ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE CIENCIAS

DESARROLLO DE ALGORITMOS PARA EL ANÁLISIS Y PROCESAMIENTO DE IMÁGENES DE RESONANCIA MAGNÉTICA PARA DIFERENCIAR LOS TEJIDOS PRESENTES EN EL HÍGADO GRASO

TRABAJO PREVIO A LA OBTENCIÓN DEL TÍTULO DE FÍSICA

PROYECTO DE INVESTIGACIÓN

ALEJANDRA ESTEFANÍA ICHINA LÓPEZ

alejandra.ichina@epn.edu.ec

DIRECTOR: NICOLÁS ALEJANDRO VÁSQUEZ PAZMIÑO, PH.D

nicolas.vasquez@epn.edu.ec

Quito, septiembre de 2021

Abstract

In recent years, advancement and development of different imaging techniques like X-Rays, Computed Tomography Scan, Ultrasound and Magnetic Resonance (MR) has facilitated the diagnosis of diseases. Due to its high resolution, MR imaging has gained prominence in the observation and diagnosis of multiple conditions that occur in important organs such as the liver. Consequently, the use of MR imaging increases and the data sets collected grow quickly, making the analysis more difficult and extensive. For this reason, it is necessary to implement methods that improve diagnostic accuracy and decrease analysis time. In this project, we use an approach based on Machine Learning to differentiate zones in the liver with fatty tissue from normal tissue, with images obtained from MR scanners. This problem is solved within the category of supervised learning, whose algorithm is implemented using multi-layer perceptron-type neural networks with texture radiomic characteristics as input data. The regions of interest of the liver are segmented using 3D-Slicer, the characteristics are extracted using PyRadiomics and the neural networks are built from the Scikit-Learn library. Our best results guarantee shorter analysis with an accuracy of about 70 %. However, precision is about 60 % which causes that the number of false positives (FP) and false negatives (FN) must be analyzed carefully with respect to the number of true positives (TP) and true negatives (TN) obtained.

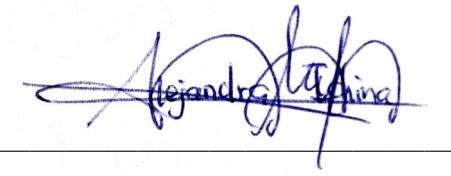
Resumen

En los últimos años, el avance y desarrollo de diferentes técnicas de adquisición de imágenes como Rayos X, Tomografía Computarizada, Ultrasonido y Resonancia Magnética (RM) ha facilitado el diagnóstico de enfermedades. Debido a su alta resolución, la imagen de RM ha ganado protagonismo en la observación y diagnóstico de múltiples condiciones que se presentan en órganos importantes como el hígado. En consecuencia, cada vez se realizan más imágenes por RM y la cantidad de datos se incrementa rápidamente, haciendo su análisis más difícil y extenso. Por esta razón, es necesario implementar métodos que mejoren la precisión del diagnóstico y disminuyan los tiempos de análisis. En este trabajo se usa herramientas basadas en Machine Learning para diferenciar las zonas del hígado con tejido graso y normal, con imágenes de RM. Este problema se resuelve dentro de la categoría de aprendizaje supervisado de clasificación, cuyo algoritmo es implementado usando redes neuronales de tipo perceptrón multicapa con características radiómicas de textura como datos de entrada. Las regiones de interés del hígado son segmentadas empleando el software 3D-Slicer, las características se extraen con PyRadiomics y las redes neuronales se construyen usando Scikit-Learn. Los mejores resultados obtenidos garantizan tiempos de análisis cortos con una exactitud de alrededor del 70 %. Sin embargo, los resultados de precisión alcanzan un 60 % provocando así que el número de falsos positivos (FP) y falsos negativos (FN) deban ser analizados cuidadosamente respecto al número de verdaderos positivos (VP) y verdaderos negativos (VN) obtenidos

Declaración de autoría

Yo, Alejandra Estefanía Ichina López, declaro bajo juramento que el trabajo aquí escrito es de mi autoría, que no ha sido previamente presentado para ningún grado o calificación profesional y que he consultado todas las referencias bibliográficas citadas en este documento.

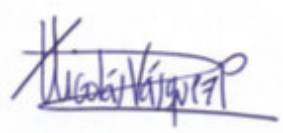
A través de la presente declaración cedo mis derechos de propiedad intelectual, correspondientes a este trabajo, a la Escuela Politécnica Nacional según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y normativa institucional vigente.



Alejandra Estefanía Ichina López

Certificación

Certifico que el presente trabajo fue desarrollado por ALEJANDRA ESTEFANÍA ICHINA LÓPEZ, bajo mi supervisión



Nicolás Vásquez Pazmiño, Ph.D
Director del Proyecto

Agradecimientos

Agradezco profundamente a mis padres que con su apoyo incondicional y perseverancia por salir adelante me motivaron a cumplir cada una de mis metas y me dieron las herramientas necesarias para emprender mi camino en la ciencia. A mis hermanos, Danilo y Alex por ser mi orgullo y ejemplo a seguir, pero sobre todo por acompañarme en cada etapa de la vida y permitirme disfrutarla al máximo.

A Esteban, Yessenia, Pedro y Daniela, mis primeros amigos en la carrera de Física, que con sus ocurrencias vivimos y gozamos los mejores momentos. Sin ustedes la carrera no hubiese sido esta montaña rusa de emociones.

A todos los amigos que conocí en el transcurso de este camino, en especial a Edison, Carlos y Ricardo. Su amistad me permitió sobrellevar momentos difíciles pero sobre todo comprender que un verdadero amigo siempre se alegrará de tus logros y compartirá los suyos.

A mis amigos de Ambato, Michelle, Gabriel y Álvaro. Gracias por tantos años de amistad y por comprender que, a pesar de la distancia siempre estaremos para apoyarnos y aconsejarnos. Sus palabras de admiración me han impulsado a emprender nuevos retos.

A todos mis profesores de la carrera de Física, en especial a Nicolás que más allá de ser mi director de tesis y profesor, fue un amigo. Siempre estaré agradecida por la confianza que pusiste en mi y que me permitió llevar a cabo este trabajo. Tu labor con la sociedad es admirable y llevaré presente cada uno de tus consejos con el fin de mejorar el nivel educativo de nuestro país. También agradezco infinitamente al Ph.D Eduardo Moreno y al grupo de física médica de la Benemérita Universidad Autónoma de Puebla (BUAP) que me dieron la oportunidad de llevar a cabo esta tesis durante mi estancia de investigación.

A Diego, por ser el mejor compañero de laboratorio, amigo y le petit ami. Gracias por las enseñanzas y el tiempo compartido. No hay palabras suficientes para agradecerte por el apoyo brindado en mi etapa más difícil. Fue tu comprensión, amor y paciencia lo que me ayudó a sanar y mejorar en cada aspecto de mi vida. Te admiro enormemente y eres mi inspiración para trazar mi camino como Física.

DEDICATORIA

A mis abuelitos

A mis padres, Víctor e Isabel

A mis hermanos, Danilo, Alex y Emily

A mi apoyo incondicional, Diego

A todas las personas interesadas en la ciencia

Índice general

Abstract I

Resumen II

Declaración de autoría III

Certificación IV

Lista de Figuras IX

Lista de Tablas XII

Introducción 1

1 Antecedentes 2

1.1. Técnicas de adquisición de imágenes médicas 3

1.1.1. Rayos X 3

1.1.2. Tomografía Computarizada (TC) 4

1.1.3. Ultrasonido (US) 5

1.1.4. Imagen por Resonancia Magnética (IRM) 5

1.2. Análisis de IRM 8

1.3. El rol de IRM en el hígado 9

1.4. Técnicas de segmentación hepática 10

1.4.1. Segmentación hepática manual 10

1.4.2. Segmentación hepática semiautomática 11

1.4.3. Segmentación hepática automática 11

2 Marco Teórico 13

2.1. Radiómica 13
2.1.1. Análisis radiómico de textura 15
2.2. Aprendizaje automático o Machine Learning 18
2.2.1. Redes neuronales 18
2.2.2. Deep Learning 20
2.2.3. Redes Neuronales Convolucionales (CNN) 21

3 Metodología **24**
3.1. Adquisición y registro de imágenes 24
3.2. Segmentación de la región de interés 25
3.3. Extracción de características radiómicas 30
3.4. Cuantificación de grasa en las zonas segmentadas 31
3.4.1. Análisis de la caída de la intensidad de señal 32
3.5. Análisis multivariante de características 32
3.6. Construcción de las reded neuronales 39

4 Resultados **42**
4.1. Redes con sistema PCA de dos componentes 42
4.2. Redes con sistema PCA de tres componentes 43
4.3. Discusión 43

5 Conclusiones **53**

Bibliografía **55**

A Repositorio: Datos y Programación **65**

B Características radiómicas **67**

Lista de Figuras

1.1. Radiografía del tórax de una persona de 38 años, tomado de [6] 4
1.2. Tomografía computarizada del abdomen de una persona de 32 años que contiene la anotación de una lesión, tomado de [8] 4
1.3. Ultrasonido de la sección abdominal. Se visualiza el hígado, el estómago y la flecha muestra una úlcera gástrica , tomado de [10] 5
1.4. Núcleo de hidrógeno en precesión alrededor del campo magnético externo (B_0). 6

2.1. Ejemplo del cálculo de la GLCM en una imagen de 4x4 píxeles (a), con 3 niveles de grises (b). El número de veces que aparece un par de píxeles en dirección horizontal es asignado en la matriz de co-ocurrencia (c). Figura adaptada de [34] 16
2.2. Estructura de una red neuronal artificial. 19
2.3. Modelo de una neurona artificial. Figura adaptada de [56] 19
2.4. Esquema de la arquitectura de una red neuronal convolucional básica (CNN). 23

3.1. Imágenes de resonancia magnética del abdomen de una mujer de 58 años en fase y fuera de fase ponderadas en T_1 obtenidas en un escáner GE Signa EXCITE de 1.5 Teslas. La imagen fuera de fase muestra la pérdida de señal debido a la presencia de depósito de grasa en el hígado. 25
3.2. 3D Slicer. Visualización de una IRM coronal de la zona abdominal. La zona verde representa el proceso de segmentación usando las herramientas del software. 26

3.3. Segmentación completa del hígado. a) Segmentación aplicada a la imagen OP. b) Segmentación aplicada a la imagen IP usando la máscara creada de la segmentación OP. 27
3.4. Segmentación por zonas en una imagen de resonancia magnética OP. . 28
3.5. Histograma de la segunda zona de interés segmentada del paciente número 1. El eje vertical representa los niveles de intensidad de la zona 29
3.6. Histograma del hígado segmentado en la imagen OP del paciente número 2. En el eje y se aprecia la normalización 29
3.7. Figura general del proceso de adquisición, segmentación, extracción de características y análisis de datos. 30
3.8. Gráfico de dispersión de las características extraídas mediante PyRadiomics. 33
3.9. Matriz de correlación de las características extraídas de imágenes de resonancia magnética del hígado. 35
3.10. Análisis de componentes principales (PCA) en 2D. En el gráfico de dispersión se puede apreciar que no existe separabilidad lineal entre los elementos con grasa y con hígado sano. 38
3.11. Análisis de componentes principales (PCA) en 3D. Se puede apreciar como están distribuidos los datos en torno a estas tres componentes. Confirmamos que no hay separabilidad lineal entre las variables. La primera y segunda componente son las mismas que el PCA en 2D y la tercera componente fue agregada ya que su peso en la varianza explicada es significante en comparación con las nueve componentes restantes. . . 39

4.1. Histograma de la cantidad de datos que se encuentran en la fase de entrenamiento (0.80) y prueba (0.2). 42
4.2. Histograma de la cantidad de datos que se encuentran en la fase de entrenamiento (0.70) y prueba (0.3). 43
4.3. Matriz de confusión de la fase de entrenamiento (a) y prueba (b) que muestra los VP, FP, VN y FN para una red neuronal con conjuntos de entrenamiento y prueba en una proporción de 80 a 20. 46
4.4. Matriz de confusión de la fase de entrenamiento (a) y prueba (b) que muestra los VP, FP, VN y FN para una red neuronal con conjuntos de entrenamiento y prueba en una proporción de 70 a 30. 48

4.5. Matriz de confusión de la fase de entrenamiento (a) y prueba (b) que muestra los VP, FP, VN y FN para una red neuronal con conjuntos de entrenamiento y prueba en una proporción de 80 a 20. 49
4.6. Matriz de confusión de la fase de entrenamiento (a) y prueba (b) que muestra los VP, FP, VN y FN para una red neuronal con conjuntos de entrenamiento y prueba en una proporción de 70 a 30. 51

Lista de Tablas

3.1. Características radiómicas originales de los cinco primeros segmentos. . 36
3.2. Estandarización de las características radiómicas de los cinco primeros segmentos. 36

4.1. Exactitud de las redes en la fase de entrenamiento y prueba. 44
4.2. Informe de clasificación de scikit-learn en la fase de entrenamiento de la red neuronal 2 con los siguientes parámetros: función de activación: relu, alpha: 0.0001, tamaño de las capas ocultas: (50, 100, 50), tasa de aprendizaje: constante. 45
4.3. Informe de clasificación de scikit-learn en la fase de prueba de la red neuronal 2 con los siguientes parámetros: función de activación: relu, alpha: 0.0001, tamaño de las capas ocultas: (50, 100, 50), tasa de aprendizaje: constante. 46
4.4. Informe de clasificación de scikit-learn en la fase de entrenamiento de la red neuronal 4 con los siguientes parámetros: función de activación: tanh, alpha: 0.0001, tamaño de las capas ocultas: (2, 50, 3), tasa de aprendizaje: constante. 47
4.5. Informe de clasificación de scikit-learn en la fase de prueba de la red neuronal 4 con los siguientes parámetros: función de activación: tanh, alpha: 0.0001, tamaño de las capas ocultas: (2, 50, 3), tasa de aprendizaje: constante. 47
4.6. Informe de clasificación de scikit-learn en la fase de entrenamiento de la red neuronal 6 con los siguientes parámetros: función de activación: relu, alpha: 0.0001, tamaño de las capas ocultas: (20,), tasa de aprendizaje: constante. 48

4.7. Informe de clasificación de scikit-learn en la fase de prueba de la red neuronal 6 con los siguientes parámetros: función de activación: relu, alpha: 0.0001, tamaño de las capas ocultas: (20,), tasa de aprendizaje: constante. 49
4.8. Informe de clasificación de scikit-learn en la fase de entrenamiento de la red neuronal 8 con los siguientes parámetros: función de activación: relu, alpha: 0.0001, tamaño de las capas ocultas: (2, 50, 3), tasa de aprendizaje: constante. 50
4.9. Informe de clasificación de scikit-learn en la fase de prueba de la red neuronal 8 con los siguientes parámetros: función de activación: relu, alpha: 0.0001, tamaño de las capas ocultas: (2, 50, 3), tasa de aprendizaje: constante. 50

B.1. Características radiómicas extraídas con la librería pyRadiomics de las 208 regiones de interés del hígado obtenido mediante imágenes de resonancia magnética. 74

Introducción

A lo largo del tiempo, ha sido necesario implementar nuevas técnicas de detección de forma no-invasiva o mínimamente invasiva que revelen información sobre la anatomía del cuerpo. Las imágenes médicas generadas a partir de diferentes principios físicos han llegado a ser indispensables en el diagnóstico y tratamiento de patologías como el Alzheimer, las enfermedades cardíacas o las afecciones causadas por tumores en los órganos internos. No obstante, su interpretación llega a ser una tarea complicada y demanda de mucho tiempo ya que, son cientos de datos los que deben ser analizados. Por esta razón, los investigadores han optado por recurrir a diferentes técnicas computacionales, como las técnicas basadas en Machine Learning, que permitan procesar, simular y analizar las imágenes en menor tiempo y con menos esfuerzo. Su uso permite realizar múltiples tareas como la identificación, clasificación o cuantificación de patrones en grandes conjuntos de datos.

Si bien el procesamiento y análisis de imágenes médicas mejora aceleradamente, existen diferentes formas de obtener los resultados deseados ya que no existe solamente una técnica de adquisición o de análisis y dependerá del estudio que deseemos realizar. En este trabajo de titulación se hace énfasis en analizar imágenes de resonancia magnética para estudiar la grasa existente en el hígado en ciertas regiones de interés. El trabajo se llevará a cabo mediante la extracción de características radiómicas dentro de dichas regiones.

El capítulo 1 describe brevemente la imagen médica y sus técnicas de adquisición. Además, se centra en el análisis de la imagen por resonancia magnética (IRM) y su rol en la estructura hepática. Finalmente, se detallan las técnicas de segmentación hepática en la resonancia magnética. En el capítulo 2 se detalla los conceptos de radiómica y aprendizaje automático, claves para el desarrollo de este trabajo. La metodología empleada tanto para la segmentación de regiones de interés y la extracción de características es explicada en el capítulo 3. También se lleva a cabo el análisis de dichas características, primordial para la creación de la red neuronal. En el capítulo 4 se presentan los resultados y por último el capítulo 5 contiene las conclusiones y recomendaciones del presente trabajo.

Capítulo 1

Antecedentes

La imagen médica produce imágenes visibles de la estructura interna del cuerpo con fines científicos y médicos mediante la creación de una estructura de datos anatómicos y fisiológicos. Su procesamiento es un campo interdisciplinario complejo que comprende diferentes áreas científicas que van, desde las matemáticas hasta la física y la medicina[1]. Por esta razón, al hablar de imágenes médicas nos referimos al conjunto de técnicas necesarias para su adquisición, reconstrucción, visualización y análisis con el fin de revelar, diagnosticar o examinar lesiones, anomalías y patologías. Por lo tanto, se trata de una herramienta fundamental en la toma de decisiones médicas. En su estudio más amplio incorpora campos como radiología, tomografía, endoscopia, termografía, fotografía médica y microscopía (por ejemplo, para investigaciones patológicas en humanos).

