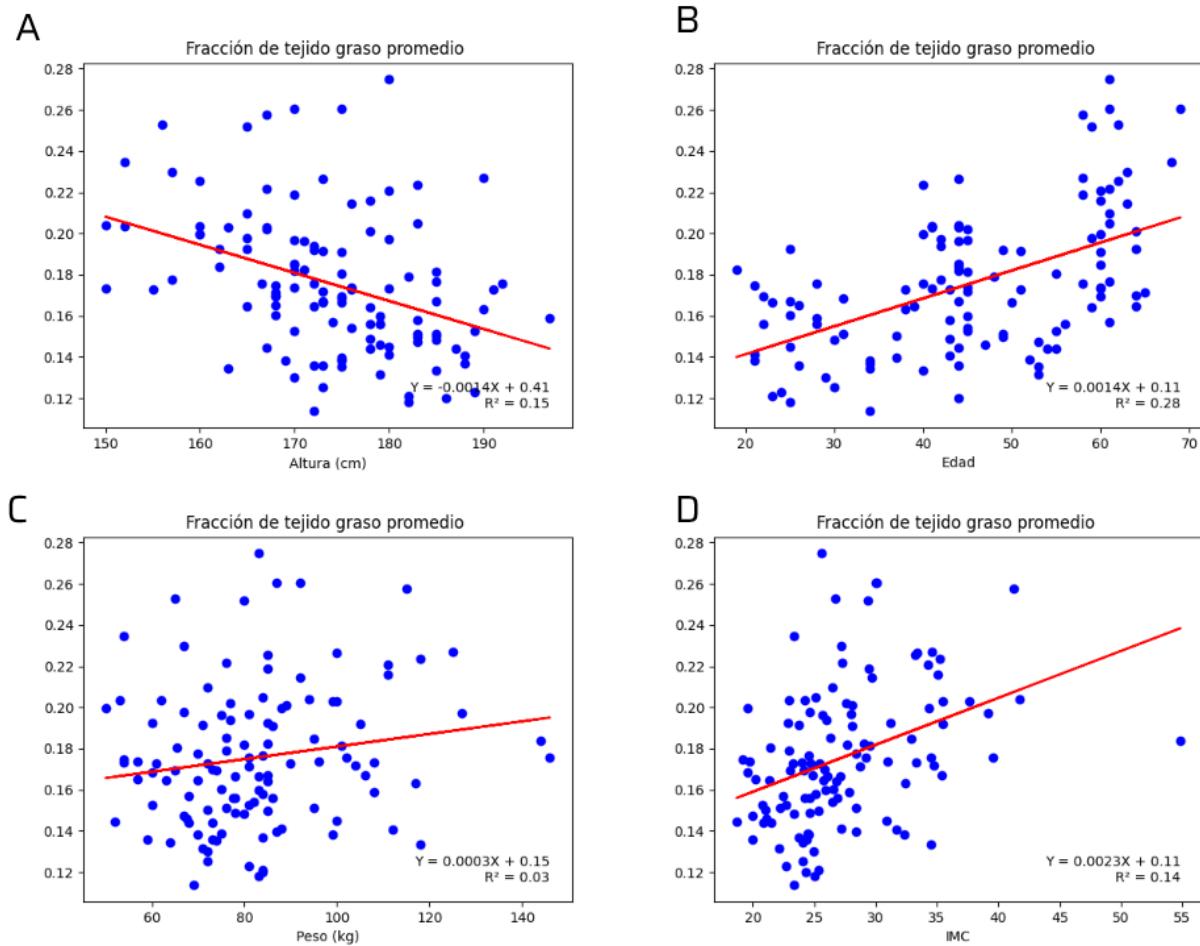
Una vez segmentadas y procesadas las imágenes, se calcularon estadísticas no paramétricas para cada métrica de salud muscular. A continuación, se presentan los resultados obtenidos para fracción de tejido graso y tamaño en términos de  $\text{cm}^3$ .

### 4.2.1 Fracción de tejido graso volumétrico

En la Figura 4.2, se presentan los resultados de fracción de tejido graso adquiridos con el modelo 3D, donde se puede observar que la infiltración de tejido graso es considerablemente mayor en el conjunto ES+M, mientras que para los demás músculos ronda cerca del 15%. Por otro lado, se presenta una infiltración significativamente mayor en todos los músculos izquierdos comparado con los derechos. Los valores p de la prueba de Mann-Whitney fueron los siguientes,  $p<0.01$  para los P, I y CL y  $p=0.02$  para los ES+M.



*Figura 4.2 Grafico de violín de los valores de FF para todos los sujetos, para cada músculo. En azul se grafican los músculos izquierdos, mientras que en rosa los derechos.*



*Figura 4.3. Regresiones lineales de fracción de tejido graso contra altura, edad, índice de masa corporal (IMC) y Peso. En círculos azules se muestran las muestras individuales, las rectas en rojo corresponden a la regresión lineal y debajo a la derecha de cada gráfico se muestran los valores de  $R^2$  de la regresión.*

Adicionalmente, calculamos regresiones lineales entre la fracción de tejido graso promedio (entre izquierda y derecha para cada individuo) de todos los músculos de la zona lumbar y las variables demográficas y antropométricas disponibles (altura, peso, edad e IMC) con el objetivo de encontrar variables asociadas a mayores niveles de infiltración grasa. En la Figura 4.3 se puede observar que los coeficientes de determinación ( $R^2$ ) de las regresiones lineales son muy bajos para todas las variables, siendo la edad el factor más influyente dentro de estas variables. Este resultado coincide con estudios previos que indicaron diferencias de infiltración grasa en este grupo muscular entre grupos de distintas edades, además de ser un patrón observado en el proceso de sarcopenia [23].

#### 4.2.1.1 Diferencias entre grupos muestrales

Se compararon los valores de fracción de tejido graso entre hombres y mujeres, y entre sujetos activos y sedentarios. Diagramas de violín para estas dos comparaciones se pueden observar en la Figura 4.4. Empleando pruebas de Mann-Whitney, se evaluó si había diferencias significativas entre ambos grupos, usando un  $p < 0.05$  para identificar diferencias significativas.

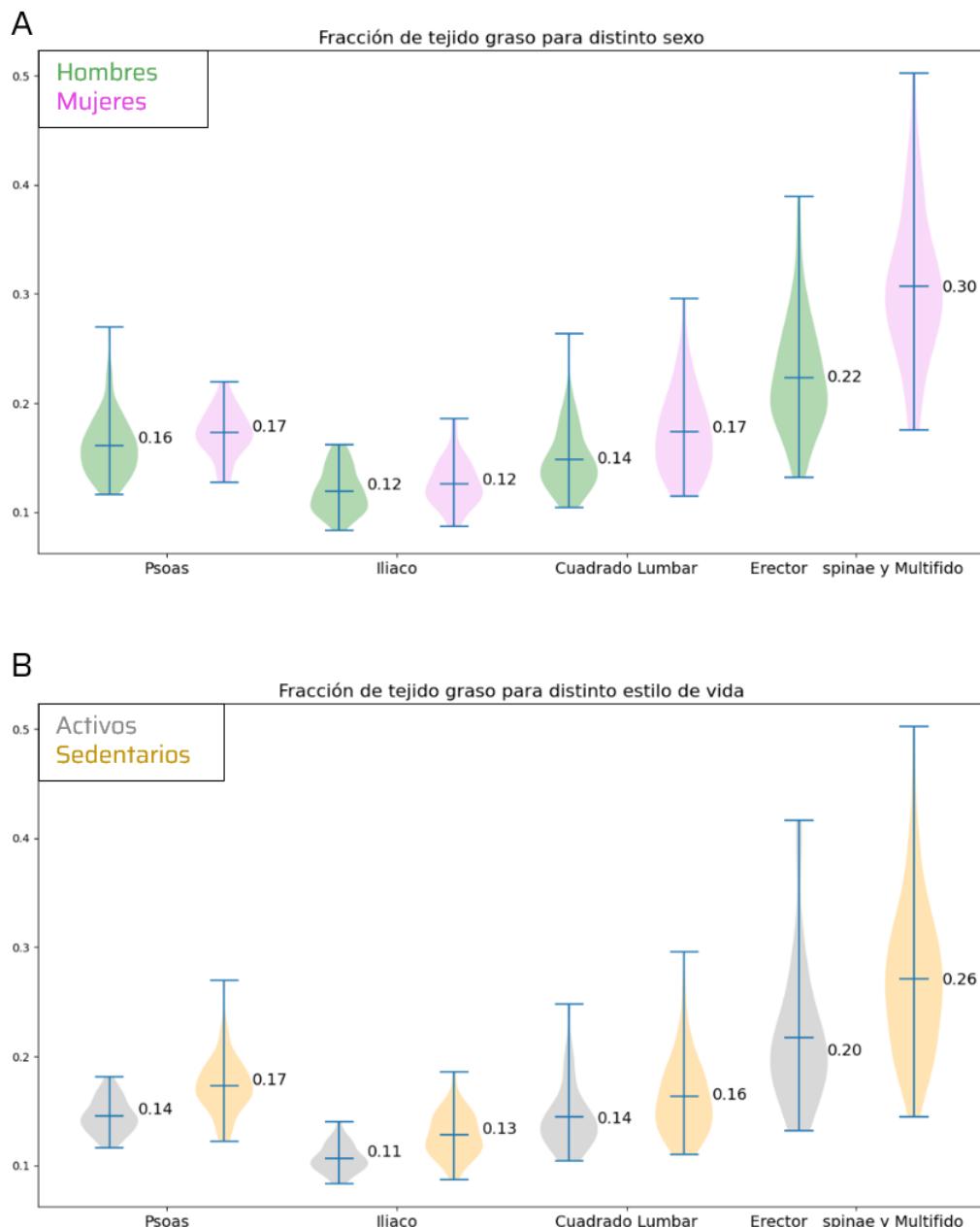


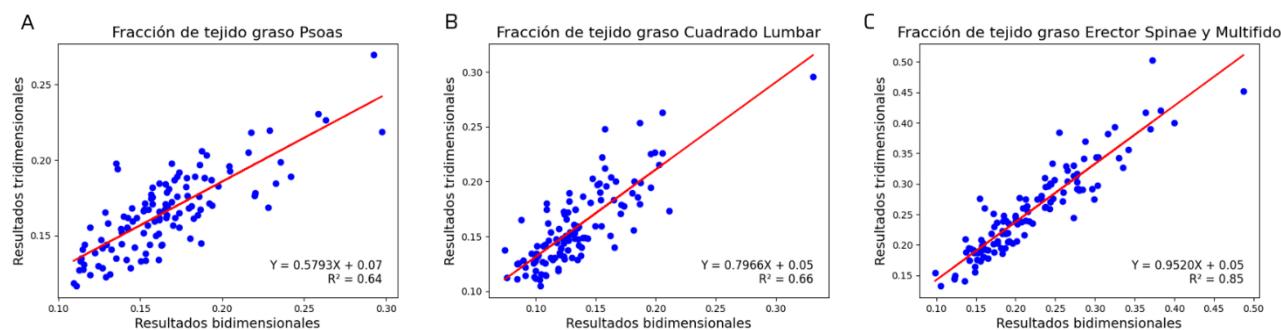
Figura 4.4. Gráficos violín para la fracción de tejido graso entre hombres y mujeres y entre sujetos activos y sedentarios. El color verde representa a los hombres y el violeta a las mujeres mientras que el gris a los sujetos activos y el naranja a los sedentarios.

Como se puede ver en la Figura 4.4, las mujeres mostraron mayor infiltración grasa en los músculos P ( $p<0.01$ ), CL ( $p<0.01$ ) y ES+M ( $p<0.01$ ), mientras que no hubo diferencias significativas para el Ilíaco ( $p=0.10$ ). Las diferencias para el conjunto ES+M han sido también encontradas en otros trabajos [23] [40].

A la hora de comparar la infiltración grasa entre los sujetos activos y sedentarios, estos últimos mostraron una mayor infiltración en todos los músculos ( $p<0.01$ ).

#### **4.2.1.2 Relación con resultados de sección transversal a la altura de L4**

Junto con los resultados volumétricos de fracción de tejido graso, también se calcularon los resultados para el corte axial a la altura de L4 utilizando el modelo de segmentación 2D. A partir de estos, se crearon regresiones lineales que vinculan los resultados 3D y 2D a la altura de altura L4 (Figura 4.5), con el objetivo de estudiar si usar métricas 2D como sustitutas de la métrica más precisa 3D es factible.