Los avances tecnológicos ofrecen un rápido desarrollo en los procesos de adquisición de datos y su interpretación, ofreciendo resoluciones cada vez más altas y tiempos de adquisición más rápidos. Esto conduce a la obtención de una gran cantidad de datos que deben procesarse e interpretarse rápido y con eficacia para lograr un diagnóstico acertado[1]. Debido a que esa gran cantidad de datos puede presentarse en diferentes formatos y dimensiones, su manejo puede ser un gran desafío. Así, se ha adoptado un sistema de comunicación y archivo de imágenes (PACS) que proporcione un almacenamiento, recuperación, circulación y acceso a imágenes adquiridas de diferentes modalidades y que estén simultáneamente a disposición en múltiples sitios[2]. Si bien PACS adopta cualquier formato de imagen, para que esta sea legible por los demás, es necesario implementar un formato estándar denominado DICOM (Digital Imaging and Communication On Medicine), para evitar la tarea

repetitiva de convertir imágenes de un formato a otro. DICOM engloba los procesos de transferencia, almacenamiento y reproducción de datos y está diseñado para cubrir los aspectos funcionales de la medicina y satisfacer sus aplicaciones prácticas [3].

Antes de involucrarnos con el objeto de estudio de esta tesis, es necesario realizar un breve recuento de algunas técnicas de obtención de imágenes médicas, con el fin de obtener una idea más amplia acerca del desarrollo de la imagen médica. Además, es importante introducir brevemente el análisis en la resonancia magnética que será de utilidad en el siguiente capítulo.

1.1 Técnicas de adquisición de imágenes médicas

Las imágenes médicas son generadas por equipos de imágenes digitales que usan radiación ionizante y no ionizante, capaz de penetrar los tejidos. Su intensidad depende de la parte del cuerpo examinada y de la técnica de adquisición utilizada, formando así señales que serán recibidas por los diferentes escáneres.

1.1.1 Rayos X

A finales del siglo XIX, Röentgen descubrió un tipo de radiación electromagnética al trabajar con un tubo de rayos catódicos, que denominó rayos X. En el cátodo los electrones se producen a través de un proceso de emisión térmica y se aceleran mediante una diferencia de potencial (ΔV), luego los electrones golpean el ánodo y producen rayos X [4].

Cuando el haz de rayos X atraviesa el cuerpo humano, algunos son absorbidos o dispersados produciendo una atenuación del haz que depende de la densidad del tejido. En una radiografía (figura 1.1), las estructuras densas ocasionan mayor atenuación y se muestran como gris o blanco, mientras aquellas menos densas provocan una menor atenuación y aparecen más oscuras[5].

Anteriormente, las películas de rayos X se procesaban en cuartos oscuros, en la actualidad, las imágenes radiográficas se almacenan digitalmente, al igual que las imágenes de las demás técnicas.

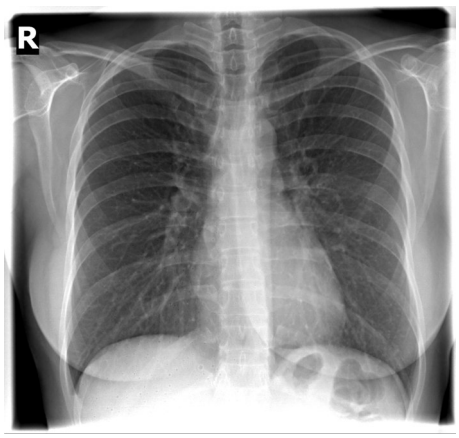


Figura 1.1: Radiografía del tórax de una persona de 38 años, tomado de [6]

1.1.2 Tomografía Computarizada (TC)

TC es una técnica basada en rayos X que produce una imagen tridimensional del interior del cuerpo adquirida mediante imágenes transversales en forma de cortes (figura 1.2). A diferencia de la técnica convencional de rayos X, TC usa una fuente de rayos X que gira alrededor del objeto a examinar y posee detectores al lado opuesto de la fuente, de manera que se capte la radiación que atraviesa al objeto [5]. La computadora es la encargada de almacenar esta información y construir una imagen en escala de grises. El uso del computador proporciona una gama más amplia de tonalidades de grises lo cual proporciona un buen contraste entre los diferentes tipos de tejidos blandos [7].

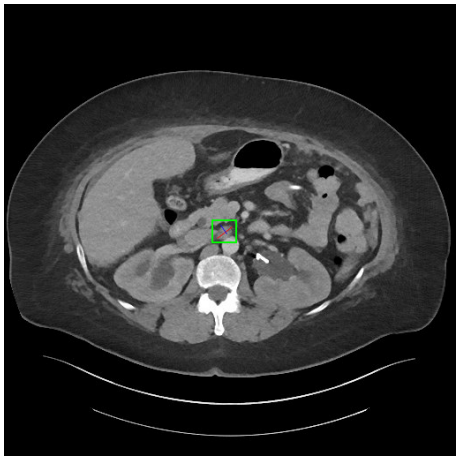


Figura 1.2: Tomografía computarizada del abdomen de una persona de 32 años que contiene la anotación de una lesión, tomado de [8]

1.1.3 Ultrasonido (US)

El ultrasonido es una técnica no invasiva que no requiere radiación ionizante. Usa ondas longitudinales mecánicas cuya frecuencia está por encima de los límites auditivos humanos con un rango superior a 20kHz. Estas ondas se originan por las vibraciones de materiales elásticos y se propagan libremente a través de los tejidos. Cuando esto sucede, algunas ondas son absorbidas y otras reflejadas. Las ondas reflejadas se conocen como *ecos* y son recibidas y convertidas en señales eléctricas para luego pasar a ser señales digitales que finalmente mostrarán una imagen 2D de las estructuras escaneadas en un monitor[4] (figura 1.3).

La frecuencia de las ondas que se usan en la imagen médica de US oscila entre los 2 Hz y los 60 MHz[9] por lo que a diferencia de las demás técnicas, se basa en un espectro de vibraciones acústicas.

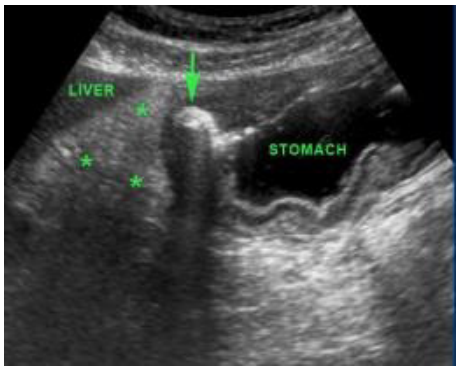


Figura 1.3: Ultrasonido de la sección abdominal. Se visualiza el hígado, el estómago y la flecha muestra una úlcera gástrica , tomado de [10]

1.1.4 Imagen por Resonancia Magnética (IRM)

Si bien la resonancia magnética (RM) fue descrita por primera vez por Bloch y Purcell en 1946[11], no fue sino hasta 1973 donde Lauterbur y Mansfiel desarrollaron la imagen por resonancia magnética, haciéndolos acreedores al Premio Nobel de medicina en 2003[12].

IRM es un método no invasivo que utiliza radiación electromagnética no ionizante. Se produce al emplear campos magnéticos de radiofrecuencia (RF) sobre núcleos atómicos sometidos a un campo magnético fuerte[13]. Los núcleos atómicos poseen una propiedad denominada espín nuclear que depende del número de protones y

neutrones. Los espines de estas partículas pueden aparearse con otra de su misma clase cuando su espín es antiparalelo y contribuye con 1/2 al valor neto del número cuántico de espín del núcleo, el cual puede ser nulo, entero o semi-entero, de acuerdo al número de partículas que lo conforman.

Como el cuerpo humano está constituido por un 60 % de agua, el IRM está especialmente adaptado para núcleos de hidrógeno [14]. Los espines nucleares de los átomos de hidrógeno en ausencia de un campo magnético externo (B_0) están orientados aleatoriamente y poseen una magnetización neta nula ($M = 0$). Si se aplica un campo magnético externo definido en una dirección del espacio, que por convención es el eje z, los espines tienden a alinearse en esta dirección y producen una magnetización neta M_z. A continuación, el eje de la magnetización experimenta precesión, es decir, el eje gira alrededor de la dirección del campo externo, como se observa en la figura 1.4. La magnetización precesa a una frecuencia fija que depende de la naturaleza del núcleo y de la fuerza del campo magnético estático y está cuantificada por la ecuación de Larmor

$$w_0 = \gamma B_0, \tag{1.1}$$

donde γ es la constante giromagnética y w_0 la frecuencia o precesión de Larmor que está en el rango de radiofrecuencias (RF) en el espectro electromagnético.

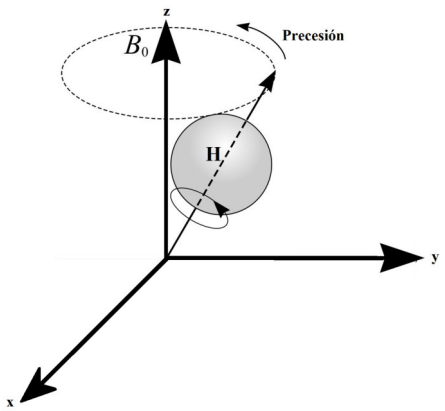


Figura 1.4: Núcleo de hidrógeno en precesión alrededor del campo magnético externo (B_0).

Al aplicar un segundo campo de radiofrecuencias perpendicular a B_0 en forma de pulsos cortos, los núcleos se excitarán ocasionando una transición desde los niveles de energía más altos a los más bajos y viceversa. Al retirar esta pulsación, inicia

el proceso de relajación, término utilizado para referirse al proceso en el cual el sistema regresará a su estado de equilibrio térmico luego de absorber energía de RF. Como resultado, se obtienen dos tipos de relajación: la longitudinal y la transversal, representadas por T_1 y T_2, respectivamente.

T_1 representa el tiempo en que la magnetización longitudinal

$$M_z = M_z(0)(1 - e^{-t/T_1}), \tag{1.2}$$

alcanza $(1 - 1/e)$ o aproximadamente el 63 % de su valor máximo y va desde varios segundos, en fluidos corporales, hasta menos de 300 [ms] en estructuras como la grasa [15]. Estas diferencias de tiempo son la razón principal del contraste de las imágenes en los tejidos, por ejemplo, la grasa que posee valores de T_1 cortos aparece más brillante en imágenes ponderadas en T_1[11]. Por otra parte, T_2 es usado para describir el tiempo necesario en el cual la componente de la magnetización transversal en el plano xy

$$M_{xy} = M_{xy}(0)(e^{-t/T_2}), \tag{1.3}$$

cae $1/e$ o aproximadamente el 37 % de su valor inicial que, a pesar de ser independiente de T_1, ocurre simultáneamente. En este caso, las sustancias que tiene T_2 largos, como el agua, aparecen más brillantes que aquellos con T_2 cortos.

En términos de IRM, las imágenes en T_1 representan planos anatómicos normales y grasos, mientras que, T_2 es utilizado para detectar anomalías como traumas, infecciones o cáncer debido a que, estos aumentan el contenido de agua de los tejidos sólidos y, por lo tanto, también aumenta el valor de T_2[14]. Finalmente, para localizar y codificar espacialmente la señal RM de una región se requiere el uso de gradientes magnéticos[1], de manera que, el espectro proporcionado por estas señales sea analizado usando una transformada de Fourier, lo que permitirá la producción de imágenes del tejido.

La imagen por resonancia magnética posee varios tipos de aplicaciones clínicas que dependen del diagnóstico deseado. Esta técnica es ampliamente utilizada para localizar tumores cerebrales, lesiones nerviosas o problemas en la sección torácica y abdominal, donde el objetivo principal es analizar órganos como el hígado, corazón,

[1] El gradiente de campo magnético es la variación del campo en función de la posición. En la RM codifica el punto exacto de donde proviene la señal para formar la imagen.

pulmones, páncreas y vesícula biliar. Además, se ha establecido como una de las herramientas más usadas en varias áreas de la medicina debido a la resolución y contraste de la imagen, los cuales pueden manipularse al variar ciertos parámetros que contribuyen al tipo y calidad de información. Igualmente, la técnica está sujeta a continuas innovaciones a través del desarrollo de su hardware y el uso de agentes de contraste. Actualmente, se explora y optimiza el uso de campos magnéticos de 3 [T] que mejorarían aún más la importancia de la resonancia magnética en el diagnóstico rutinario [11], [16].

1.2 Análisis de IRM

Al usar IRM tenemos como resultado imágenes digitales compuestas por píxeles. Estos asignan un valor de intensidad a cada canal de la imagen, denominado nivel de gris, para imágenes en escala de grises o de un solo canal. Por esta razón, procesar y analizar imágenes se convierte en una tarea compleja ya que, existen factores como la similitud del nivel de gris entre órganos adyacentes, los bordes difusos o el bajo contraste entre algunas estructuras que dificultan su valoración. Para conseguir este análisis se requiere conocer los principios físicos usados para adquirir las imágenes y el contenido de cada una de ellas ya que con esto se podrá realizar la segmentación[2] de las mismas para luego extraer sus características e interpretar los resultados.

La segmentación de imágenes médicas es una técnica que permite visualizar estructuras del cuerpo humano, así como realizar comparaciones volumétricas y de forma entre esas estructuras[17]. Desde su introducción en IRM, se han implementado varios métodos que se clasifican en tres categorías: manuales, semiautomáticos y automáticos. La segmentación manual a menudo, es realizada por el radiólogo especialista. Por lo general, lo realiza corte por corte rodeando la región de interés o anotando la zona que posee información importante[18]. A pesar que, este método es considerado estándar, es una tarea tediosa, requiere mucha atención a los detalles, consume mucho tiempo y sobre todo, es propenso a la variabilidad inter e intra-observador, ocasionando así la falta de reproducibilidad de los resultados, es decir,

[2]La segmentación consiste en dividir una imagen digital en varios conjuntos de píxeles o regiones que poseen propiedades similares como el nivel de gris, el brillo o la textura para simplificar su visualización.

que dos observadores diferentes o el mismo no obtendrán la misma segmentación [17].

Por otro lado, el continuo progreso en la visualización computarizada y los avances en el análisis y diagnóstico asistido por computadora (CAD) [19], han permitido investigar métodos que faciliten o solucionen estos problemas. La segmentación semiautomática es parte de esto, al ayudarse de algunos algoritmos durante o al final de la segmentación, reduciendo el esfuerzo y tiempo del experto[20]. Sin embargo, la variabilidad entre observadores no será eliminada por completo debido a que, la interacción con el usuario sigue presente.

La segmentación automática no requiere la interacción del usuario y generalmente es más rápida que los métodos mencionados anteriormente. Además, se ha vuelto popular gracias al uso del aprendizaje automático donde la segmentación la realizan redes neuronales entrenadas con parámetros iniciales (su concepto y entrenamiento serán descritos en el capítulo 2). Este método de segmentación es reproducible para los mismos datos y sus resultados dependen de quien automatiza la segmentación y sus errores son sistemáticos en lugar de incidentales como en los casos anteriores[18].

Generalmente, la segmentación manual es usada para la validación de los métodos semiautomáticos y automáticos a pesar de mostrar ventajas sobre ella. Actualmente, los algoritmos automáticos son el centro de atención y están siendo adoptados en el ámbito clínico para la diferenciación de tejidos y detección de tumores.

1.3 El rol de IRM en el hígado

El hígado es uno de los órganos más importantes y está a cargo de varias funciones vitales como la síntesis de proteínas o la desintoxicación de sustancias químicas para su eliminación[21]. En los últimos años, las enfermedades hepáticas crónicas (CLD, por sus siglas en inglés) como la cirrosis o el hígado graso son cada vez más comunes. Las enfermedades hepáticas se han convertido en un problema de salud mundial, sobre todo, porque la prevalencia del hígado graso ha aumentado considerablemente en los últimos 30 años, acompañada de enfermedades como la obesidad, diabetes y síndrome metabólico [22]. Por esta razón, los médicos han optado por utilizar sistemas de diagnóstico y de planificación quirúrgica[23]. Las imágenes médicas desempeñan un papel importante en la detección de las diferentes enfermedades hepáticas, lo que

permite una mejor evaluación con ayuda de los grandes avances en las técnicas de adquisición. El ultrasonido, la tomografía computarizada y la resonancia magnética son las técnicas usadas de forma habitual para valorar estas enfermedades hepáticas, ya que evalúan los cambios morfológicos del hígado [24]. Sin embargo, IRM es superior respecto a TC y US porque posee una mayor resolución espacial y de contraste gracias a los desarrollos técnicos tales como, campos magnéticos más potentes, sensores más sensibles y cambios en el diseño de la secuencia de pulsos, que permiten la adquisición de imágenes de alta resolución del hígado[25], [26].

A pesar de las ventajas mencionadas, hay que considerar que ciertos factores como la forma variable del hígado, las intensidades no homogéneas dentro del mismo, la homogeneidad de intensidad y bordes poco claros entre los órganos abdominales adyacentes (estómago, riñones y corazón) hacen que el análisis de las imágenes, sobre todo en la segmentación, sea una tarea desafiante pero atractiva para la investigación. Existe una gran cantidad de métodos que ayudan en la segmentación hepática e incluyen crecimiento de regiones, cortes de gráficos, modelos estadísticos, histogramas basados en niveles de grises o máquina de vectores de soporte (Support Vector Machines, SVM) [23]. Los métodos de segmentación hepática más comunes se detallan a continuación.

1.4 Técnicas de segmentación hepática

Como se mencionó antes, la segmentación puede ser manual, semiautomática o totalmente automática. Para el caso hepático en la resonancia magnética se han desarrollado diferentes técnicas para cada uno de estos métodos.

1.4.1 Segmentación hepática manual

Es realizada mediante el contorno de píxeles a lo largo del límite del hígado, que una vez identificado pasa por un software de post-procesamiento para generar su volumen. Entre las técnicas de contorno usadas en este tipo de segmentación están el contorno activo [27], el enfoque livewire [28] y la interpolación de formas [29]. En el contorno activo, el especialista dibuja un contorno aproximado del hígado con un cursor, el cual debe ajustarse a los bordes externos del hígado. Livewire se basa en colocar un punto de partida en el límite del hígado, de manera que, a medida que el

cursor se mueve se mostrará una ruta desde la posición actual hasta el punto inicial que se ajustará al límite del hígado[30]. En la interpolación de formas, se traza un número limitado de contornos sobre la forma del hígado, de manera que, el número de imágenes a contornear es reducido a pocos cortes importantes.