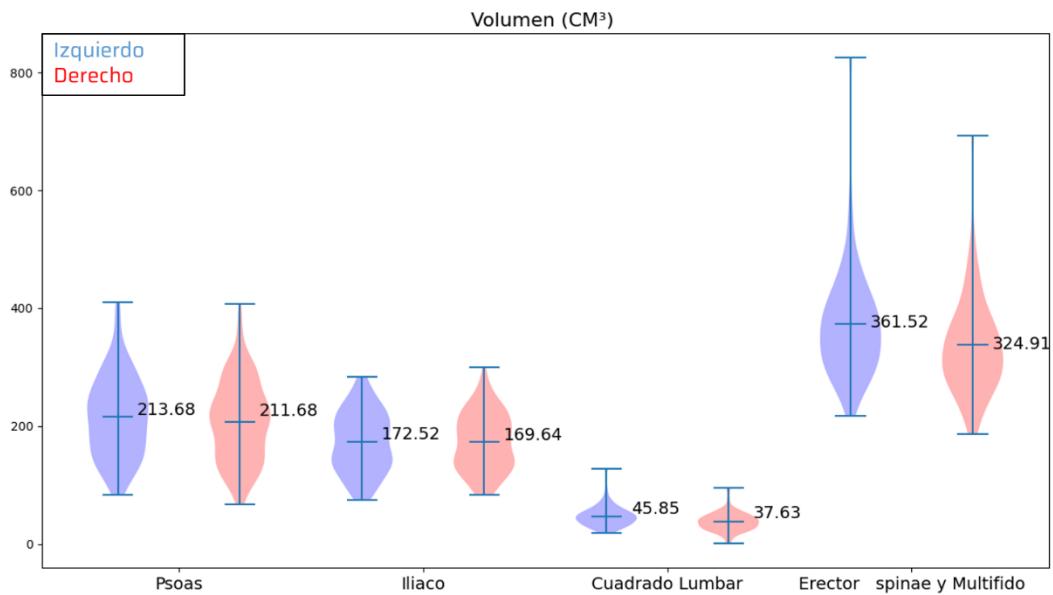


*Figura 4.5. Regresiones lineales entre resultados para la sección transversal a la altura de L4 y volumétricos. En círculos azules se muestran las muestras individuales, las rectas en rojo corresponden a la regresión lineal y debajo a la derecha de cada gráfico se muestran los valores de  $R^2$  de la regresión.*

Los resultados obtenidos en los ajustes lineales de la Figura 4.5 indican que hay una correlación moderada entre los resultados 2D y 3D para los Psoas ( $R^2=0.64$ ) y Cuadrados Lumbar (R<sup>2</sup>=0.66). Para el caso ES+M, las métricas 3D y 2D estuvieron altamente correlacionadas ( $R^2=0.85$ ), indicando que se podría predecir los valores de infiltración de tejido graso de todo el conjunto a partir de un único corte axial a la altura de L4.

#### **4.2.2 Tamaño de musculatura de zona lumbar**

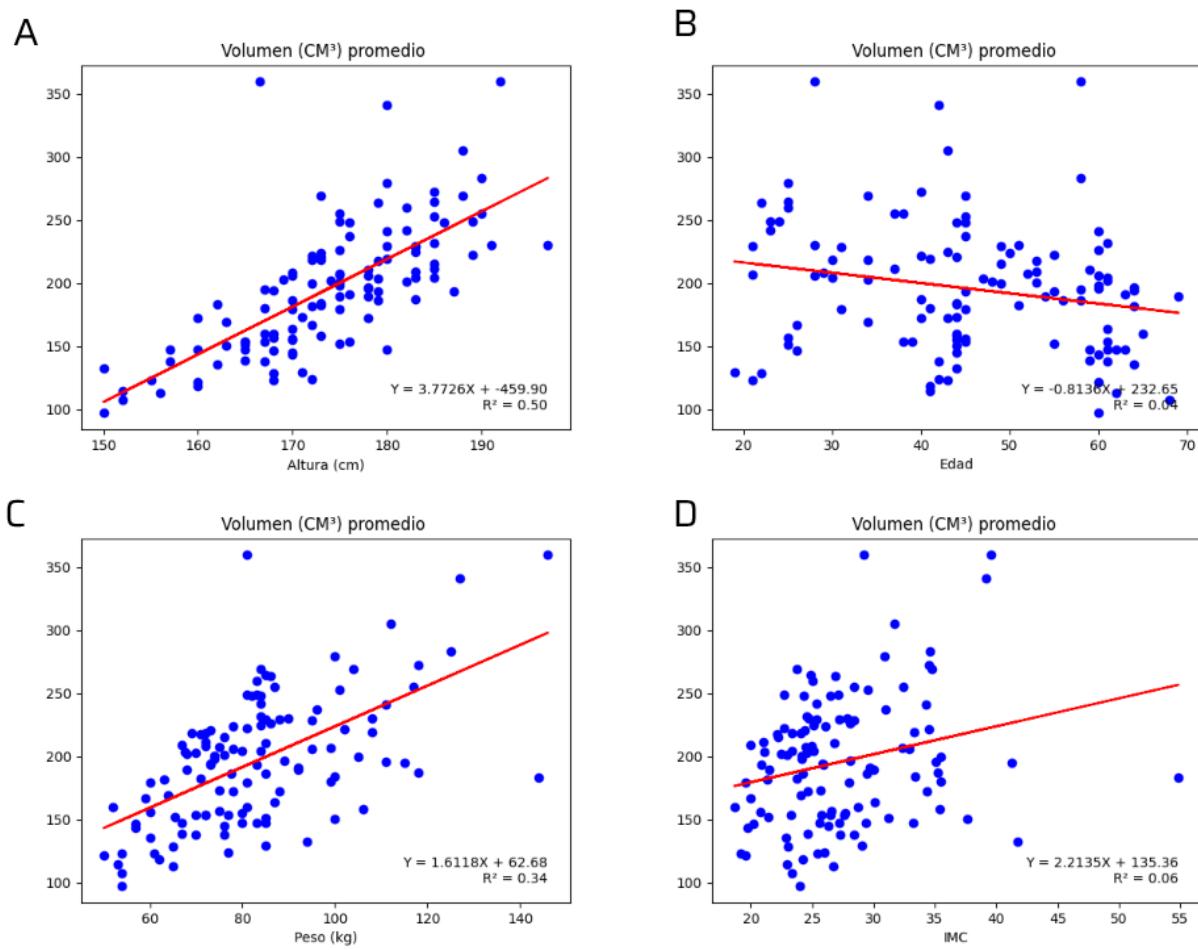
Al igual que punto anterior, se presentan los resultados del tamaño de la musculatura lumbar normalizado por el peso, obtenidos utilizando el modelo de segmentación 3D.



*Figura 4.6. Gráfico de violín de los valores del tamaño de cada uno de los músculos analizados para todos los sujetos. En azul se grafican los músculos izquierdos, mientras que en rosa los derechos.*

Los resultados obtenidos indican una paridad en cuanto al tamaño izquierdo y derecho para los músculos P ( $p=0.39$ ) e I ( $p=0.99$ ) mientras que existe una diferencia significativa para los músculos CL ( $p<0.01$ ) y el conjunto ES+M ( $p<0.01$ ) a favor del lado izquierdo. No obstante, estas diferencias pueden estar provocadas por pobres desempeños del modelo en algunas imágenes. Esto se debe a que los valores de tamaño son considerablemente más sensibles que los valores de fracción de tejido graso.