1.4.2 Segmentación hepática semiautomática

Usualmente se incluyen técnicas basadas en el nivel de gris, donde los píxeles y sus vecinos son clasificados según su intensidad. Funciona correctamente en hígados homogéneos, sin embargo, para hígados enfermos requiere una interacción mayor por parte del usuario debido a fugas que pueden presentarse en los bordes. También se usa la técnica de cortes de gráficos, donde el usuario toma algunos píxeles del primer plano, es decir, el hígado y otros de fondo, de manera que, separa estas áreas mediante cortes.

1.4.3 Segmentación hepática automática

La segmentación automática del hígado puede distribuirse en las siguientes categorías: métodos basados en la textura y en los bordes o gradientes.

1.4.3.1 Métodos basados en bordes o gradientes

El borde es una característica básica de la imagen y se describe como el límite que existe entre dos regiones que destacan las propiedades del nivel de gris, es decir, representa el final de un área y el comienzo de otra. Este tipo de métodos consiste principalmente en encontrar las discontinuidades de los objetos de interés en la imagen[31]. Posee un enfoque derivado y otro de ajuste de patrones, este último usa una serie de funciones de aproximación de bordes en una zona pequeña. En cambio, en el enfoque derivado los píxeles de borde son usados para encontrar derivadas bidimensionales [32]. Los operadores más populares usados en la segmentación de IRM son: el operador de Sobel, de Roberts, de Prewitt y de Canny.

1.4.3.2 Métodos basados en la textura

El análisis de textura proporciona información cuantitativa de la estructura interna de los objetos, por ejemplo, tejidos del cuerpo humano como el hígado. En otras palabras, la textura representa los parámetros matemáticos que describen la

distribución de intensidades del nivel de gris, mostrando la regularidad estructural del hígado en la imagen y se describe como regular, irregular, suave o rugosa [33]. Su aplicación es común en el análisis e interpretación de imágenes médicas y consta de ciertos pasos que pueden elegirse de acuerdo con el problema a tratar. Entre estos están, la definición de la región de interés, la extracción, selección y clasificación de características dentro de la región. Por lo tanto, los resultados dependerán de la metodología establecida durante el proceso [34]. Generalmente, el análisis de texturas es el método radiómico que puede extraer características definidas matemáticamente de diferentes imágenes y está enfocado en métodos estructurales, estadísticos, de transformación y basados en modelos. El término radiómica se refiere a la extracción de características de imágenes para decodificar una patología en los tejidos permitiendo crear un conjunto de datos[35].

El presente trabajo de tesis abarca el problema usando texturas basadas en métodos estadísticos. Estos representan la textura considerando las distribuciones y relaciones entre los niveles de gris en la imagen [33]. En el siguiente capítulo se describirá de una manera más concisa el concepto de radiómica junto a sus propiedades, además, se detallará el uso y manejo de redes neuronales.

Capítulo 2

Marco Teórico

La medicina es una de las áreas que ha experimentado un aumento constante de digitalización de la información. La adquisición de datos a través de imágenes médicas, ha dado lugar a una gran acumulación de datos multidimensionales, volviendo indispensable la utilización de técnicas de inteligencia artificial (IA), principalmente, big data y el aprendizaje automático o en inglés, machine learning. Su objetivo principal es facilitar las tareas repetitivas a las que los médicos especialistas se enfrentan, de manera que, sean una herramienta de ayuda en la toma de decisiones. Como se mencionó anteriormente, una técnica novedosa que permite analizar las imágenes médicas a partir de estas técnicas de inteligencia artificial se denomina radiómica.

2.1 Radiómica

La radiómica es un campo de investigación diseñado para extraer la mayor cantidad de características cuantitativas de las imágenes digitales, y luego colocarlas en una base de datos donde serán extraídas para producir una hipótesis[36]. Su concepto ha sido ampliamente aplicado en el área de oncología, sin embargo, los análisis radiómicos no se realizan únicamente en regiones tumorales, sino también, en tejidos normales[37]. Principalmente, se basa en que las imágenes médicas proporcionan información específica de las enfermedades que no es perceptible por el ojo humano pero que pueden extraerse mediante una cuantificación de textura. Esto se logra mediante la extracción matemática de la distribución espacial de las intensidades y las interrelaciones de los píxeles [38], [39].

Los trabajos desarrollados por Claude Shannon[40] y ampliados por Haralick[41]

y Galloway[42] dieron inicio al análisis de texturas, al incorporar diferentes medidas estadísticas para incluir operaciones de matrices de niveles de gris. Una de sus primeras aplicaciones fue en el análisis de fotografías aéreas, pero con el tiempo la técnica fue ampliada a otros campos, logrando ser aplicada en el análisis de las imágenes médicas [43]. Actualmente, es usada en muchas áreas de investigación, sobre todo, en aquellas donde existe un mayor uso de herramientas computacionales y capacidad de almacenamiento digital.

Llevar a cabo un estudio radiómico de imágenes médicas, incluye ciertos procesos como: *(i)* adquisición de imágenes, *(ii)* segmentación de la zona de interés, *(iii)* extracción de características y *(iv)* selección de características. Cada una de las etapas mencionadas posee sus propios desafíos y puede tratarse por separado. Sin embargo, los resultados dependerán de las variaciones en cualquiera de estos procesos. Por lo tanto, al final será necesario estandarizar el proceso alcanzando así una metodología concreta. El primer y segundo paso fueron detallados en el capítulo anterior, el tercer paso dependerá de la región de interés segmentada en el paso anterior y debe realizarse en función de las propiedades de textura y forma para producir una cantidad de características radiómicas seleccionadas por el usuario. Esta extracción de características permite describir cuantitativamente las regiones de interés seleccionadas. Finalmente, las características son clasificadas para generar un modelo que pueda por ejemplo diferenciar a un tejido normal de aquel con una lesión.

En esta tesis se abordará y discutirá la extracción de características basada en métodos de textura estadística. Cabe mencionar que existen otros métodos como aquellos enfocados en filtrado y morfología que pueden ser ampliados en [44] y [45], respectivamente. Debido a que existen varias formas y fórmulas diferentes para calcular esas características, es recomendable cumplir con las directrices de la Iniciativa de Estandarización de Biomarcadores de Imágenes [46] puesto que, en la práctica, la extracción de características se refiere a simplemente presionar el botón "ejecutar" y esperar los resultados una vez que finalice el cálculo.

2.1.1 Análisis radiómico de textura

Este tipo de características usa herramientas estadísticas para caracterizar la distribución espacial del nivel de gris de una imagen y se dividen en estadísticos de primer, segundo y orden superior.

2.1.1.1 Estadísticos de primer orden

Los estadísticos de primer orden describen la distribución de píxeles individuales de la imagen y se basan en histogramas $P(i)$ definidos como el número de píxeles con nivel gris i sobre el número total de píxeles en la región de interés. El histograma reduce a esta región de interés a valores únicos de media, varianza, entropía, skewness y kurtosis[36], [47].

Basados en la definición del histograma, la media μ_1 y los momentos centrales μ_k están dados por

$$\mu_1 = \sum_{i=0}^{L-1} i^1 P(i), \tag{2.1}$$

$$\mu_k = \sum_{i=0}^{L-1} (i - \mu_1)^k P(i), \tag{2.2}$$

donde L representa el número de niveles de grises posibles.

La varianza, la skewness y la kurtosis dados por μ_2, μ_3 y μ_4, respectivamente son los momentos centrales usualmente utilizados. La varianza describe el ancho del histograma al medir la desviación de los niveles de grises de la media, la raíz cuadrada de su valor se conoce como la desviación estándar típica. La skewness mide el grado de asimetría del histograma alrededor de la media, mientras que la kurtosis caracteriza el pico relativo o la planitud de la distribución del histograma en comparación con la distribución normal[48].

2.1.1.2 Estadísticos de segundo orden

Describen las interrelaciones estadísticas entre píxeles vecinos con valores de contraste similares o diferentes. La información puede extraerse de la matriz de co-ocurrencia de niveles de grises (con sus siglas en inglés, GLCM) definida por Halarick [49], la cual cuantifica la presencia de píxeles con la misma intensidad en una distancia a lo largo de una dirección fija.

Usualmente la matriz de co-ocurrencia $P(i,j)$ está construida en cuatro orientaciones espaciales (0°, 45°, 90° y 135°) y su número de filas y columnas es igual al número de niveles de grises L de la imagen. Cada elemento (i,j) representará la frecuencia con la que el píxel con nivel de gris i esta relacionado espacialmente con el píxel de nivel de gris j. La figura 2.1 muestra un ejemplo de la construcción de una GLCM cuya relación entre píxeles es calculada horizontalmente hacia la derecha, es decir, su orientación espacial es 0°.

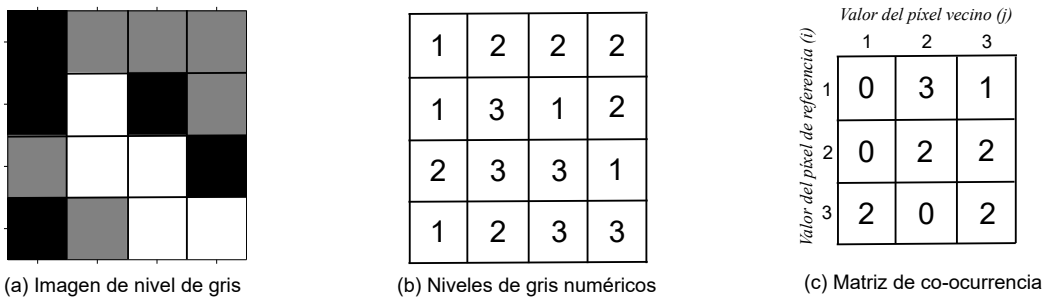


Figura 2.1: Ejemplo del cálculo de la GLCM en una imagen de 4x4 píxeles (a), con 3 niveles de grises (b). El número de veces que aparece un par de píxeles en dirección horizontal es asignado en la matriz de co-ocurrencia (c). Figura adaptada de [34]

Después de calcular esta matriz se pueden extraer características estadísticas de segundo orden como la energía o segundo momento angular, la entropía, el momento de diferencia inversa (MDI), el contraste y la correlación. Estas características están descritas de la siguiente forma,

$$\text{Energía} = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(P(i,j))^2, \tag{2.3}$$

$$\text{Entropía} = -\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}P(i,j)\log_2(P(i,j)), \tag{2.4}$$

$$\text{MDI} = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\frac{1}{1+(i-j)^2}P(i,j), \tag{2.5}$$

$$\text{Contraste} = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(i-j)^2P(i,j), \tag{2.6}$$

$$\text{Correlación} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{(ij)P(i,j) - \mu_{1i}\mu_{1j}}{\sigma_i \sigma_j}, \tag{2.7}$$

donde μ y σ son los valores de la media y la desviación estándar de P_i y P_j.

La energía por su parte mide la uniformidad de textura de la imagen al cuantificar el número de píxeles repetidos. La entropía es una medida de la aleatoriedad de la imagen, en cambio MDI mide la homogeneidad de la imagen local y asume valores grandes para diferencias de tonos de grises más pequeñas. Por otro lado, el contraste es una medida de la variación del tono de gris de un conjunto contiguo de píxeles, mientras que, la correlación muestra la dependencia lineal del tono de gris de los píxeles [48], [50], [51].

2.1.1.3 Estadísticos de orden superior

Comparan las diferencias y relaciones entre varios píxeles y son características extraídas luego de aplicar filtros o trasformaciones matemáticas a las imágenes, con el fin de identificar patrones repetitivos y no repetitivos, reducir ruidos y mejorar los bordes . En este grupo se incluye el análisis fractal, los funcionales de Minkowski y las transformadas laplacianas, gaussianas y wavelets [52].

La definición y uso de las características, filtros y transformaciones matemáticas que permiten el procesamiento de imágenes se remontan hace años atrás pero la radiómica innovó su aplicación, al basarse en el uso simultáneo de una gran cantidad de parámetros extraídos de un tejido, los cuales son procesados bajo la hipótesis de que, al combinar adecuadamente estas características se encontrarán propiedades del tejido claves para el diagnóstico.

No obstante, existen dificultades al momento de elegir cuántos y cuales parámetros son necesarios extraer de las imágenes, debido a que todos estos nos permitirán generar varias variables a partir de una sola imagen. Por esta razón, se realiza una selección de características que mejoren el rendimiento de predicción y proporcionen una mejor comprensión del proceso evitando características con información repetida. Debido a que esta tarea produce demasiadas variables y es un gran desafío para los métodos tradicionales, se ha implementado la idea del uso del aprendizaje automático cuya función principal es detectar y diferenciar patrones al utilizar datos de entrenamiento (características) para aprender a hacer predicciones [53].

2.2 Aprendizaje automático o Machine Learning

Este término introducido por Arthur Samuel en 1959 [54], es usado para describir una rama de la inteligencia artificial que permite producir modelos capaces de clasificar, predecir y estimar una situación a partir de un conjunto de datos de entrada. Como se ha descrito en las secciones anteriores, los avances han permitido que las técnicas de aprendizaje automático sean claves y de gran importancia en el análisis de las imágenes médicas, ya sea en su procesamiento, registro, segmentación, extracción o selección de características, o a su vez, que formen parte en todo este proceso. Se tiene dos categorías básicas de aprendizaje automático: el aprendizaje supervisado y no supervisado. Generalmente, el estudio de imágenes médica emplea el aprendizaje supervisado siempre y cuando se cuente con un conjunto de datos etiquetados de entrada y de salida.

En imágenes médicas, las redes neuronales artificiales son la base principal del aprendizaje automático y también de un subcampo nuevo de este, conocido como aprendizaje profundo o Deep Learning (DL). Los algoritmos desarrollados a partir de redes neuronales están estructurados en capas compuestas por nodos interconectados. Por ejemplo, en el caso de imágenes médicas, las entradas pueden ser características radiómicas o si usamos redes neuronales convolucionales (la base de deep learning) podrían ser las propias imágenes [55].

2.2.1 Redes neuronales

Las redes neuronales artificiales están inspiradas en la estructura neuronal del cerebro. Su objetivo es aprender de la información de entrada para optimizar la salida mediante la coordinación del procesamiento interno [56]. Su estructura más simple consta de una capa de entrada que alimenta a la red, una oculta donde se procesan los datos y una de salida donde se presentarán los resultados del problema, es decir, la información fluirá secuencialmente de la capa de entrada a la oculta y de la oculta a la capa de salida. Dicha estructura puede observarse en la figura 2.2. Cada capa contiene cierta cantidad de nodos o neuronas artificiales interconectados con otras neuronas que dependen de los pesos[1] de la red neuronal. Las neuronas son

[1]Los pesos son parámetros numéricos que determinan el impacto con la que cada una de las neuronas afecta a la otra.

los elementos de procesamiento de cada capa y los pesos representan las conexiones entre las neuronas[57].

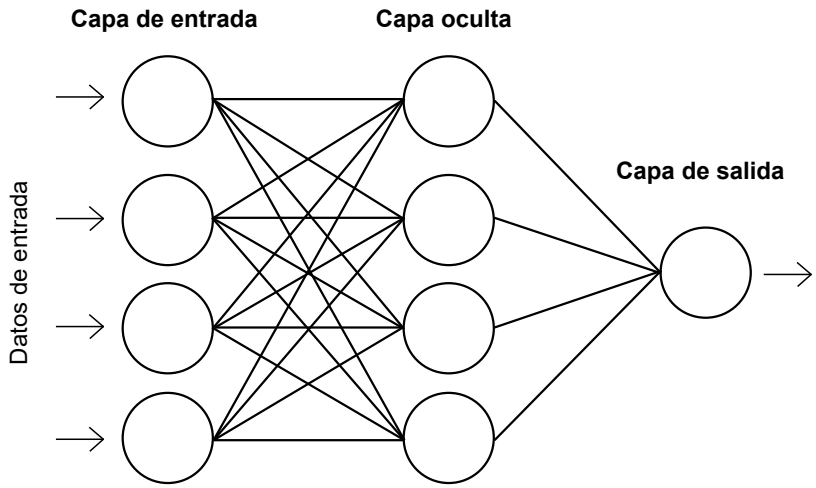


Figura 2.2: Estructura de una red neuronal artificial.

La estructura básica de una neurona artificial está representada en la figura 2.3, donde x_1, x_2, ... ,x_n son las entradas a la neurona y Y la salida. Cada entrada está multiplicada por su peso w_i y habitualmente se asocia un parámetro b que se denomina umbral (en inglés bias) que permite ajustar de mejor manera la predicción de los datos. La suma de todo esto pasa por una función de activación f que muestra la respuesta de una neurona a la señal de entrada total que se recibe.

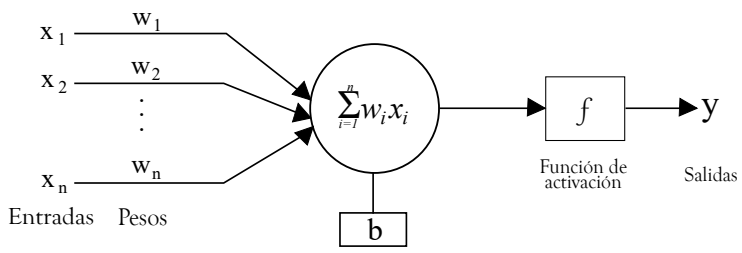


Figura 2.3: Modelo de una neurona artificial. Figura adaptada de [56]

Como resultado, la relación entre la entrada y la salida está descrita de la siguiente manera [58]:

$$Y = f\left(\sum_{i=1}^{n} w_i x_i + b\right). \tag{2.8}$$

El número de nodos y capas ocultas depende del problema que se intenta resolver. Para imágenes médicas hay diferentes arquitecturas de redes neuronales disponibles, pero la red neuronal de alimentación directa o feed-forward es una de las más comunes ya que las entradas y salidas son numéricas y los pares de vectores proporcionan una base clara para el entrenamiento supervisado [56]. En este tipo de redes, las conexiones son unidireccionales, es decir, la información pasa a través de la red en una única dirección, desde la capa de entrada, a la oculta hasta la de salida. Un tipo de red de alimentación directa comúnmente usada es el *perceptrón multicapa*, compuesto de tres o más capas y que utiliza el algoritmo de aprendizaje de retropropagación, el cual minimiza el error entre la salida generada por la red y la salida objetivo al propagar el error de regreso a la red a través de cada neurona, actualizando así los pesos de conexión.

Por otra parte, las funciones de activación pueden ser lineales y no lineales. En caso de usar una función lineal nos referimos a un modelo simple de regresión lineal a diferencia de las no lineales que facilitan la adaptación del modelo con una variedad de datos y por eso son las más utilizadas. Si consideramos el algoritmo de retropropagación, las funciones deben ser continuas y diferenciables por lo tanto, trabajaremos con funciones no lineales. Entre ellas podemos mencionar a la función de activación lineal rectificada (ReLU) (ecuación 2.9), la función logística (sigm) (ecuación 2.10) y la tangente hiperbólica (tanh) (ecuación 2.11).

$$\text{ReLU(x)} = \max(0, x), \tag{2.9}$$

$$\text{sigm(x)} = \frac{1}{1 + e^{-x}}, \tag{2.10}$$

$$\tanh\text{(x)} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}. \tag{2.11}$$

2.2.2 Deep Learning

Como un subcampo de Machine Learning, en estos últimos años, Deep Learning (DL) ha demostrado resultados y desarrollos importantes en tareas como, reconocimiento de voz, facial, seguimiento visual y diagnóstico médico [59]. Su diferencia

con las técnicas convencionales de aprendizaje automático, sobre todo en imágenes médicas es importante puesto que, el primer paso para identificar un objeto usualmente es extraer sus características, lo cual es usualmente realizado por un experto. Por lo tanto, él es el encargado de decidir qué características serán importantes para implementar en el algoritmo, pero en ocasiones se corre el riesgo de incluir características redundantes y poco útiles. DL fusiona este proceso de extracción y la toma de decisiones intentando aprender las características por sí solo a partir de los datos, para así llevar a cabo una tarea determinada, provocando una reducción de experiencia en el área donde se determinarán las características [60].

Existen varios tipos de algoritmos de DL como por ejemplo, redes neuronales convolucionales (por sus siglas en inglés, CNN), redes neuronales profundas (DNN), redes neuronales recurrentes (RNN), etc. Sin embargo, las CNN son el tipo de arquitectura más popular en el análisis de imágenes médicas cuya estructura se detallará a continuación.

2.2.3 Redes Neuronales Convolucionales (CNN)

Las redes neuronales convolucionales están inspiradas en estructuras multicapas: una capa de entrada, capas ocultas y una de salida. Como se mencionó anteriormente, las CNN son populares en el análisis de imágenes médicas, esto debido a que conservan las relaciones espaciales de los datos que son importantes por ejemplo, para diferenciar donde un tejido normal interactúa con otro cancerígeno [61].

Dentro de la capa oculta, CNN contienen múltiples capas convolucionales y de activación, intercaladas por capas de agrupación (pooling). Además, poseen capas totalmente conectadas que calculan los resultados finales. Conceptualmente, la extracción de características de las imágenes que realizan las capas de convolución y agrupación, es similar al proceso de extracción en el análisis radiómico.

La convolución está definida como una operación de dos funciones para producir una nueva función. En imágenes médicas, una función consta de los valores de entrada en cierta posición, es decir, los píxeles y la segunda función es un filtro o núcleo (kernel). Todos estos valores están representados como matrices de números sobre los cuales se aplicará la operación de convolución. Cada filtro comparte los mismos pesos en todas las posiciones de entrada debido a que, se desplaza de posición en

posición hasta cubrir completamente la imagen y además porque una característica que aparece en una parte de las imágenes, probablemente aparecerá en otras. Por tanto, al aplicar el filtro en toda la imagen se puede reducir el número de parámetros necesarios para crear un modelo. Por ejemplo, si un filtro es capaz de detectar líneas horizontales, este podría usarse para detectar todas estas líneas en cualquier parte. Finalmente, las operaciones de convolución y la aplicación de este grupo de filtros producirá un mapa de características, seguido de funciones de activación que permitirán que las características seleccionadas pasen a la siguiente capa [62].

Luego de la capa de convolución, se aplican funciones de activación no lineales como, la unidad lineal rectificada (ReLU). Esta función reemplaza los valores negativos de entrada por ceros, permitiendo así, simplificar y acelerar los cálculos y el entrenamiento[63].

La capa de agrupación se ubica en medio de la capa de convolución y de la función de activación. Su función principal es combinar características similares para reducir el número de parámetros y cálculos en la red de manera que los tiempos de entrenamiento sean reducidos. En otras palabras, permite obtener un mapa de características de menor resolución. Las tareas de agrupación generalmente se calculan usando la función de agrupación máxima, la cual toma el valor de entrada más grande dentro del filtro y descarta el resto de valores.

La última capa de una red neuronal convolucional es la capa totalmente conectada, la cual integra todas las características obtenidas por medio de la capa de convolución y agrupación al conectar cada neurona de la capa anterior a cada neurona de la capa totalmente conectada. La estructura simple de una CNN se ilustra en la figura 2.4. Finalmente, la salida realizará tareas de regresión o clasificación a través de la función softmax que se usa para normalizar la salida al convertir el vector de números en una distribución de probabilidad cuyo total suma 1. Esta función generalmente se usa para calcular las pérdidas esperadas al entrenar un conjunto de datos. La fórmula softmax es

$$\sigma(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}, \tag{2.12}$$

donde x_i representa a los elementos del vector de entrada y el término del denominador nos asegura la normalización.

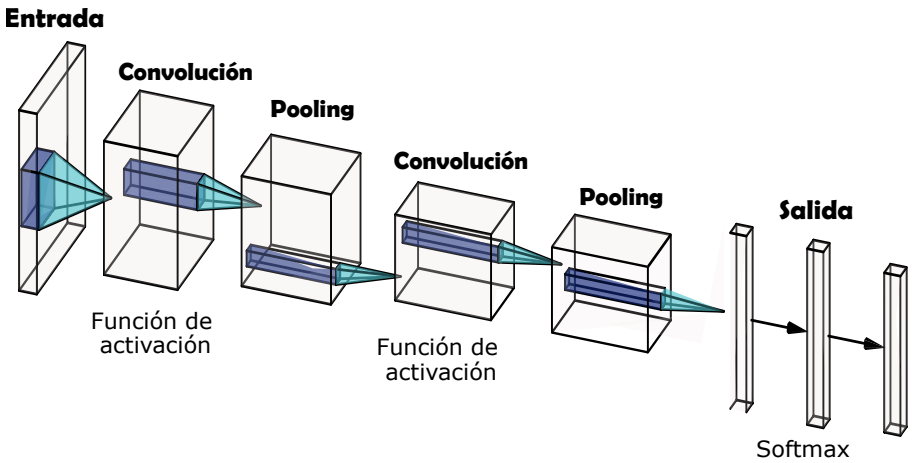


Figura 2.4: Esquema de la arquitectura de una red neuronal convolucional básica (CNN).

El desarrollo y validación de un algoritmo de DL normalmente consiste de un conjunto de datos de entrenamiento, validación y prueba [53]. Durante el entrenamiento de la red, los datos de entrenamiento y validación se envían a una función de pérdida que mide el error. Luego, este error es propagado hacia atrás para que las CNN ajusten sus parámetros con el fin de reducir los errores. La fase de entrenamiento continuará hasta que la función de pérdida haya alcanzado un mínimo y posteriormente se seleccionará un modelo. Una vez que todos los parámetros de dicho modelo sean fijados, su rendimiento será evaluado mediante el conjunto de datos de prueba.

Básicamente el diseño de una red neuronal depende de varios factores como, la comprensión de la tarea que se intenta resolver, los requisitos que debe cumplir, la memoria que se consumirá o simplemente los datos que se utilizarán para alimentar la red y puede conseguirse mediante la utilización de un software que construya y entrene redes neuronales. Al principio, para generar una red se utilizaba componentes como AlexNet[64] y Lenet[65]. Sin embargo, hoy en día existen una cantidad de marcos de software como TensorFlow, Scikit-Learn, Pytorch, Theano y Keras que pueden describir las arquitecturas de redes neuronales de una manera eficiente. Además, algunos como TensorFlow permiten el uso de los recursos informáticos disponibles, como las unidades de procesamiento gráficos (GPU). De igual manera, es importante destacar que este tipo de software es gratis, de código abierto y se encuentra en constante desarrollo [62].

Capítulo 3

Metodología

3.1 Adquisición y registro de imágenes

La muestra está constituida por 26 pacientes que poseen hígados normales y con grasa. Las imágenes médicas fueron obtenidas en un escáner de resonancia magnética GE Signa EXCITE de 1.5 Teslas en formato DICOM. Son imágenes en dos dimensiones ponderadas en T_1 en fase (IP) y fuera de fase (OP). Cada archivo DICOM comprende de 60 cortes coronales cubriendo así, la totalidad de la zona abdominal. El grosor de corte es de 6 mm con un espacio entre rebanadas de 7mm. El tiempo medio de adquisición es menor a 20 minutos lo que lo hace tolerable para los pacientes. El tamaño de la matriz de 14 imágenes es 512×512 píxeles, mientras que las 12 restantes tienen un tamaño de 256×256 píxeles.

Las imágenes fueron adquiridas en las secuencias en fase en tiempos de eco $TE = 4{,}4$ ms y fuera de fase con $TE = 2{,}2$ ms, es decir, los protones de agua y grasa estarán en fase a 4,4 ms y fuera de fase 2,2 ms antes o después del primer tiempo de eco en fase. La razón principal del uso de estas secuencias, se debe a que, permiten identificar de forma no invasiva las señales de los protones de agua y grasa en función de su frecuencia de resonancia [66]. Por lo tanto, el contenido de grasa en los tejidos del abdomen, puede ser cuantificado con las imágenes OP e IP al calcular la fracción de señal de la grasa (FSF) (sección 3.4). La diferencia de la intensidad en las imágenes OP e IP (figura 3.1), se debe a que, en las OP los protones de la grasa y del agua se encuentra a 180° uno con respecto al otro, ocasionando que sus intensidades se resten, mientras que, en las imágenes IP las intensidades se suman.

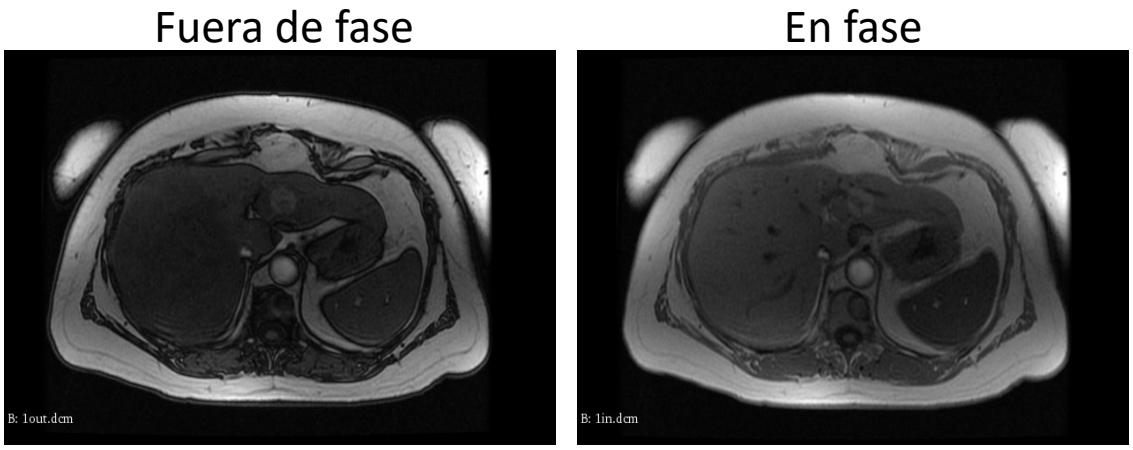


Figura 3.1: Imágenes de resonancia magnética del abdomen de una mujer de 58 años en fase y fuera de fase ponderadas en T_1 obtenidas en un escáner GE Signa EXCITE de 1.5 Teslas. La imagen fuera de fase muestra la pérdida de señal debido a la presencia de depósito de grasa en el hígado.

De los 60 cortes obtenidos, se ha elegido aquel que proporcione una imagen completa del hígado, necesario para la segmentación total del mismo. Cada uno de los archivos DICOM han sido almacenados en el formato NRRD (nearly raw raster data). La principal ventaja del uso de este formato, es que los archivos son anónimos, por lo tanto, no contienen información confidencial de los pacientes. Además, en el formato NRRD el escaneo médico es almacenado en un solo archivo, a diferencia que cada archivo DICOM se compone de varios directorios que pueden contener cientos de archivos individuales. En consecuencia, el formato NRRD nos permitirá llevar un mejor manejo en la transferencia de los datos médicos y a su vez protegerá la privacidad de los pacientes.

3.2 Segmentación de la región de interés

La técnica de segmentación nos permite separar una imagen en múltiples regiones de interés de acuerdo al estudio que se requiera realizar. Para el estudio de la grasa presente en el hígado hemos utilizado segmentación manual y semiautomática. La segmentación manual fue útil para reconocer la ubicación y el límite preciso del órgano, es decir, para segmentar el hígado completo. En cambio, la segmentación semiautomática fue usada para marcar zonas en el hígado con diferente intensidad de grises. Este proceso se llevó a cabo, gracias a que, las imágenes OP nos permiten

observar de mejor manera las zonas más blanquecinas y oscuras que se presentan en la imagen.

La segmentación fue procesada en un software de código abierto y gratuito, llamado 3D Slicer [67] disponible en Windows o Ubuntu. 3D Slicer es comúnmente usado para analizar e investigar las imágenes médicas mediante su visualización. Su principal ventaja es que, posee un interfaz de programación en el lenguaje de código Python, de manera que, reduce el tiempo de análisis de las imágenes, pues en lugar de trabajar una a una, podemos tomar a todas como un solo conjunto.

Mediante el software 3D Slicer, las imágenes en formato DICOM fueron transformadas a NRRD, luego el hígado de las 26 imágenes OP fue segmentado completamente con la herramienta Segmentation al dibujar el contorno del mismo, como se puede observar en la figura 3.2.

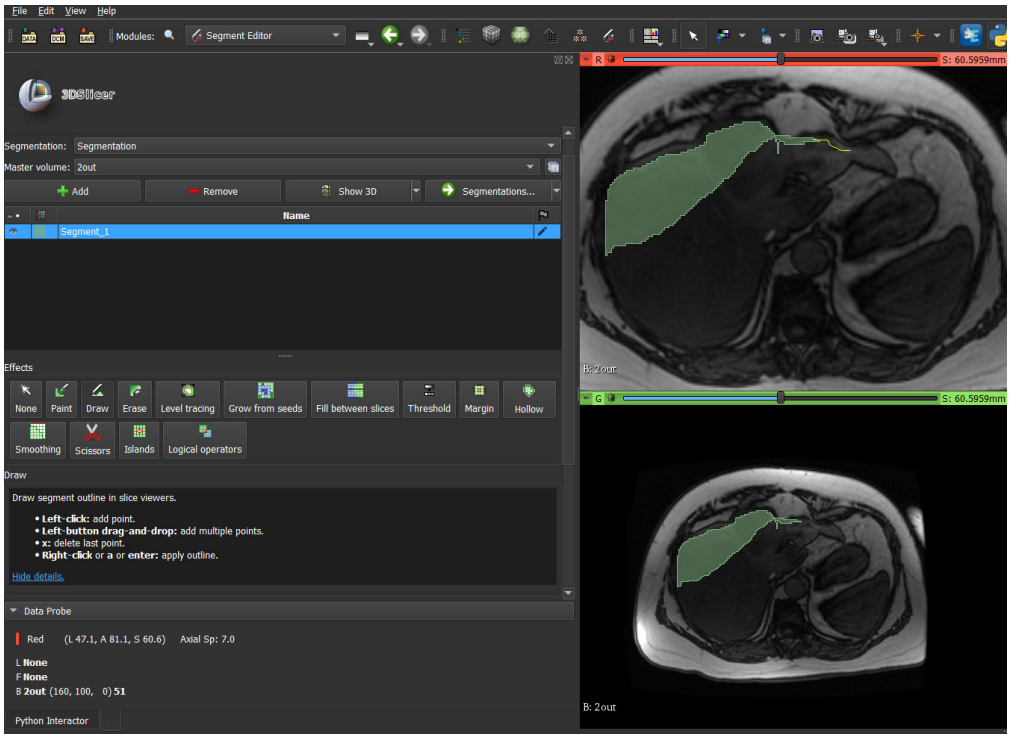


Figura 3.2: 3D Slicer. Visualización de una IRM coronal de la zona abdominal. La zona verde representa el proceso de segmentación usando las herramientas del software.

Con la ayuda de la ventana de interacción se implementó un programa capaz de convertir cada una de las segmentaciones obtenidas en una máscara (Apéndice A), con el fin de aplicar la misma segmentación en las imágenes IP que será de

utilidad para calcular la FSF más adelante. Cabe mencionar que cada segmentación es independiente debido a que, el hígado es un órgano sin morfología definida, lo cual hace que los resultados varíen en cada paciente, por lo tanto, se han realizado 26 segmentaciones completas diferentes. La imagen 3.3 es un ejemplo de la segmentación aplicada sobre el hígado de las imágenes OP e IP.

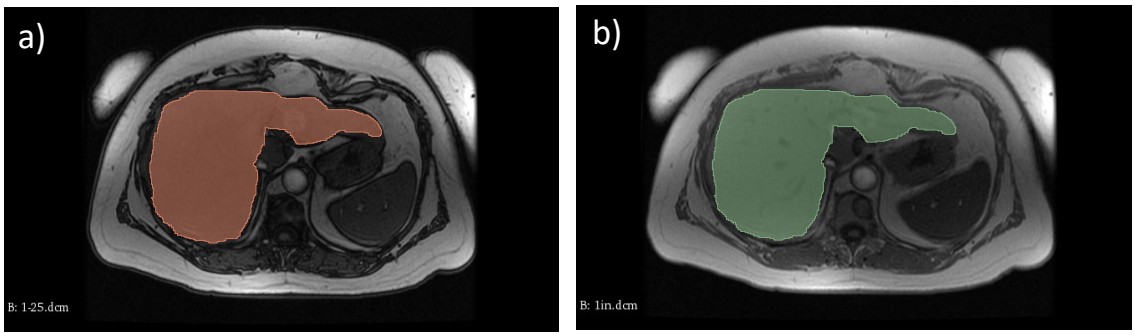


Figura 3.3: Segmentación completa del hígado. a) Segmentación aplicada a la imagen OP. b) Segmentación aplicada a la imagen IP usando la máscara creada de la segmentación OP.