Nuevamente, se calcularon las regresiones lineales para los resultados del tamaño muscular con respecto a las variables antropométricas y demográficas de los sujetos.



*Figura 4.7. Regresiones lineales de tamaño contra altura, edad, índice de masa corporal (IMC) y Peso. En círculos azules se muestran las muestras individuales, las rectas en rojo corresponden a la regresión lineal y debajo a la derecha de cada gráfico se muestran los valores de  $R^2$  de la regresión.*

A diferencia de los resultados obtenidos en el punto anterior, los coeficientes de determinación son altos para las variables altura ( $R^2=0.50$ ) y peso ( $R^2=0.34$ ) (Figura 4.7). No es así para la edad, el cual es el factor más influente en cuanto a la infiltración de tejido graso, pero no para el tamaño.

#### **4.2.2.1 Diferencias entre grupos muéstros**

Al igual que antes, se realizaron pruebas de Mann-Whitney para determinar si había diferencias entre los grupos estudiados en términos de volumen muscular, normalizados por peso y luego por altura.

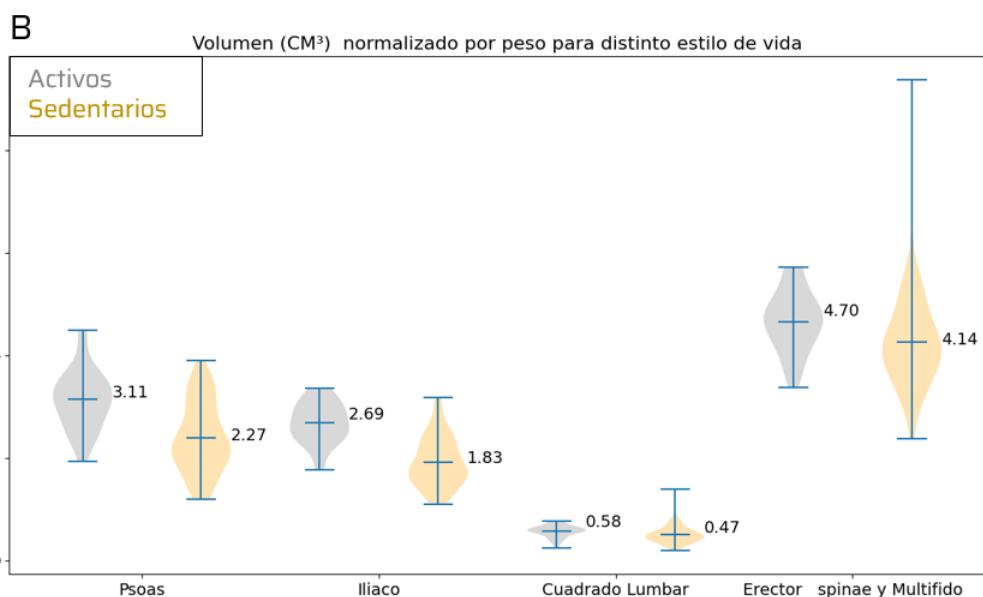
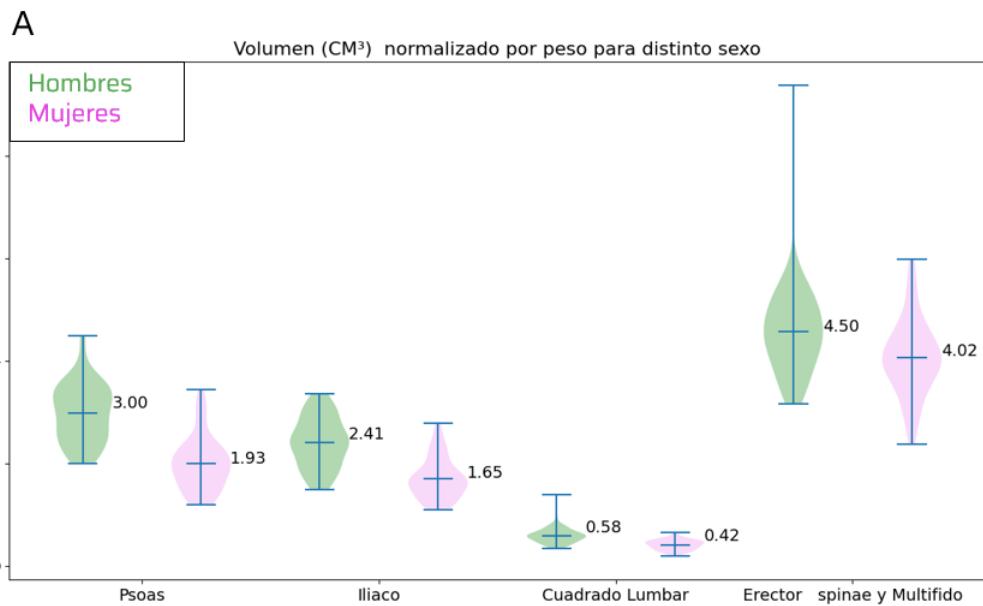


Figura 4.8. Gráficos violín para el volumen (cm<sup>3</sup>) normalizado por peso entre hombres y mujeres y entre sujetos activos y sedentarios. El color verde representa a los hombres y el violeta a las mujeres mientras que el gris a los sujetos activos y el naranja a los sedentarios.

En la Figura 4.8 encontramos que el tamaño de toda la musculatura lumbar normalizada por peso es significativamente mayor en hombres ( $p<0.01$ ) y en sujetos activos ( $p<0.01$ ).