Una vez identificado el hígado, aplicamos segmentaciones en forma circular alrededor del hígado de acuerdo a los diferentes niveles de grises que se presentaron en el mismo. Este tipo de segmentación es usualmente utilizada para analizar las caídas de intensidad de señal con ayuda de imágenes OP e IP (más información en [68]). Para realizar la segmentación, utilizamos el código *segmentos e histogramas.txt* disponible en el Apéndice A, el cual fue variando de acuerdo a la posición espacial de secuencia en fase y fuera de fase. Por cada imagen OP se tomaron 8 regiones de interés diferentes de manera que, la mayor parte del área total del hígado quede cubierta (ver figura 3.4), consiguiendo así un total de 208 datos segmentados. Usando el mismo proceso anterior correspondiente a la segmentación total, se creó una máscara de la segmentación para aplicarla a las imágenes IP.

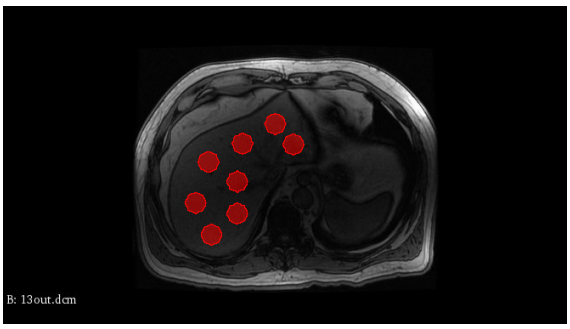


Figura 3.4: Segmentación por zonas en una imagen de resonancia magnética OP.

El software 3D Slicer es de gran utilidad si exploramos todas sus herramientas. Una de ellas calcula ciertos parámetros como la media, la mediana, la desviación estándar y también el número de píxeles existentes en la zona segmentada, los cuales son indispensables en el cálculo de histogramas. Los parámetros estadísticos fueron extraídos en tablas de datos con el fin de compararlos con aquellos valores que se obtendrán al realizar la extracción de características.

Finalmente, con el código *segmentos e histogramas.txt* en la interfaz de Python de 3D Slicer, se obtuvieron los diferentes histogramas a partir de las regiones de interés segmentadas para las imágenes OP (figura 3.5). El número de bins fue determinado mediante la regla de Sturge [69] que asume distribuciones aproximadamente normales y está dada por,

$$k = 1 + \log_2 n, \tag{3.1}$$

donde k es el número de bins y n el número de píxeles de la zona segmentada.

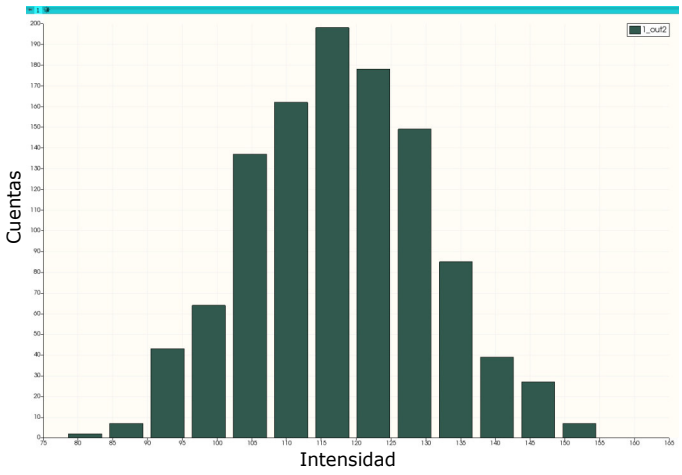


Figura 3.5: Histograma de la segunda zona de interés segmentada del paciente número 1. El eje vertical representa los niveles de intensidad de la zona

Debido a que los tamaños de las imágenes son diferentes (respecto a su número de píxeles), es recomendable normalizar los histogramas, puesto que, nuestro interés es conocer los porcentajes de intensidad de cada muestra, mas no los conteos absolutos. Para la normalización, utilizamos los datos extraídos del software 3D Slicer. Esos datos fueron graficados en Gnuplot usando el código del *graf.pro* detallado en el Apéndice A. En la figura 3.6 podemos observar un ejemplo del histograma normalizado con su respectiva función de distribución, la cual se aproxima a una distribución unimodal. El repositorio del Apéndice A contiene los datos necesarios para generar los histogramas de las zonas segmentadas de las imágenes OP.

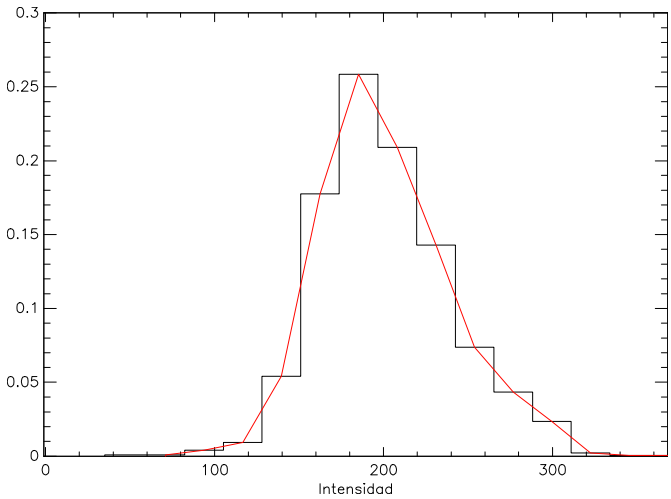


Figura 3.6: Histograma del hígado segmentado en la imagen OP del paciente número 2. En el eje y se aprecia la normalización

3.3 Extracción de características radiómicas

Luego de segmentar las imágenes, se llevó a cabo la extracción de características radiómicas que serán utilizadas para cuantificar los niveles de gris dentro de la región de interés segmentada. Las características fueron extraídas utilizando el paquete PyRadiomics [70] en Python. Este paquete nos permite extraer los datos radiómicos de imágenes médicas usando tres pasos principales: *(i)* carga de la imagen base y de la máscara obtenida a través de la segmentación, *(ii)* cálculo de características de acuerdo al criterio del usuario y *(iii)* retorno de resultados.

La figura 3.7 muestra el proceso general que se debe llevar a cabo para analizar las imágenes médicas. En primer lugar, las imágenes fueron extraídas y están segmentadas. Luego, las diferentes características se extraen usando PyRadiomics tomando en cuenta el tipo de características y en caso de existir, los filtros. Finalmente, estás características deben ser analizadas.

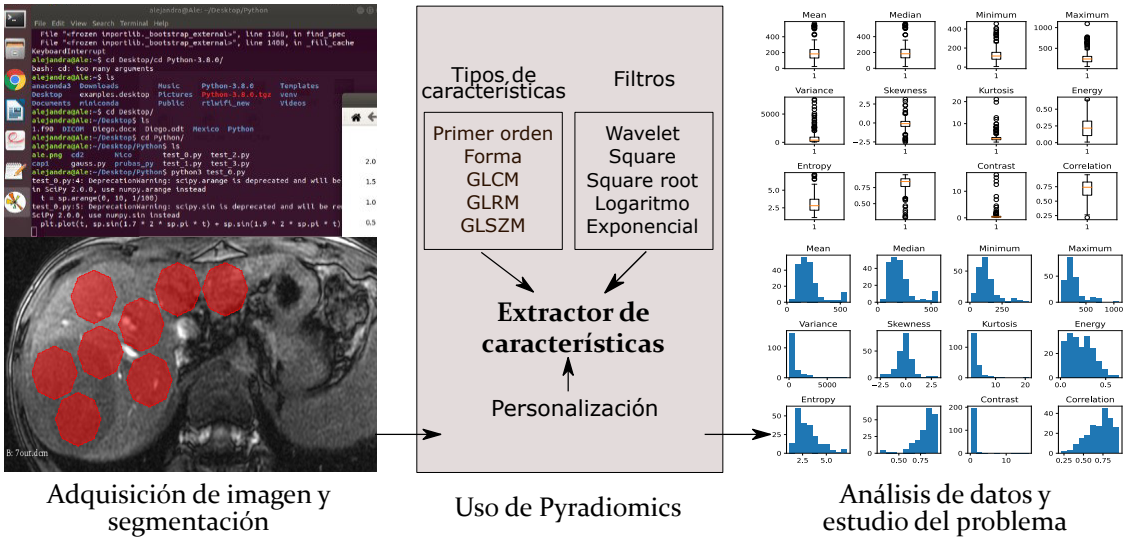


Figura 3.7: Figura general del proceso de adquisición, segmentación, extracción de características y análisis de datos.

Naganawa et. al. [71] realizó un estudio previo aplicado en tomografías computarizadas, donde utiliza características extraídas mediante histogramas para predecir la esteatohepatitis no alcohólica, una enfermedad hepática que se caracteriza por la acumulación de grasa en el órgano. Naganawa recomendó que los estudios futuros se realicen utilizando resonancia magnética como fuente de información, como es el caso del presente trabajo. Sin embargo, hemos ampliado el análisis al trabajar con

estadísticos de segundo orden, los cuales hasta ahora han sido usados solamente en la detección de tumores. Debido a que en la actualidad, la investigación radiómica no posee patrones estandarizados y unificados, la información de las características radiómicas está basada en la iniciativa de estandarización de biomarcadores de imagen [46].

La extracción de características mediante PyRadiomics puede ser personalizada. La referencia [70] muestra cómo realizar este proceso de acuerdo a cada necesidad. En este trabajo de titulación se modifica el código principal. (La información esta disponible en sección *Instalación y modificación de PyRadiomics* del Anexo A), de manera que, se obtiene los valores de la media, mediana, mínimo, máximo, varianza, skewness, kurtosis, energía, entropía, MDI, contraste y correlación, es decir, algunos estadísticos de primer y segundo orden cuyas definiciones están detalladas en el capítulo 2 sección 2. Por lo tanto, tenemos 12 características para los 208 datos disponibles en el Apéndice B.

3.4 Cuantificación de grasa en las zonas segmentadas

Ciertos estudios muestran que existe una buena correlación entre el uso de la técnica de resonancia magnética y la histología de los pacientes con acumulación de grasa en el hígado, enfermedad también conocida como esteatosis hepática. Esta especificidad se encuentra entre los siguientes rangos $(76,7 - 90\,\%)$ y $(87,1 - 91\,\%)$. Hablamos de esteatosis o hígado graso cuando este posee un $5\,\%$ de grasa y se clasifica en leve $(5 - 33\,\%)$, moderado $(34 - 66\,\%)$ y grave $(\geq 67\,\%)$ [66].

Una de las ventajas de la técnica RM es que permite detectar la esteatosis aún cuando el hígado tiene solo un $3\,\%$ de contenido graso [72], la cual puede ser calculada mediante diferentes técnicas. La imagen codificada por desplazamiento químico o método de Dixon es una de ellas y se basa en la diferencia de las frecuencias de resonancia entre las señales de los protones de agua y grasa que componen el hígado. Debido a que es necesario etiquetar nuestros datos entre hígado graso y normal, hemos decidido trabajar con esta técnica ya que varios estudios [73] han encontrado una alta correlación entre el porcentaje de grasa histológica y el porcentaje calculado con imágenes de RM con desplazamiento en fase y fuera de fase. Por lo tanto, la

clave de este cálculo es poseer imágenes tanto en fase como fuera de fase, las cuales están disponibles en este trabajo.

3.4.1 Análisis de la caída de la intensidad de señal

Se ha comprobado que, la presencia de grasa en el hígado provoca una pérdida de intensidad de señal en las imágenes IP (figura 3.1), en cambio, en el hígado normal las imágenes OP e IP poseen intensidades similares. Las señales de grasa y agua se cancelan en los tiempos de eco fuera de fase y se suman en los tiempo en fase. Debido a este fenómeno es posible calcular la grasa hepática al comparar las imágenes secuenciales OP e IP [74].

La fracción de señal de grasa (FSF) se puede calcular aplicando la siguiente fórmula:

$$FSF = \frac{S_{IP} - S_{OP}}{2S_{IP}}, \tag{3.2}$$

donde S_{IP} y S_{OP} corresponden a las intensidades medidas en la región de interés en las imágenes IP y OP, respectivamente. Para calcular S_{IP} y S_{OP} se tomaron los valores medios de cada una de las secciones segmentadas registrados en el procedimiento de la sección anterior y finalmente se aplicó la ecuación 3.2. Una vez obtenido el valor de la FSF, éste fue multiplicado por 100 para determinar el porcentaje de grasa presente en cada una de las regiones. Los datos fueron etiquetados en dos grupos: hígado normal, si el porcentaje de grasa es $<5\,\%$ e hígado graso si su porcentaje es $\geq 5\,\%$

Hay que tener presente que cada región de interés debe ser la misma para las imágenes IP y OP, es decir, la segmentación debe ubicarse en el mismo punto y tener el mismo tamaño. El programa 3D Slicer, nos facilitó este procedimiento al crear la máscara de segmentación a partir de las imágenes OP.

3.5 Análisis multivariante de características

El paso preliminar del análisis es la exploración visual de datos a partir de gráficos o tablas que los resuman, puesto que, de esta manera es más sencillo captar los puntos principales que en comparación a varias líneas de datos sin procesar. Los diagramas de dispersión son comúnmente usados en dicha visualización y sirven

principalmente para identificar una relación por pares, tres o más características al realizar diferentes combinaciones. Además, permiten evaluar si los datos tienen una correlación lineal con otras variables. La figura 3.8 muestra la matriz de dispersión de las 12 características extraídas de las zonas segmentadas del hígado. En la diagonal se visualiza el histograma de cada una ellas. Se observa que existe una correlación lineal predominante entre algunas características como por ejemplo, media-mediana, varianza-entropía, energía-entropía, entropía-MDI y contraste-MDI. Sin embargo, ciertos pares de características no poseen una correlación lineal fuerte.

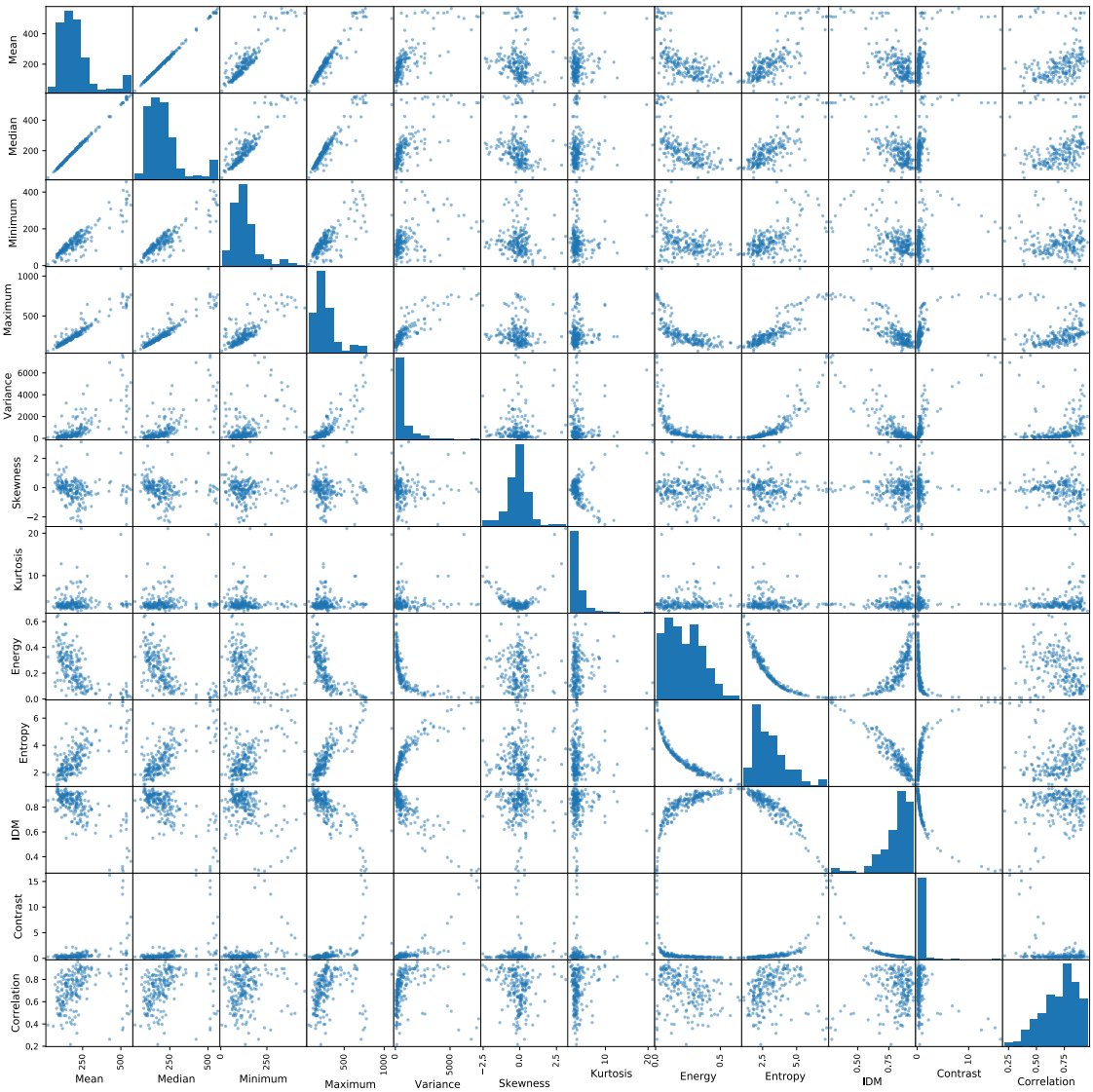


Figura 3.8: Gráfico de dispersión de las características extraídas mediante PyRadiomics.