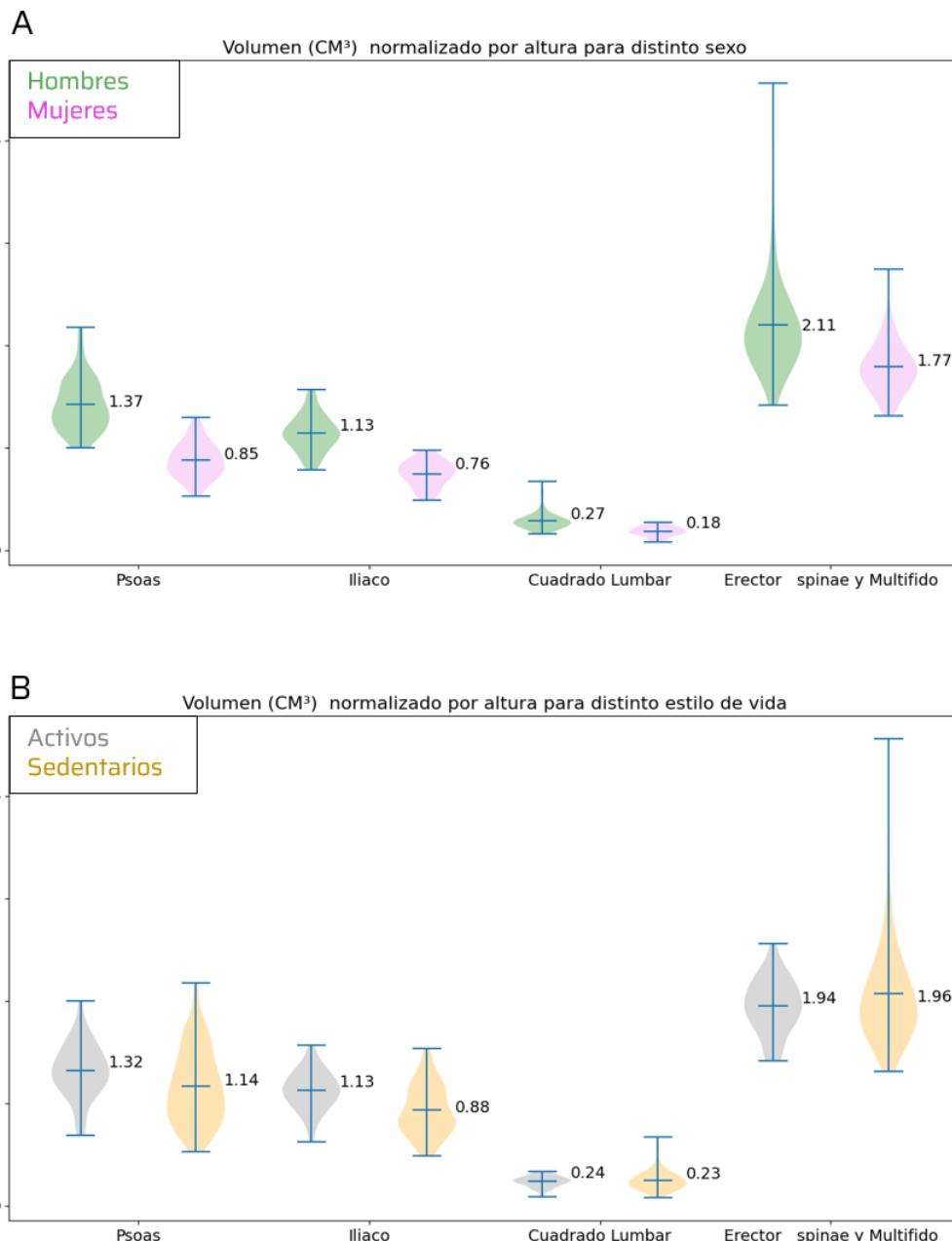


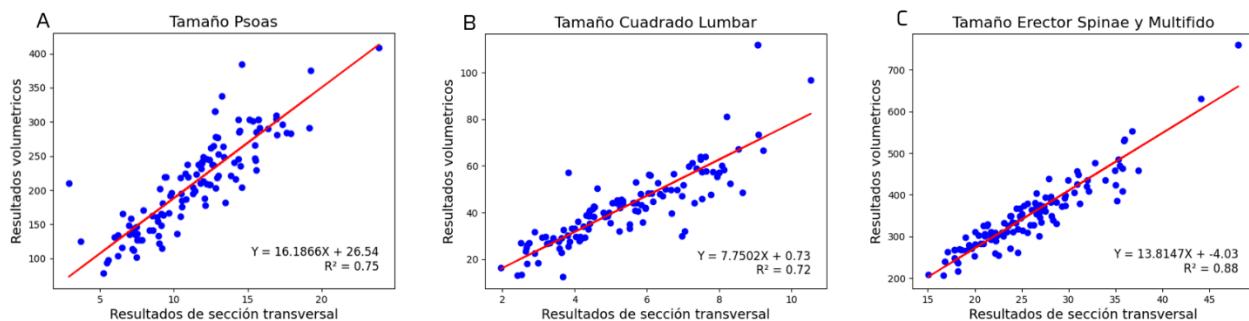
Figura 4.9. Gráficos violín para el volumen (cm<sup>3</sup>) normalizado por altura entre hombres y mujeres y entre sujetos activos y sedentarios. El color verde representa a los hombres y el violeta a las mujeres mientras que el gris a los sujetos activos y el naranja a los sedentarios.

En la Figura 4.9-A se presenta una situación similar a la anterior, los hombres poseen una musculatura normalizada por altura significativamente mayor ( $p<0.01$ ) en todos los músculos. No obstante, las diferencias entre tamaño normalizado por altura

para distintos estilos de vida no fueron significativas en los CL ( $p=0.76$ ) ni en los conjuntos ES+M ( $p=0.60$ ) mientras que si lo fueron para el P ( $p=0.015$ ) y el I ( $p<0.01$ ).

#### 4.2.2.2 Relación con resultados de sección transversal a la altura de L4

Por último, se obtuvieron resultados de tamaño en términos de sección de área transversal a la altura de L4 utilizando el modelo de segmentación 2D, y se realizó un ajuste lineal de estos resultados con los valores de los tamaños volumétricos.



*Figura 4.10. Regresiones lineales entre resultados de sección transversal a altura L4 y volumétricos. En círculos azules se muestran las muestras individuales, las rectas en rojo corresponden a la regresión lineal y debajo a la derecha de cada gráfico se muestran los valores de  $R^2$  de la regresión.*

En la Figura 4.10 se presentan las regresiones lineales entre los resultados de tamaño 2D y 3D. En comparación con los resultados presentados en el punto 4.2.1.2, los coeficientes de determinación son más altos en todos los músculos (un  $R^2=0.75$  para el P y  $R^2=0.72$  para el CL) y su magnitud indicaría que es viable predecir el volumen de los músculos utilizando la ecuación lineal estimada y el área a la altura de L4, principalmente cuando se evalúa el conjunto ES+M ( $R^2=0.88$ ).

## **5. Discusión y conclusiones**

En este último capítulo se presenta la discusión de los resultados obtenidos en la sección anterior y las conclusiones de todo el trabajo.

### **5.1 Discusión**

Si bien existen múltiples trabajos que utilizan redes neuronales para la segmentación automática de los músculos de zona lumbar de manera 2D, este es uno de los primeros que también implementa algoritmos de segmentación 3D y obtiene buenos resultados. Por otro lado, la posibilidad de evaluar automáticamente la calidad muscular en términos de fracción de tejido graso y tamaño a partir de estos algoritmos, permite el estudiar de manera rápida grandes grupos muestrales (como el trabajo realizado en la sección anterior) con el fin de obtener valores de referencia.

Comenzando con los resultados en términos de puntaje Dice obtenidos en el set de validación por los modelos entrenados, encontramos que las segmentaciones realizadas por el modelo 2D de máximo desempeño fueron mejores que las del modelo 3D en todos los músculos presentes en el corte axial. Si comparamos el desempeño del modelo 2D contra el de otros trabajos, encontramos que los puntajes Dice son tan buenos o mejores en los músculos coincidentes (se debe a que muchos estudios no involucran los cuadrados lumbares u optan por segmentar el Erector Spinae y Multifido de manera separada) [41] [42] [43]. Por parte del modelo 3D, el desempeño solo puede ser comparado con el presentado por Wesselink et al. el cual es un modelo basado en la arquitectura UNet 3D capaz de segmentar los Psoas, Erector Spinae y los Multifidos de manera separada [43]. Teniendo en cuenta esto encontramos que el desempeño de nuestro modelo es ligeramente peor a la hora de segmentar los Psoas, pero en la contracara, tiene la capacidad de segmentar los cuadrados lumbares e iliacos. Para finalizar, las diferencias de desempeño con el modelo de Wesselink se pueden deber a que el modelo entrenado en este trabajo segmenta más etiquetas y de esta manera se deteriora el desempeño particular en cada una de estas o a una escasez de imágenes manualmente segmentadas en el set de datos. Es por esto último que se sugiere continuar con el trabajo segmentando mayor cantidad de imágenes permitiendo crear un set de testeo y obtener un modelo de mejor desempeño y más fiable.

Por parte de los resultados poblacionales que obtuvimos en la sección anterior, podemos decir que existen diferencias significativas de infiltración de tejido graso entre

los músculos izquierdos y derechos de zona lumbar siendo los izquierdos los que presentan mayor infiltración en todos los músculos. Por otro lado, también se encontró que la infiltración de tejido graso es significativamente mayor para las mujeres y las personas sedentarias. Teniendo en cuenta las variables antropométricas que se disponían, se observó que la infiltración de tejido graso está más vinculada con la edad que con el peso, la altura o el índice de masa corporal lo cual coincide con lo presentado en múltiples trabajos [10] [24].

En cuanto a los resultados obtenidos para el tamaño de los músculos de zona lumbar encontramos que los izquierdos son ligeramente más grandes que los derechos, que los músculos normalizados por peso y por altura son más grandes para los hombres que para las mujeres en toda la musculatura y que en términos generales, ocurre lo mismo entre sujetos sedentarios y activos siendo los segundos los que cuentan con una musculatura de mayor tamaño.

Para finalizar, se puede decir que el buen desempeño de ambos modelos en términos de puntaje Dice implican una buena precisión a la hora de calcular los valores de calidad muscular. Por otro lado, es factible realizar una evaluación del conjunto ES+M en términos de tamaño y fracción de tejido graso en un corte axial de altura L4 y estimar el valor de todo el conjunto en la zona lumbar. Esto no ocurre para los Psoas y Cuadrados Lumbares que, si bien tienen una gran correlación entre los resultados de sección transversal y volumétricos, estimar uno a partir del otro no sería fiable.

## 5.2 Conclusiones

En este trabajo se logró crear herramientas capaces de evaluar automáticamente la calidad muscular de los músculos de la zona lumbar en términos de tamaño e infiltración de tejido graso de manera volumétrica y de sección de área transversal a la altura de L4 a partir de algoritmos de *machine learning*. Las herramientas desarrolladas están basadas en RNC de segmentación de imágenes con arquitectura U-Net 2D y 3D. Las medianas (IQR) del modelo 2D fueron, en términos de puntajes Dice, de 0.951 (0.035) para el Psoas, 0.929 (0.043) para el Cuadrado Lumbar y 0.962 (0.018) para el conjunto Erector Spinae más Multifido. Por parte del modelo 3D, fueron de 0.921 (0.024) para el Psoas, 0.881 (0.031) para el Iliaco, 0.847 (0.044) para el Cuadrado lumbar y 0.926 (0.010) para el conjunto Erector Spinae más Multífido.

Estos modelos fueron entrenados utilizando un set de datos manualmente segmentado compuesto de 58 imágenes Dixon In-Phase 2D en altura L4 y 26 imágenes Dixon In-Phase 3D que abarcan desde la primera vértebra lumbar hasta el trocánter menor del fémur. Estas imágenes fueron manualmente segmentadas utilizando la herramienta de creación de máscaras de *Simpleware*, luego registradas y en el caso de las imágenes 3D submuestreadas a un cuarto de su tamaño original. Debido a que se contaba con muy pocas imágenes, estas fueron aumentadas utilizando transformaciones lineales y no lineales para crear una gran cantidad de imágenes artificiales. Por otro lado, los modelos finales de este trabajo fueron entrenados de manera local con una placa de video Nvidia RTX 3090 de 24 GB de VRAM mientras que los experimentos y evaluaciones de desempeño de los modelos 2D fueron entrenados con una placa de video Nvidia RTX 2060 de 6 GB de VRAM.

A partir de estas herramientas, se procesaron las imágenes de 118 sujetos que formaron parte de un estudio para analizar el impacto de la actividad física en la salud muscular en el Reino Unido. A partir de este procesamiento se obtuvieron los valores de tamaño e infiltración de tejido graso de manera 3D y 2D en altura L4 y en conjunto con los datos antropométricos y realizando pruebas de Mann-Whitney se obtuvieron los siguientes resultados poblacionales.

En cuanto a la fracción de tejido graso, se encuentran infiltraciones significativamente mayores en las mujeres que en los hombres en todos los músculos ( $p<0.01$ ) excepto el Iliaco ( $p=0.10$ ). Si se compara a los sujetos sedentarios con los ciclistas encontramos nuevamente infiltraciones de tejido graso significativamente superiores para el primer grupo en todos los músculos ( $p<0.01$ ). Por otro lado, al intentar relacionar la fracción de tejido graso con las variables antropométricas de edad, altura, peso e IMC, encontramos coeficientes de determinación bajos ( $R^2<0.3$ ) pero, de todas estas variables, la infiltración se encuentra más bien vinculada con la edad ( $R^2=0.28$ ). Por ultimo al vincular los valores 3D con los valores 2D de infiltración de tejido graso, se concluyó que es factible realizar una evaluación en un corte axial a la altura de L4 que sea equivalente a una evaluación de todo el tejido para el grupo ES + M ( $R^2=0.85$ ).