El coeficiente de correlación de Pearson [75] puede cuantificar dicha correlación al ser una medida del grado de relación entre dos variables relacionadas linealmente. Este coeficiente solo asume valores entre -1 y 1. Hablamos de una correlación muy fuerte cuando el coeficiente tiene valores entre 0.9 a 1, de una correlación fuerte para valores entre 0.7 a 0.9, la correlación es moderada si los valores van desde 0.5 a 0.7, la correlación es débil cuando los valores están entre 0.3 a 0.5 y finalmente los valores de 0 a 0.3 indican una correlación insignificante. Esto para valores tanto positivos como negativos.

La figura 3.9 es una matriz de correlación que contiene los coeficientes de correlación entre las diferentes características. Esta matriz se utiliza para resumir los datos y es la cuantificación de la correlación obtenida en la figura 3.8. Además, la figura está representada mediante una gráfico de calor que nos ayuda a visualizar de mejor manera los coeficientes. Los colores más claros y más oscuros representan una correlación muy fuerte. Podemos observar que la media, la varianza, la energía, la entropía y el MDI son las características que poseen una mayor correlación respecto al resto de características. Por otro lado, las características como la skewness, kurtosis, contraste y correlación poseen una correlación débil. Sobre todo el parámetro estadístico correlación no se relaciona linealmente con ninguna otra característica.

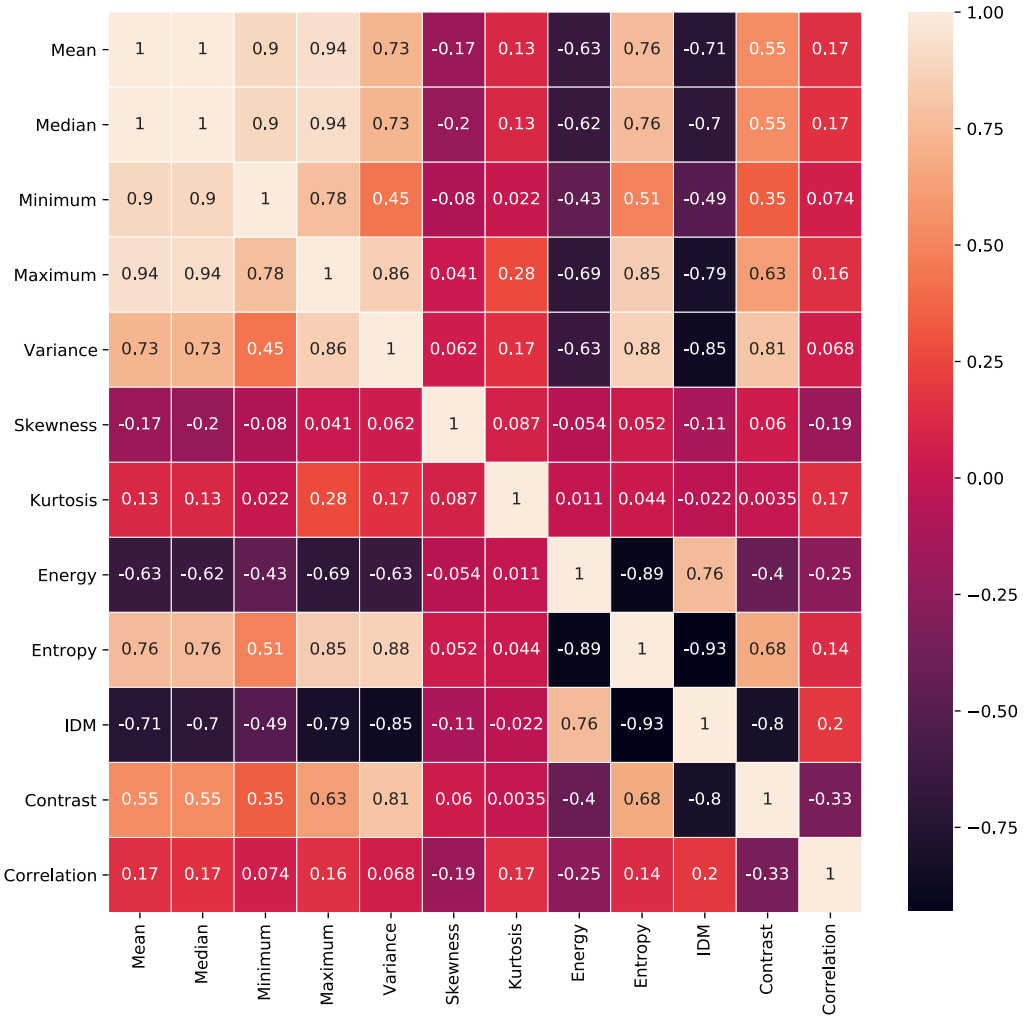


Figura 3.9: Matriz de correlación de las características extraídas de imágenes de resonancia magnética del hígado.

Este análisis es importante para comprender las relaciones existentes entre las diferentes características de manera que, seamos capaces de crear un modelo de aprendizaje automático.

Los conjuntos de datos del mundo real tienden a tener muchas características que pueden considerarse como una dimensión en el espacio. Como consecuencia, trabajamos con datos de alta dimensión cuya visualización completa es difícil. En este caso nosotros contamos con características, que estrictamente hablando serían 12 dimensiones. Sin embargo, para lograr ver el conjunto de datos como un todo sin perder mucha información, recurrimos a la reducción de la dimensionalidad. Uno de los métodos de reducción más conocidos es el Análisis de Componentes Principales

(PCA) que intenta preservar las partes que tienen más variación de los datos y elimina aquellas partes no esenciales con menos variación, al simplificar el conjunto de datos seleccionando algunas componentes principales de las características originales [76]. En general, su propósito es encontrar la mejor representación de baja dimensión en un conjunto de datos multivariados sin perder mucha información.

Por otra parte, hay que considerar que la escala de algunas características es mayor y si estas no son estandarizadas predominarán sobre el resto. La estandarización consiste en cambiar la escala de las características para que tengan propiedades de una distribución normal, es decir, con media cero y desviación estándar uno. Para estandarizar las características usamos el módulo StandardScaler de la biblioteca scikit-learn de Python [77]. En el cuadro 3.1 se observan seis características originales extraídas de los 5 primeros segmentos del hígado, en cambio el cuadro 3.2 muestra las mismas seis características de dichos segmentos pero estandarizadas. Los datos completos y los detalles de la estandarización pueden encontrarse en el repositorio detallado en el Apéndice A.

	Mean	**Median**	**Minimum**	**Maximum**	**Variance**	**Skewness**
1	114.040946	115.0	37.0	160.0	219.358651	-0.814406
2	120.707650	120.0	81.0	158.0	170.157701	0.074498
3	105.027372	105.0	71.0	143.0	141.435382	0.146669
4	110.680657	110.0	78.0	142.0	130.381597	0.118146
5	113.856882	113.0	78.0	160.0	211.513702	0.395898

Cuadro 3.1: Características radiómicas originales de los cinco primeros segmentos.

	Mean	**Median**	**Minimum**	**Maximum**	**Variance**	**Skewness**
1	-0.826514	-0.817566	-1.245327	-0.739410	-0.497403	-0.862788
2	-0.767259	-0.773564	-0.661776	-0.751978	-0.534167	0.284839
3	-0.906629	-0.905571	-0.794401	-0.846232	-0.555630	0.378015
4	-0.856381	-0.861568	-0.701564	-0.852516	-0.563889	0.341191
5	-0.828150	-0.835167	-0.701564	-0.739410	-0.503265	0.699785

Cuadro 3.2: Estandarización de las características radiómicas de los cinco primeros segmentos.

Luego de la estandarización de datos, definimos el número de componentes principales. Es necesario considerar que, no hay un método único que permita identificar el número óptimo de componentes. No obstante, como el objetivo de PCA es reducir la dimensionalidad, basta con utilizar el número mínimo de componentes. En este trabajo de titulación se utiliza la *varianza explicada* como una medida para obtener el número de componentes principales ya que cuantifica la cantidad de información que puede ser atribuida a cada una de las componentes principales. Al determinar la varianza explicada obtuvimos que, las dos primeras componentes principales representan al 70 % de la varianza explicada. Además, hemos ampliado nuestro estudio al uso de tres componentes principales, obteniendo en estas tres una varianza del 80 %.

Por lo tanto, las doce características han sido reducidas a 2 y 3 dimensiones, de manera que, podamos proyectar los resultados a través de gráficos en 2D (figura 3.10) y 3D (figura 3.11) que visualmente son más inteligibles. En la figura 3.10 los ejes x e y del gráfico de dispersión son las dos primeras componentes principales obtenidas por PCA y su histograma muestra que la varianza máxima (alrededor del 57 %) se encuentra en la primera componente principal. La segunda componente principal posee la segunda varianza más alta (alrededor de 13 %) del grupo de datos. Por lo tanto, la varianza explicada se concentra en estas dos primeras componentes.

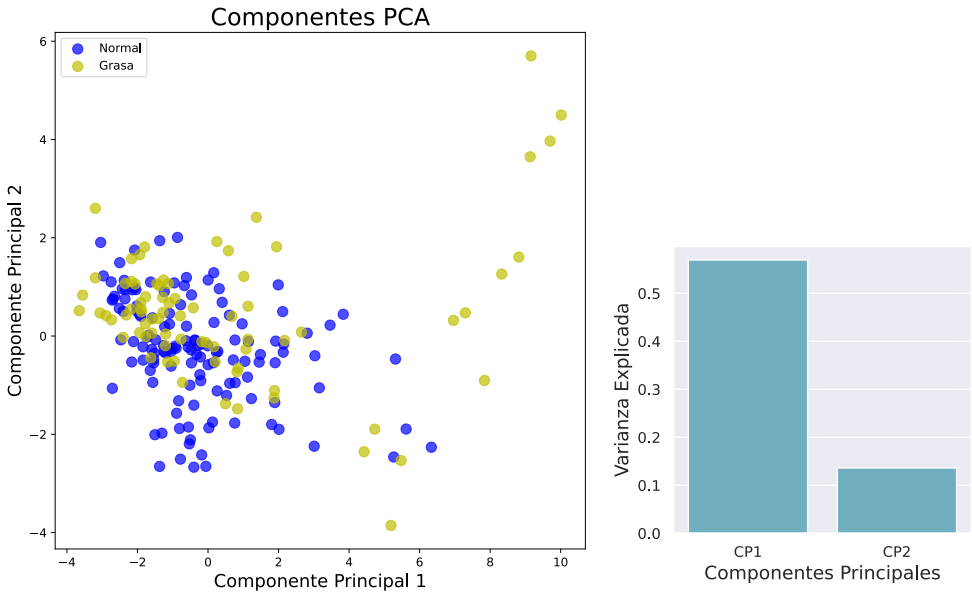


Figura 3.10: Análisis de componentes principales (PCA) en 2D. En el gráfico de dispersión se puede apreciar que no existe separabilidad lineal entre los elementos con grasa y con hígado sano.

Los ejes x, y y z de la figura 3.11 representan las tres primeras componentes principales. Al igual que en la figura anterior, en el histograma se aprecia la varianza de cada una de las componentes principales. En este caso, se ha añadido una tercera con un valor alrededor del 10 %. Las siguientes componentes (cuarta, quinta, sexta, etc) poseen una menor cantidad de información en comparación con las componentes principales elegidas y en total aportan cerca del 20 % de los datos. Se ha decidido no utilizar el resto de componentes ya que, este 20 % de la varianza restante está distribuida en las siguientes nueve componentes, es así que no aportan cambios significativos en nuestros resultados.

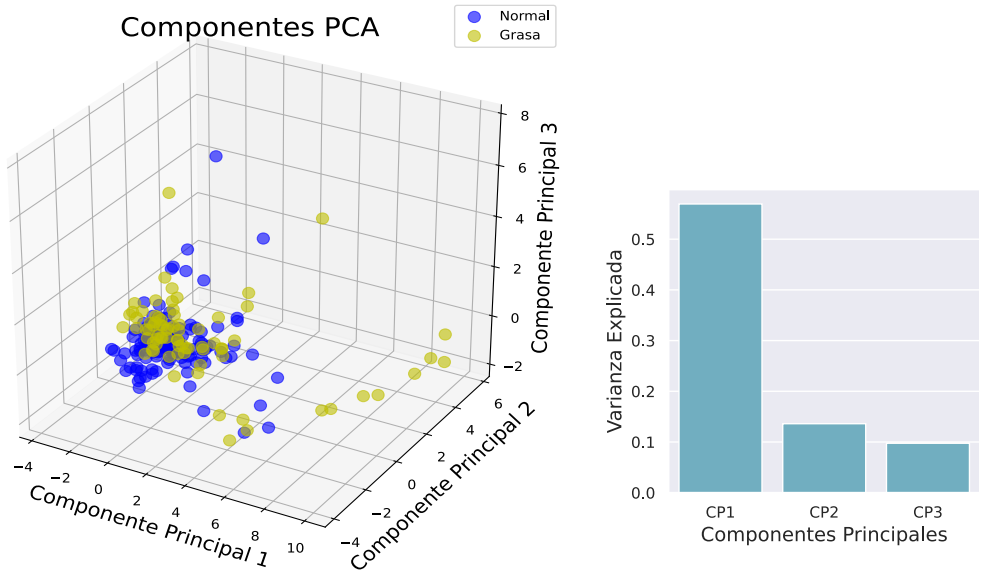


Figura 3.11: Análisis de componentes principales (PCA) en 3D. Se puede apreciar como están distribuidos los datos en torno a estas tres componentes. Confirmamos que no hay separabilidad lineal entre las variables. La primera y segunda componente son las mismas que el PCA en 2D y la tercera componente fue agregada ya que su peso en la varianza explicada es significante en comparación con las nueve componentes restantes.

Una vez reducidas las características a 2 y 3 componentes construimos nuestra red neuronal. Para esto, abordamos al problema como un modelo de aprendizaje supervisado de clasificación, ya que requerimos conocer si cierta zona de interés corresponde a una con grasa o normal y esto se realiza a través de la evaluación y la validación.

3.6 Construcción de las reded neuronales

El primer paso para la construcción de la red es transformar el tipo (grasa o normal) a notación binaria. El 0 corresponde a las zonas donde no existe un porcentaje de grasa mayor al 5 % y el 1 a las zonas con un porcentaje mayor a ese 5 %. La red fue construida usando el algoritmo de *perceptrón multicapa* ya que tenemos un conjunto de características x, determinadas mediante PCA que no son linealmente separables (figura 3.10 y 3.11) y el objetivo y que son las etiquetas de los datos (0 y 1). Adicionalmente, utilizamos el método *train_test_split* de la librería scikit-learn para dividir los datos en un conjunto de entrenamiento y uno

de prueba. Luego, con la clase MLPClassifier de la misma librería implementamos el algoritmo de perceptrón multicapa y obtuvimos una red neuronal simple. No obstante, debido a que uno de nuestros objetivos es mejorar la red neuronal utilizamos el ajuste de *hiperparámetros* diseñado para encontrar un conjunto de parámetros que produzcan un mejor rendimiento. Scikit-learn posee la clase GridSearchCV que evalúa y selecciona de manera sistemática los parámetros de cierto modelo. Para esto, es necesario proveer al modelo de un conjunto de parámetros que serán evaluados de todas las combinaciones posibles mediante validación cruzada[1].

El rendimiento de los problemas de clasificación suele ser medido usando métricas como la exactitud, precisión, sensibilidad y la puntuación F1. La exactitud es una medida general de como se comporta el modelo en todas las clases al medir el porcentaje de casos clasificados correctamente y se calcula como la relación entre el número de predicciones verdaderas y el número total de predicciones,

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN}. \tag{3.3}$$

Para este caso en particular, los *verdaderos positivos (VP)* se refieren a las regiones de interés que posee un porcentaje de grasa mayor al 5 % y se detectan como regiones con grasa. Los *falsos positivos (FP)* son las regiones sin grasa (porcentaje de grasa menor al 5 %) que se detectan como una con grasa. Por otro lado, los *verdaderos negativos (VN)* son las regiones que detectan como normales (sin grasa). Y finalmente, los *falsos negativos (FN)* hacen referencia a las regiones con grasa que se clasifican como normales. Hay que tener cuidado con los falsos negativos pues, diagnosticar que un paciente no posee ningún tipo de anomalía llega a ser sumamente peligroso para él en el futuro.

La precisión se refiere a la habilidad del modelo para predecir como positivo lo que realmente es positivo. Se calcula como la relación entre el número de predicciones clasificadas correctamente y el número total de casos clasificados como positivos (correcta o incorrectamente),

$$\text{Precisión} = \frac{VP}{VP + FP}. \tag{3.4}$$

[1]La validación cruzada se usa para estimar el rendimiento de los modelos predictivos y ayuda a evitar el sobreajuste, particularmente cuando la cantidad de datos es limitada. Su proceso consiste en realizar un número fijo de particiones de los datos, luego ejecuta el análisis en cada partición y finalmente promedia el error general.

En cambio, la sensibilidad mide la habilidad del modelo para encontrar todos los casos positivos y es la relación entre el número de casos positivos correctamente identificados y el número total de muestras positivas,

$$\text{Sensibilidad} = \frac{VP}{VP + FN}. \tag{3.5}$$

La puntuación F1 combina las medidas de precisión y sensibilidad, donde 1 es la mejor puntuación y 0 la peor y se calcula por,

$$\text{Puntuación F1} = 2 * \frac{Precisión * Sensibilidad}{Precisión + Sensibilidad}. \tag{3.6}$$

Se desarrollaron 8 redes neuronales cuya construcción está disponible en el anexo A en los archivos $Red_PCA2_(0_2)$, $Red_PCA2_(0_3)$, $Red_PCA3_(0_2)$ y $Red_PCA3_(0_3)$ con diferentes parámetros, los cuales están detallados en el siguiente capítulo.

Capítulo 4

Resultados

Todos los resultados se obtuvieron usando un estado aleatorio (random state) de 42 para el entrenamiento y prueba y con el parámetro de estratificación (stratify) activo. Además, para todas las redes neuronales el número máximo de iteraciones fue fijado en 200. El estado aleatorio nos garantiza que las divisiones de entrenamiento y prueba sean reproducibles, caso contrario cada vez que el código sea ejecutado se generaría un nuevo valor aleatorio y por consecuencia un nuevo conjunto de entrenamiento y prueba. El parámetro de estratificación permite que el entrenamiento y la prueba tengan las mismas proporciones que el conjunto de datos de entrada.