Finalmente, en cuanto al tamaño de la musculatura lumbar, se encontró que el volumen normalizado por peso es significativamente mayor en hombres ( $p<0.01$ ) y en sujetos activos ( $p<0.01$ ) en todos los músculos. Si se normaliza el tamaño por la altura encontramos que el volumen vuelve a ser significativamente mayor en hombres para todos los músculos ( $p<0.01$ ) pero no ocurre lo mismo cuando se compara ciclistas con

sedentarios. En este segundo caso encontramos que no hay diferencias significativas en los CL ( $p=0.76$ ) ni en los conjuntos ES+M ( $p=0.60$ ), mientras que sí las hubo en los músculos P ( $p=0.015$ ) y el I ( $p<0.01$ ). Al comparar las métricas antropométricas con los resultados, obtenemos que hay una correlación considerable entre la altura y el tamaño del músculo ( $R^2=0.5$ ) y una relación no tan directa entre el peso y el tamaño ( $R^2=0.34$ ). Para terminar, al vincular los valores 3D con los valores 2D, se puede decir que es factible calcular el tamaño volumétrico del conjunto ES+M a partir de un corte axial a la altura de L4 ( $R^2=0.88$ ) mientras que los otros músculos obtienen coeficientes de determinación cercanos a 0.8.

En conclusión, en este trabajo presentamos una novedosa herramienta para evaluar de forma automática la calidad de los músculos de la zona lumbar a partir de imágenes de resonancia magnética. Adicionalmente, empleando esta herramienta se obtuvieron resultados originales de valores de referencia de tamaño e infiltración graso para este grupo muscular, los cuales contribuirán a la investigación clínica y a comprender mejor las características de estos. Por último, pudimos determinar que evaluar estos músculos a partir un único corte axial como se hace muchas veces en la práctica, solo es adecuado para el grupo ES+M.

## Bibliografía

- [1] Shi, L., Yan, B., Jiao, Y., Chen, Z., Zheng, Y., Lin, Y., & Cao, P. (2022). Correlation between the fatty infiltration of paraspinal muscles and disc degeneration and the underlying mechanism. *BMC Musculoskeletal Disorders*, 23(1), 509. <https://doi.org/10.1186/s12891-022-05466-8>
- [2] Willard, F. H., Vleeming, A., Schuenke, M. D., Danneels, L., & Schleip, R. (2012). The thoracolumbar fascia: anatomy, function and clinical considerations. *Journal of Anatomy*, 221(6), 507-536. <https://doi.org/10.1111/j.1469-7580.2012.01511.x>
- [3] Mandelli, F., Nüesch, C., Zhang, Y., Halbeisen, F., Schären, S., Mündermann, A., & Netzer, C. (2021). Assessing Fatty Infiltration of Paraspinal Muscles in Patients With Lumbar Spinal Stenosis: Goutallier Classification and Quantitative MRI Measurements. *Frontiers in Neurology*, 12. <https://doi.org/10.3389/fneur.2021.656487>
- [4] Belzunce, M. A., Henckel, J., Laura, A. di, Horga, L. M., & Hart, A. J. (2023). Mid-life cyclists preserve muscle mass and composition: a 3D MRI study. *BMC Musculoskeletal Disorders*, 24(1), 209. <https://doi.org/10.1186/s12891-023-06283-3>
- [5] Belzunce, M. A., Henckel, J., di Laura, A., Horga, L. M., & Hart, A. J. (2023). Similarities and differences in skeletal muscle and body composition between sexes: an MRI study of recreational cyclists. *BMJ Open Sport & Exercise Medicine*, 9(3), e001672. <https://doi.org/10.1136/bmjsem-2023-001672>
- [6] Belzunce, M. A., Henckel, J., di Laura, A., & Hart, A. (2021). Intramuscular fat in gluteus maximus for different levels of physical activity. *Scientific Reports*, 11(1), 21401. <https://doi.org/10.1038/s41598-021-00790-w>
- [7] Belzunce, M. A., Henckel, J., di Laura, A., & Hart, A. J. (2022). Reference values for volume, fat content and shape of the hip abductor muscles in healthy individuals from Dixon MRI. *NMR in Biomedicine*, 35(2). <https://doi.org/10.1002/nbm.4636>
- [8] Belzunce, M. A., Henckel, J., Fotiadou, A., di Laura, A., & Hart, A. (2020). Automated multi-atlas segmentation of gluteus maximus from Dixon and T1-weighted magnetic resonance images. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 33(5), 677-688. <https://doi.org/10.1007/s10334-020-00839-3>
- [9] D'hooge, R., Cagnie, B., Crombez, G., Vanderstraeten, G., Dolphens, M., & Danneels, L. (2012). Increased intramuscular fatty infiltration without differences in lumbar muscle cross-sectional area during remission of unilateral recurrent low back pain. *Manual Therapy*, 17(6), 584-588. <https://doi.org/10.1016/j.math.2012.06.007>
- [10] McGregor, R. A., Cameron-Smith, D., & Poppitt, S. D. (2014). It is not just muscle mass: a review of muscle quality, composition and metabolism during ageing as determinants of muscle function and mobility in later life. *Longevity & Healthspan*, 3(1), 9. <https://doi.org/10.1186/2046-2395-3-9>

- [11] Burian, E., Rohrmeier, A., Schlaeger, S., Dieckmeyer, M., Diefenbach, M. N., Syväri, J., Klupp, E., Weidlich, D., Zimmer, C., Rummeng, E. J., Karampinos, D. C., Kirschke, J. S., & Baum, T. (2019). Lumbar muscle and vertebral bodies segmentation of chemical shift encoding-based water-fat MRI: the reference database MyoSegmenTUM spine. *BMC Musculoskeletal Disorders*, 20(1), 152. <https://doi.org/10.1186/s12891-019-2528-x>
- [12] Takashima, H., Takebayashi, T., Ogon, I., Yoshimoto, M., Morita, T., Imamura, R., Nakanishi, M., Nagahama, H., Terashima, Y., & Yamashita, T. (2018). Analysis of intra and extramyocellular lipids in the multifidus muscle in patients with chronic low back pain using MR spectroscopy. *The British Journal of Radiology*, 20170536. <https://doi.org/10.1259/bjr.20170536>
- [13] Jiménez González, M. M., Pons Serra, M., Castaño Moreno, C., Monés Jiménez, L., & Martínez Rodenas, F. (2002). Anales de medicina interna. In *Anales de Medicina Interna* (Vol. 19, Issue 9). Arán Ediciones, S.A. [https://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S0212-71992002000900009&lng=es&nrm=iso&tlang=es](https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0212-71992002000900009&lng=es&nrm=iso&tlang=es)
- [14] Faur, C., Patrascu, J. M., Haragus, H., & Anglitoiu, B. (2019). Correlation between multifidus fatty atrophy and lumbar disc degeneration in low back pain. *BMC Musculoskeletal Disorders*, 20(1), 414. <https://doi.org/10.1186/s12891-019-2786-7>
- [15] Li, H., Luo, H., & Liu, Y. (2019). Paraspinal Muscle Segmentation Based on Deep Neural Network. *Sensors*, 19(12), 2650. <https://doi.org/10.3390/s19122650>
- [16] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation.
- [17] Siccardi, M. A., Tariq, M. A., & Valle, C. (2023). Anatomy, Bony Pelvis and Lower Limb: Psoas Major.
- [18] Bordoni, B., & Varacallo, M. (2023). Anatomy, Bony Pelvis and Lower Limb, Iliopsoas Muscle.
- [19] Bordoni, B., & Varacallo, M. (2023). Anatomy, Abdomen and Pelvis, Quadratus Lumborum.
- [20] Henson, B., Kadiyala, B., & Edens, M. A. (2023). Anatomy, Back, Muscles. [21]
- [21] J. Markle, «Centeno Schultz Clinic,» 1 4 2018. [En línea]. Available: <https://centenoschultz.com/multifidus-muscle-spinal-stabilizer/>. [Último acceso: 2 12 2023].
- [22] P. collaborators, «Physiopedia,» [En línea]. Available: [https://www.physiopedia.com/Quadratus\\_Lumborum](https://www.physiopedia.com/Quadratus_Lumborum). [Último acceso: 2 12 2023].
- [23] Crawford, R. J., Filli, L., Elliott, J. M., Nanz, D., Fischer, M. A., Marcon, M., & Ulbrich, E. J. (2016). Age- and Level-Dependence of Fatty Infiltration in Lumbar Paravertebral Muscles of Healthy Volunteers. *American Journal of Neuroradiology*, 37(4), 742-748. <https://doi.org/10.3174/ajnr.A4596>
- [24] Sakai, Y., Matsui, H., Ito, S., Hida, T., Ito, K., Koshimizu, H., & Harada, A. (2017). Sarcopenia in elderly patients with chronic low back pain. *Osteoporosis and Sarcopenia*, 3(4), 195-200. <https://doi.org/10.1016/j.afos.2017.09.001>

- [25] Lins, C. F., Salmon, C. E. G., & Nogueira-Barbosa, M. H. (2021). Applications of the Dixon technique in the evaluation of the musculoskeletal system. *Radiologia Brasileira*, 54(1), 33–42. <https://doi.org/10.1590/0100-3984.2019.0086>
- [26] S. Shapiro, Encyclopedia of artificial Intelligence second edition, New Jersey: Wiley, 1992
- [27] Ramesh, A. N., Kambhampati, C., Monson, J. R. T., & Drew, P. J. (2004). Artificial intelligence in medicine. *Annals of the Royal College of Surgeons of England*, 86(5), 334–338. <https://doi.org/10.1308/147870804290>
- [28] F. Chollet, Deep Learning with Python, Manning, 2017.
- [29] I. Neutelings, «TikZ.net,» [En línea]. Available: [https://tikz.net/neural\\_networks/](https://tikz.net/neural_networks/).
- [30] J. Fu y R. Ng, «Deep Learning Wizard,» [En línea]. Available: [https://www.deeplearningwizard.com/deep\\_learning/boosting\\_models\\_pytorch/lr\\_scheduling/](https://www.deeplearningwizard.com/deep_learning/boosting_models_pytorch/lr_scheduling/). [Último acceso: 16 10 2023].
- [31] Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- [32] O. Ian, «Convolution Filters,» Medium, 9 5 2020. [En línea]. Available: <https://medium.com/@ianormy/convolution-filters-4971820e851f>. [Último acceso: 2023 10 19].
- [33] V. Dumoulin, «Getting started with CNNs by calculating LeNet-Layer manually,» philschmid, 27 2 2020. [En línea]. Available: <https://www.philschmid.de/getting-started-with-cnn-by-calculating-lenet-layer-manually>. [Último acceso: 2023 10 19].
- [34] Raina, V., Molchanova, N., Graziani, M., Malinin, A., Muller, H., Cuadra, M. B., & Gales, M. (2023). Tackling Bias in the Dice Similarity Coefficient: Introducing nDSC for White Matter Lesion Segmentation.
- [35] Ibtehaz, N., & Rahman, M. S. (2020). MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121, 74–87. <https://doi.org/10.1016/j.neunet.2019.08.025>
- [36] Bousias Alexakis, E., & Armenakis, C. (2020). Evaluation of unet and unet++ architectures in high resolution image change detection applications. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2020, 1507–1514. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-1507-2020>
- [37] Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization.
- [38] Ruder, S. (2016). An overview of gradient descent optimization algorithms.
- [39] Li, H., Wang, Y., Hong, Y., Li, F., & Ji, X. (2023). Layered mixed-precision training: A new training method for large-scale AI models. *Journal of King Saud University - Computer and Information Sciences*, 35(8), 101656. <https://doi.org/10.1016/j.jksuci.2023.101656>

- [40] Yang, Q., Yan, D., Wang, L., Li, K., Liang, W., Zhang, W., Liu, Y. D., Li, X. M., Blake, G. M., Konerth, N., Cheng, X., Tian, W., & Hind, K. (2022). Muscle fat infiltration but not muscle cross-sectional area is independently associated with bone mineral density at the lumbar spine. *The British Journal of Radiology*, 95(1134). <https://doi.org/10.1259/bjr.20210371>
- [41] Shen, H., Huang, J., Zheng, Q., Zhu, Z., Lv, X., Liu, Y., & Wang, Y. (2021). A Deep-Learning-Based, Fully Automated Program to Segment and Quantify Major Spinal Components on Axial Lumbar Spine Magnetic Resonance Images. *Physical Therapy*, 101(6). <https://doi.org/10.1093/ptj/pzab041>
- [42] Giaccone, P., D'Antoni, F., Russo, F., Volpecina, M., Mallio, C. A., Papalia, G. F., Vadalà, G., Denaro, V., Vollero, L., & Merone, M. (2023). Fully automated evaluation of paraspinal muscle morphology and composition in patients with low back pain. *Intelligence-Based Medicine*, 100130. <https://doi.org/10.1016/j.ibmed.2023.100130>
- [43] Wesselink, E. O., Elliott, J. M., Coppieters, M. W., Hancock, M. J., Cronin, B., Pool-Goudzwaard, A., & Weber II, K. A. (2022). Convolutional neural networks for the automatic segmentation of lumbar paraspinal muscles in people with low back pain. *Scientific Reports*, 12(1), 13485. <https://doi.org/10.1038/s41598-022-16710-5>