4.1 Redes con sistema PCA de dos componentes

El tamaño del conjunto de entrenamiento para *la red número 1* es 0.80 y el 0.20 restante se asignó al conjunto de prueba (figura 4.1). Esta red fue construida con los parámetros asignados por defecto de la librería scikit-learn.

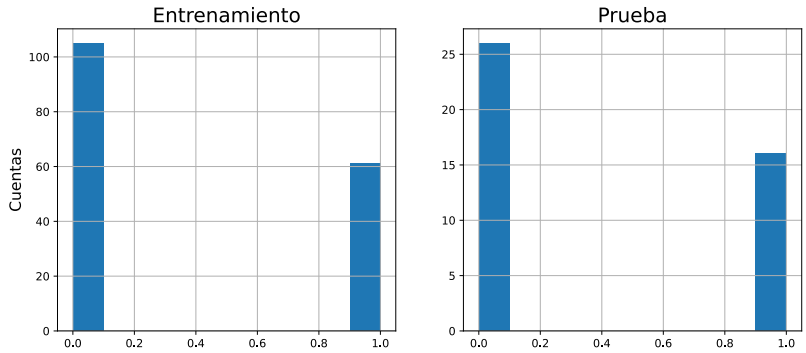


Figura 4.1: Histograma de la cantidad de datos que se encuentran en la fase de entrenamiento (0.80) y prueba (0.2).

La red neuronal 2 tiene los mismos valores de entrenamiento y prueba que la red 1 y fue construida con los parámetros obtenidos luego de realizar el ajuste de hiperparámetros. Todos los parámetros usados para las diferentes combinaciones están detallados en el código. El tamaño del entrenamiento de *la red neuronal 3* es 0.70 y la prueba 0.3 (figura 4.2) y sus parámetros son los mismos que los de la red 1. Se utilizó el mismo porcentaje de entrenamiento y prueba de la red anterior para *la red neuronal 4* y sus parámetros fueron extraídos del ajuste de hiperparámetros.

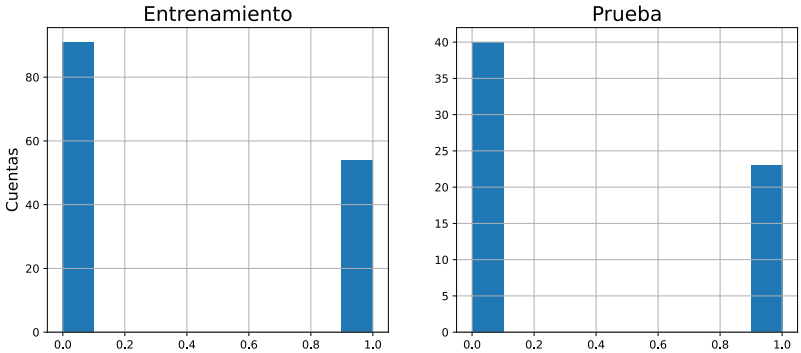


Figura 4.2: Histograma de la cantidad de datos que se encuentran en la fase de entrenamiento (0.70) y prueba (0.3).

4.2 Redes con sistema PCA de tres componentes

Las redes neuronales 5, 6, 7 y 8 fueron construidas bajo las mismas condiciones que la sección anterior solamente cambiamos el conjunto de características x. Por lo tanto, los porcentajes de entrenamiento de la red neuronal 1, 2, 3 y 4 son los mismo para la red neuronal 5, 6, 7 y 8, respectivamente. Debemos tomar en cuenta que los parámetros usados en las redes construidas bajo el ajuste de hiperparámetros son diferentes para cada caso ya que, si cambia una condición de las mencionadas anteriormente, los resultados variarán.

4.3 Discusión

Antes de introducirnos a detallar los resultados obtenidos de los diferentes modelos de redes neuronales es importante mencionar que realizamos una red neuronal preliminar con los 12 parámetros estadísticos principales donde obtuvimos

valores de exactitud y precisión del 100 %. Estos resultados indican que el modelo esta sobre ajustado y difieren completamente de los resultados reales. Por esta razón, realizamos el análisis multivariante de características detallado en la sección 3.5. Además, implementamos los algoritmo en sklearnel y no en TensorFlow (TF) porque el primero es una biblioteca de aprendizaje automático en general, en cambio TF es una biblioteca de aprendizaje profundo y para esto los datos de entrada no serían las características radiómicas extraídas sino más bien el archivo de cada una de las imágenes de resonancia magnética.

La tabla 4.1 muestra los valores de la exactitud de nuestras diferentes redes, es decir, el porcentaje de casos clasificados correctamente.

Exactitud		
Red neuronal	**Entrenamiento**	**Prueba**
1	0.6927710843373494	0.6428571428571429
2	0.8072289156626506	0.6666666666666666
3	0.6551724137931034	0.6666666666666666
4	0.7034482758620689	0.6825396825396826
5	0.7228915662650602	0.5952380952380952
6	0.6867469879518072	0.6190476190476191
7	0.7379310344827587	0.6349206349206349
8	0.7448275862068966	0.6507936507936508

Cuadro 4.1: Exactitud de las redes en la fase de entrenamiento y prueba.

Como podemos observar en la tabla 4.1, los datos son clasificados de mejor manera en la fase de entrenamiento que en la fase de prueba. Asimismo la exactitud del modelo mejora al utilizar el ajuste de hiperparámetros. Por esta razón, hemos decidido descartar las redes con menor exactitud, es decir, aquellas que fueron construidas con los parámetros por defecto de la librería (1, 3, 5 y 7) para así, analizar y comparar otros parámetros estadísticos como la precisión y la sensibilidad del resto de redes (2, 4, 6 y 8).

En las siguientes tablas, como la 4.2 se presentan los valores de la precisión, la sensibilidad, la puntuación F1 y el soporte cuyo valor es el número de elementos de cada clase que aparecen en el conjunto de datos especificados, y nos informan

si existe la necesidad de realizar un reequilibrio de datos, donde se puede aplicar técnicas como el sobre-muestreo o sub-muestreo.

El promedio macro y ponderado esta determinado para cada clase y cada uno tiene su propósito. Por ejemplo, si requerimos que a todos los números del conjunto de datos se les asigne un peso idéntico usaremos el promedio macro, en cambio si los datos están desequilibrados y queremos dar más importancia a alguna predicción manejamos el promedio ponderado.

ENTRENAMIENTO	**Precisión**	**Sensiblidad**	**Puntuación F1**	**Soporte**
0	0.91	0.81	0.86	119
1	0.62	0.81	0.70	47
promedio macro	0.77	0.81	0.78	166
promedio ponderado	0.83	0.81	0.81	166

Cuadro 4.2: Informe de clasificación de scikit-learn en la fase de entrenamiento de la red neuronal 2 con los siguientes parámetros: función de activación: relu, alpha: 0.0001, tamaño de las capas ocultas: (50, 100, 50), tasa de aprendizaje: constante.

La tabla 4.2 muestra que la clase 0 del entrenamiento tiene una mejor precisión que la clase 1 y en esta ocasión la sensibilidad es la misma. El promedio macro de la precisión nos indica que el 77 % de las veces predecimos un diagnóstico correcto. La tabla 4.3 representa los valores obtenidos al ingresar los datos en el modelo entrenado. En este caso el promedio macro de la precisión es de un 63 %, en otras palabras esto implicaría que al ingresar los datos de prueba, el 37 % de las veces predecimos que el hígado de una persona esta sano cuando en realidad posee un hígado graso.

PRUEBA	**Precisión**	**Sensiblidad**	**Puntuación F1**	**Soporte**
0	0.77	0.71	0.74	28
1	0.50	0.57	0.53	14
promedio macro	0.63	0.64	0.64	42
promedio ponderado	0.68	0.67	0.67	42

Cuadro 4.3 continuación de la página anterior

Cuadro 4.3: Informe de clasificación de scikit-learn en la fase de prueba de la red neuronal 2 con los siguientes parámetros: función de activación: relu, alpha: 0.0001, tamaño de las capas ocultas: (50, 100, 50), tasa de aprendizaje: constante.

La figura 4.3 indica el valor de los diferentes eventos en la fase de entrenamiento y prueba con sus respectivos porcentajes. En la fase de entrenamiento los VP y VN son predominantes, mientras que en la fase de prueba lo son los VP. En el entrenamiento los FN son mayores que los FP y entre los dos constituyen cerca del 20 % de total. Por otra parte, en la prueba los FN y VP tienen el mismo porcentaje, lo que implicaría que tenemos la misma probabilidad de diagnosticar correctamente un hígado graso o de equivocarnos al afirmar que un paciente tiene un hígado normal cuando en realidad es graso.

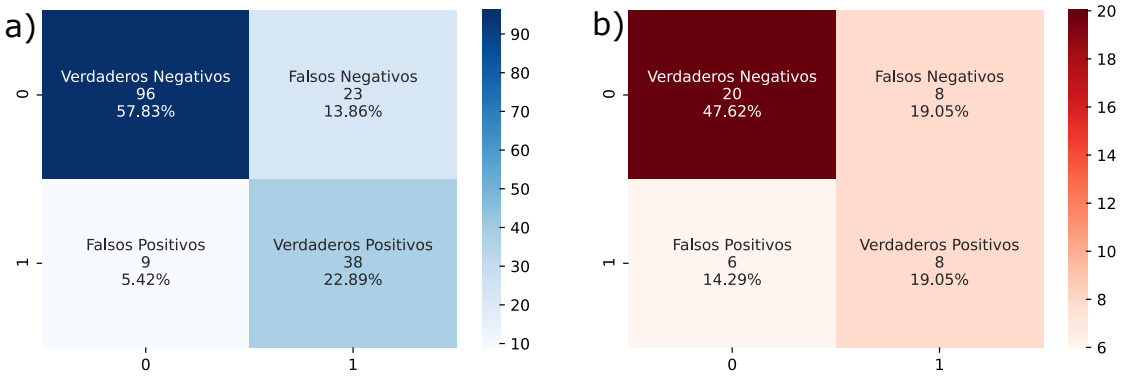


Figura 4.3: Matriz de confusión de la fase de entrenamiento (a) y prueba (b) que muestra los VP, FP, VN y FN para una red neuronal con conjuntos de entrenamiento y prueba en una proporción de 80 a 20.

El porcentaje de FN y FP será analizados en los diferentes modelos de redes neuronales, pues nuestro propósito es equivocarnos lo menos posible en el diagnóstico. Además, hay que considerar que los FN pueden llegan a ser perjudiciales y hasta mortales para un paciente.

La tabla 4.4 muestra que en el entrenamiento el 64 % de las veces predecimos hígados grasos correctamente y que el porcentaje de identificar VP Y FP es del 70 %. En el caso de la prueba (tabla 4.5) la precisión es del 61 % y su sensibilidad del 66 %. A diferencia del modelo de la red neuronal 2, los resultados del entrenamiento y la

prueba de la red 4 no varían en gran medida. No obstante, la precisión de la red 2 es mayor que de esta red neuronal.

ENTRENAMIENTO	Precisión	Sensiblidad	Puntuación F1	Soporte
0	0.90	0.71	0.79	116
1	0.37	0.69	0.48	29
promedio macro	0.64	0.70	0.64	145
promedio ponderado	0.79	0.70	0.73	145

Cuadro 4.4: Informe de clasificación de scikit-learn en la fase de entrenamiento de la red neuronal 4 con los siguientes parámetros: función de activación: tanh, alpha: 0.0001, tamaño de las capas ocultas: (2, 50, 3), tasa de aprendizaje: constante.

PRUEBA	Precisión	Sensiblidad	Puntuación F1	Soporte
0	0.88	0.70	0.78	50
1	0.35	0.62	0.44	13
promedio macro	0.61	0.66	0.61	63
promedio ponderado	0.77	0.68	0.71	63

Cuadro 4.5: Informe de clasificación de scikit-learn en la fase de prueba de la red neuronal 4 con los siguientes parámetros: función de activación: tanh, alpha: 0.0001, tamaño de las capas ocultas: (2, 50, 3), tasa de aprendizaje: constante.

Los FN del entrenamiento y prueba de la red neuronal 4 (figura 4.4) es la segunda clase más prevaleciente. Si bien los VN ocupan más de la mitad del total, el porcentaje de los FN es alto pues es casi 1/4 del total, lo que significa que es más probable afirmar que un paciente no tiene hígado graso cuando en realidad si lo tiene, que afirmar que un paciente con hígado graso lo tiene. Los FP están reducidos, sin embargo, lo ideal sería que los FN no sean muchos.

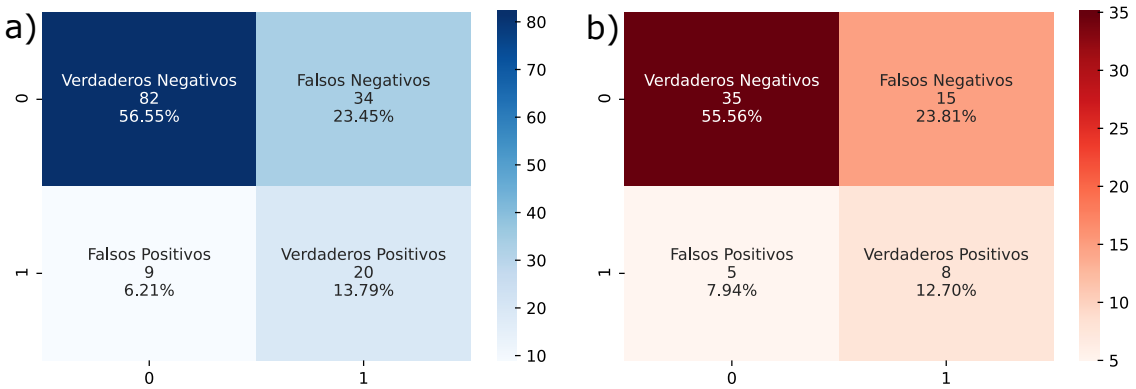


Figura 4.4: Matriz de confusión de la fase de entrenamiento (a) y prueba (b) que muestra los VP, FP, VN y FN para una red neuronal con conjuntos de entrenamiento y prueba en una proporción de 70 a 30.

La red neuronal 6 construida con tres componentes principales cuya fase de entrenamiento y prueba tienen en una proporción de 80 y 20 alcanza una precisión del 60 % en la primera y del 54 % en la segunda. En comparación con la red 2 y 4, la precisión de la 6 en la prueba es menor, por lo tanto, decimos que solo el 54 % de las veces predecimos hígados grasos correctamente.

ENTRENAMIENTO	**Precisión**	**Sensiblidad**	**Puntuación F1**	**Soporte**
0	0.92	0.69	0.79	141
1	0.28	0.68	0.40	25
promedio macro	0.60	0.68	0.59	166
promedio ponderado	0.83	0.69	0.73	166

Cuadro 4.6: Informe de clasificación de scikit-learn en la fase de entrenamiento de la red neuronal 6 con los siguientes parámetros: función de activación: relu, alpha: 0.0001, tamaño de las capas ocultas: (20,), tasa de aprendizaje: constante.

PRUEBA	**Precisión**	**Sensiblidad**	**Puntuación F1**	**Soporte**
0	0.88	0.64	0.74	36
1	0.19	0.50	0.27	6
promedio macro	0.54	0.57	0.51	42
promedio ponderado	0.79	0.62	0.67	42

Cuadro 4.7 continuación de la página anterior

Cuadro 4.7: Informe de clasificación de scikit-learn en la fase de prueba de la red neuronal 6 con los siguientes parámetros: función de activación: relu, alpha: 0.0001, tamaño de las capas ocultas: (20,), tasa de aprendizaje: constante.

Para esta red neuronal el número de FN es mayor que los VP y FP en la fase de entrenamiento y prueba y en esta última constituyen un porcentaje considerable, aproximadamente 31 % (figura 4.5). Además, en la prueba los FP y VP tienen el mismo porcentaje, lo que conlleva a afirmar que existe la misma probabilidad de diagnosticar hígados grasos que son realmente grasos e hígados grasos que en realidad no lo son. En la prueba, el porcentaje de falsos totales es cerca del 39 %, por lo tanto, este modelo no es bueno si lo comparamos con el de la red 2 y 4, sobre todo por el valor alto que existe en los FN.

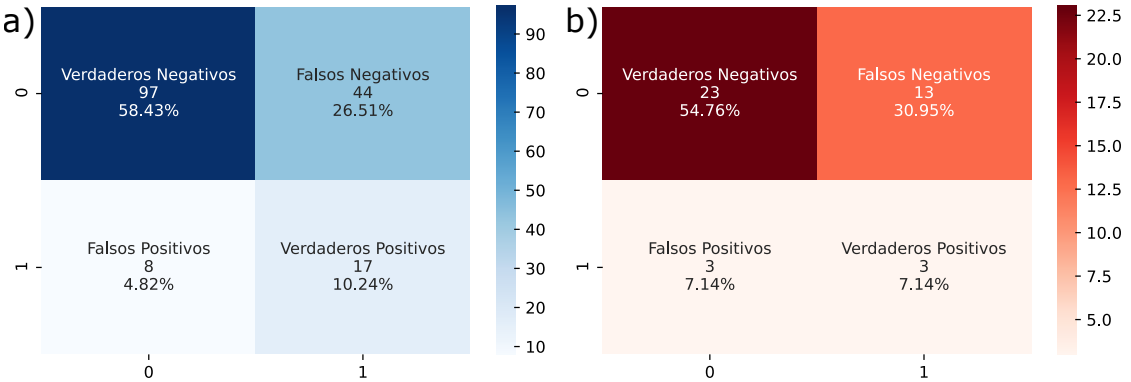


Figura 4.5: Matriz de confusión de la fase de entrenamiento (a) y prueba (b) que muestra los VP, FP, VN y FN para una red neuronal con conjuntos de entrenamiento y prueba en una proporción de 80 a 20.

Al usar una proporción de 70-30 para las fases entrenamiento-prueba de la red neuronal 8 usando tres componentes principales obtuvimos una precisión del 68 % para la primera y un 56 % para la segunda fase. En comparación con el resto de modelos, tenemos que los resultados (respecto a la precisión) de la red 8 son mejores que la red neuronal 6, en cambio la red neuronal 4 tiene una precisión menor en el entrenamiento pero mayor en la prueba. Por otro lado la red neuronal 2 es mejor en las dos fases que la red neuronal 8.

ENTRENAMIENTO	Precisión	Sensiblidad	Puntuación F1	Soporte
0	0.95	0.73	0.82	118
1	0.41	0.81	0.54	27
promedio macro	0.68	0.77	0.68	145
promedio ponderado	0.84	0.74	0.77	145

Cuadro 4.8: Informe de clasificación de scikit-learn en la fase de entrenamiento de la red neuronal 8 con los siguientes parámetros: función de activación: relu, alpha: 0.0001, tamaño de las capas ocultas: (2, 50, 3), tasa de aprendizaje: constante.

PRUEBA	Precisión	Sensiblidad	Puntuación F1	Soporte
0	0.90	0.67	0.77	54
1	0.22	0.56	0.31	9
promedio macro	0.56	0.61	0.54	63
promedio ponderado	0.80	0.65	0.70	63

Cuadro 4.9: Informe de clasificación de scikit-learn en la fase de prueba de la red neuronal 8 con los siguientes parámetros: función de activación: relu, alpha: 0.0001, tamaño de las capas ocultas: (2, 50, 3), tasa de aprendizaje: constante.

En la figura 4.6 se observa que hay más VN que VP y además, los FN son mayores que los VP. Los FN en el entrenamiento no difieran en una cantidad muy grande de los VP pero en la prueba los FN son aproximadamente el triple de los VP. Si bien los resultados del entrenamiento no son malos, los de la prueba no reflejan lo mismo porque el número de VP es muy pequeño ya que ni siquiera corresponde a un 10 % del total.

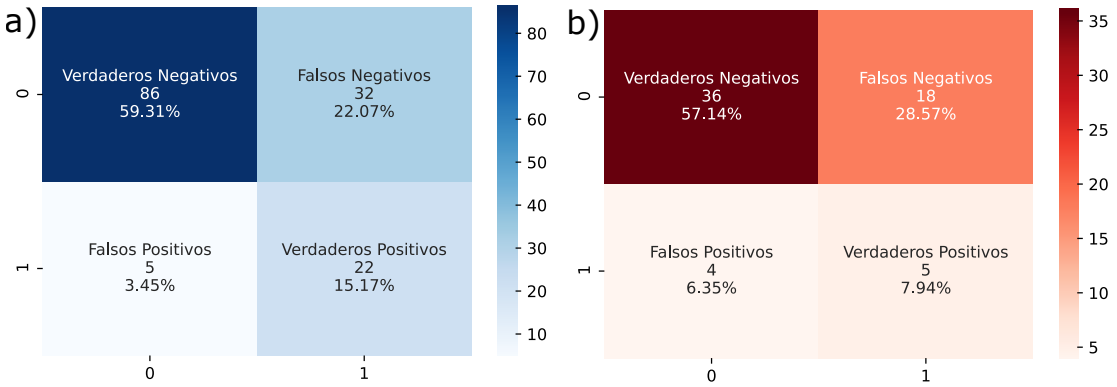


Figura 4.6: Matriz de confusión de la fase de entrenamiento (a) y prueba (b) que muestra los VP, FP, VN y FN para una red neuronal con conjuntos de entrenamiento y prueba en una proporción de 70 a 30.

Luego de analizar los diferentes modelos usados para construir las redes neuronales, tenemos que la red neuronal 2 en su fase de prueba tiene los mejores resultados con una exactitud del 66 %, una precisión del 63 % y una sensibilidad y puntuación F1 del 64 %. Este resultado no puede ser comparado con algún estudio previo pues no hay evidencia del desarrollo de métodos de clasificación con redes neuronales artificiales para la identificación del hígado graso en imágenes de resonancia magnética. Sin embargo, varios estudios han utilizado el aprendizaje profundo con imágenes de ultrasonido para la clasificación de la esteatosis. Byra et al ([78]) usaron una red neuronal convolucional (CNN) previamente entrenada con imágenes no médicas para diagnosticar la enfermedad del hígado graso. Evaluaron a 55 pacientes y obtuvieron una exactitud del 96 % y una sensibilidad del 100 %. Reddy et al ([79]) utilizó un enfoque de aprendizaje profundo similar en 157 imágenes de ultrasonido, usando el diagnóstico radiológico como la verdad básica y alcanzaron una exactitud del 91 % y una sensibilidad del 95 %.

Si bien, los resultados de exactitud y sensibilidad de los trabajos citados alcanzan un mejor rendimiento que nuestros resultados, debemos considerar que, la exactitud del ultrasonido para detectar esteatosis, es del 93 % cuando el hígado posee más del 33 % de grasa, en cambio, la resonancia magnética tiene una exactitud cercana al 100 % y puede detectar desde un 3 % de contenido de grasa [80]. A pesar que, nuestros resultados no lograron rendimientos similares a los de Byra et al y Reddy et al, hay que considerar que es complicado comparar los diversos estudios debido a

diferencias en las muestras y en las técnicas de adquisición y de clasificación usadas, así como también en los parámetros de cada red neuronal.

Capítulo 5

Conclusiones

Los diferentes tonos de grises que se derivan de una gran cantidad de datos hacen del análisis de la imagen por resonancia magnética una tarea complicada que consume mucho tiempo. El uso de un software de segmentación de regiones de interés, la radiómica y sobre todo el de redes neuronales para la implementación de modelos predictivos permiten que el tiempo de análisis disminuya. Con el empleo del software 3D-Slicer, el paquete PyRadiomics y las redes neuronales del tipo perceptrón multicapa se concluye lo siguiente:

El uso de los siguientes parámetros estadísticos: media, mediana, mínimo, máximo, varianza, skewness, kurtosis, energía, entropía, momento de diferencia inversa (MDI), contraste y correlación constituyen un grupo de características de textura útil para el desarrollo de redes neuronales. Sin embargo, el análisis multivariante de estos parámetros es fundamental debido a que un número correcto de parámetros de entrada (características) evitará un sobre ajuste de los datos en la implementación de la red, es decir, valores de exactitud y precisión del 100 %. Este resultado fue visible en una prueba preliminar realizada con las 12 características.

El proceso de estandarización es necesario cuando la escala de ciertas características es mayor que otras, por ejemplo, la diferencia de escalas entre la media y la skewness (tabla 3.1). En caso de pasar por alto esta etapa corremos el riesgo de perder mucha información pues aquellas características con mayor escala predominarán al resto y se derivan en problemas al momento de entrenar las redes neuronales.

Las técnicas de Machine Learning como las redes neuronales son de utilidad para manejar grandes cantidades de datos. El algoritmo de perceptrón multicapa para problemas de clasificación es simple y fácil de construir y los tiempos de ejecución de estos algoritmos no superan los 3 minutos, por lo tanto, pueden realizarse un

sinnúmero de pruebas en intervalos cortos de tiempo.

Tras el análisis, podemos concluir que la mejor predicción para este conjunto de datos se obtiene al reducir las características a dos dimensiones, usando un 80 % de los datos para el entrenamiento de la red neuronal y el 20 % restante para probar el modelo, donde fijamos el estado aleatorio para que los resultados sean reproducibles. La reducción de dimensionalidad de características mediante el método de componentes principales (PCA) nos permite utilizar un número mínimo de componentes sin perder mucha información y además es útil para visualizar el comportamiento del conjunto de datos como es el caso de las figuras 3.10 y 3.11. Por otro lado, dividir el conjunto de datos original en entrenamiento y prueba nos permite comprobar si el modelo generado funciona. Hay que considerar que el conjunto de datos de prueba debe ser lo suficientemente grande para obtener resultados significativos y representativos en comparación a la muestra total. Caso contrario, tendríamos sobre ajustes o sub ajustes que nos llevan a modelos demasiado simples y poco predictivos.

La red neuronal que alcanza los mejores resultados fue construida con una función de activación relu, un alpha de 0.0001, una tasa de aprendizaje constante y con un tamaño de capas ocultas de (50, 100, 50). La exactitud de este modelo es del 80 %, su precisión es 77 % y su sensibilidad 81 % en la zona de entrenamiento, en la prueba la exactitud alcanza un 66 %, la precisión un 63 % y la sensibilidad 64 %. El número de verdaderos negativos y positivos constituyen el 80 % y los falsos positivos y negativos el 20 % restante en el entrenamiento. En la prueba el primer grupo alcanza un 67 % y el segundo un 33 %. Además, comprobamos que el uso de hiperparámetros mejora los resultados de la red neuronal, sin embargo, se recomienda usar un número mayor de parámetros para realizar más combinaciones y una base de datos más grande. Además, se puede utilizar técnicas de sobremuestreo que posiblemente solucionaría el desequilibro existente entre los datos que corresponden a una zona sana y una con grasa. Otro variable de estudio puede ser la eliminación de características que no aportan información valiosa al modelo, para esto debemos usar técnicas diferentes y un análisis más profundo en cada una de las características a usar.

Bibliografía

[1] A. Patyuchenko, "Medical Image Processing: From Formation to Interpretation," *Analog Devices*, págs. 1-4, 2019.

[2] R. Choplin, J. Boehme y D. Maynard, "Picture archiving and communication systems: an overview," *RadioGraphics*, vol. 12, n.º 1, págs. 127-129, 1992. DOI: 10.1148/radiographics.12.1.1734458.

[3] X. W. Gao, Y. Qia y R. Hui, "The state of the art of medical imaging technology: from creation to archive and back," *The Open Medical Informatics Journal*, vol. 5, n.º 1, págs. 73-85, 2011. DOI: 10.2174/1874431101105010073.

[4] Y. M. Y. Abdallah y T. Alqahtani, *Research in Medical Imaging Using Image Processing Techniques*. IntechOpen, 2019, págs. 1-6. DOI: 10.5772/intechopen.84360.

[5] D. A. Lisle, *Imaging for students*, Cuarta edición. Hodder Arnold, 2012, págs. 1-12.

[6] A. Bustos, A. Pertusa, J. M. Salinas y M. de la Iglesia Vayá, "PadChest: A large chest X-ray image dataset with multi-label annotated reports," *Medical Image Analysis*, vol. 66, págs. 1-11, 2020. DOI: 10.1016/j.media.2020.101797.

[7] C. Guy y D. Ffytche, *An Introduction to The Principles of Medical Imaging*, Edición revisada. London: Imperial College Press, 2005, págs. 1-19.

[8] K. Yan, X. Wang, L. Lu y R. M. Summers, "DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning," *Journal of Medical Imaging*, vol. 5, n.º 3, pág. 036 501, 2018. DOI: 10.1117/1.JMI.5.3.036501.

[9] A. Vargas, L. M. Amescua-Guerra, M. A. Bernal y C. Pined, "Principios físicos básicos del ultrasonido, sonoanatomía del sistema musculoesquelético y

artefactos ecográficos," *Acta Ortopédica Mexicana*, vol. 22, n.º 6, págs. 361-373, 2008.

[10] J. Puylaert, *US of the GI tract - Normal Anatomy*, 2020. dirección: https://radiologyassistant.nl/abdomen/ultrasound/lk-jg-1-1#normal-anatomy-stomach.

[11] V. Grover, J. Tognarelli, M. Crossey, J. Cox, S. T. Robinson y M. McPhail, "Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians," *Journal of Clinical and Experimental Hepatology*, vol. 5, n.º 3, págs. 246-255, 2015. DOI: 10.1016/j.jceh.2015.08.001.

[12] K. Möllenhof, A. M. Oros-Peusquens y J. N. Shah, *Introduction to the Basics of Magnetic Resonance Imaging*. Molecular Imaging in the Clinical Neurosciences, 2012, págs. 1-6. DOI: 10.1007/7657_2012_56.

[13] R. Trivedi, L. Saba y J. S. Suri, *3D Imaging Technologies in Atherosclerosis*. Morgan Kaufmann Publishers Inc., 2015, págs. 153-163. DOI: 10.1007/978-1-4899-7618-5.

[14] M. Lakrimi, A. M. Thomas, G. Hutton, M. Kruip, R. Slade, P. Davis, A. J. Johnstone, M. J. Longfield, H. Blakes, S. Calvert, M. Smith y C. A. Marshall, "The principles and evolution of magnetic resonance imaging," *Journal of Physics: Conference Series*, vol. 286, pág. 012 016, 2011. DOI: 10.1088/1742-6596/286/1/012016.

[15] A. Bjørnerud, "The Physics of Magnetic Resonance Imaging," en *Quantitative MRI of the brain: Measuring changes caused by disease*, John Wiley & Sons, 2008, págs. 7-50. DOI: 10.1002/0470869526.

[16] Y. X. Wáng, X. Wang, P. Wu, Y. Wang, W. Chen, H. Chen y J. Li, "Topics on quantitative liver magnetic resonance imaging," *Quantitative imaging in medicine and surgery*, vol. 9, n.º 11, págs. 1840-1890, 2019. DOI: 10.21037/qims.2019.09.18.

[17] K. Karsch, Q. He e Y. Duan, "A Fast, Semi-Automatic Brain Structure Segmentation Algorithm for Magnetic Resonance Imaging," *Quantitative imaging in medicine and surgery*, págs. 297-302, 2019. DOI: 10.1109/BIBM.2009.40.

[18] P. Abolmaesumi, S. Ahmad, W. Anderson, A. Asman, A. Ben-Cohen, E. Boctor, S. Bodenstedt, E. Breton, X. Cao, A. Carass y M. J. Cardoso, *Handbook of Medical Image Computing and Computer Assisted Intervention*. Academic Press, 2020, págs. 429-441.

[19] G. Dougherty, "Image analysis in medical imaging: Recent advances in selected examples," *Biomedical Imaging and Intervention Journal*, vol. 6, n.º 3, 2010. DOI: 10.2349/biij.6.3.e32.

[20] N. G. Castillo, "Contributions to Automatic and Unsupervised MRI Brain Tumor Segmentation: A new Fuzzy Aproach," Tesis doct., Universitat Politècnica de Catalunya, 2010.

[21] E. Göçeri, M. Gürcan y O. Dicle, "Fully automated liver segmentation from SPIR image series," *Computers in Biology and Medicine*, vol. 53, págs. 265-278, 2014. DOI: 10.1016/j.compbiomed.2014.08.009.

[22] E. Chartampilas, "Imaging of nonalcoholic fatty liver disease and its clinical utility," *Hormones*, vol. 17, n.º 1, págs. 69-81, 2018. DOI: 10.1007/s42000-018-0012-x.

[23] S. Luo, X. Li y J. Li, "Review on the Methods of Automatic Liver Segmentation from Abdominal Images," *Journal of Computer and Communications*, vol. 2, n.º 2, págs. 1-7, 2014. DOI: 10.4236/jcc.2014.22001.

[24] R. P. Mathew y S. Venkatesh, "Review on the Methods of Automatic Liver Segmentation from Abdominal Images," *Applied Radiology*, vol. 48, n.º 5, págs. 13-20, 2019.

[25] L. Vu, J. Morelli y J. Szklaruk, "Basic MRI for the liver oncologists and surgeons," *Journal of hepatocellular carcinoma*, vol. 5, págs. 37-50, 2018. DOI: 10.2147/JHC.S154321.

[26] E. Göçeri, "A comparative evaluation for liver segmentation from spir images and a novel level set method using signed pressure force function," Tesis doct., Izmir Institute of Technology, 2013.

[27] M. Kass, A. Witkin y D. Terzopoulos, "Snakes: Active Contour Models," *International Journal of Computer Vision*, vol. 1, n.º 4, págs. 321-331, 1988. DOI: 10.1007/BF00133570.

[28] A. X. Falcão y J. K. Udupa, "A 3D generalization of user-steered live-wire segmentation," *Medical Image Analysis*, vol. 4, n.º 4, págs. 389-402, 2000. DOI: 10.1016/s1361-8415(00)00023-2.

[29] A. Gotra, L. Sivakumaran, G. Chartrand, K.-N. Vu, F. Vandenbroucke-Menu, C. Kauffmann, S. Kadoury, B. Gallix, J. de Guise y A. Tang, "Liver segmentation: indications, techniques and future directions," *Insights into imaging*, vol. 8, n.º 4, págs. 377-392, 2017. DOI: 10.1007/s13244-017-0558-1.

[30] F. Malmberg, E. Vidholm e I. Nyström, "A 3D Live-Wire Segmentation Method for Volume Images Using Haptic Interaction," *Discrete Geometry for Computer Imagery*, vol. 4245, págs. 663-673, 2006. DOI: 10.1007/11907350_56.

[31] A. S. Kulkarni1 y R. S. Kamathe, "MRI Brain Image Segmentation by Edge Detection and Region Selection," *International Journal of Technology and Science*, vol. 1, n.º 2, 2014.

[32] R. Nagaraju y J. Reddy, "Image segmentation in medical imaging: state of the art," *International Journal of Recent Scientific Research*, vol. 8, n.º 7, págs. 18 903-18 906, 2017. DOI: 10.24327/IJRSR.

[33] A. Materka, "Texture analysis methodologies for magnetic resonance imaging," *Dialogues in clinical neuroscience*, vol. 6, n.º 2, págs. 243-250, 2004. DOI: 10.31887/DCNS.2004.6.2/amaterka.

[34] A. Larroza, V. Bodí y D. Moratal, *Texture Analysis in Magnetic Resonance Imaging: Review and Considerations for Future Applications, Assessment of Cellular and Organ Function and Dysfunction using Direct and Derived MRI Methodologies*. IntechOpen, 2016, págs. 75-88. DOI: 10.5772/64641.

[35] V. Parekh y M. A. Jacobs, "Radiomics: a new application from established techniques," *Expert Rev Precis Med Drug Dev.*, vol. 1, n.º 2, págs. 207-226, 2016. DOI: 10.1080/23808993.2016.1164013.

[36] R. J. Gillies, P. E. Kinahan y H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data," *Radiology*, vol. 278, n.º 2, págs. 563-577, 2016. DOI: 10.1148/radiol.2015151169.

[37] P. Lambin, R. Leijenaar, T. Deist, J. Peerlings, E. Jong, J. van Timmeren, S. Sanduleanu, R. Larue, A. Even, A. Jochems, Y. van Wijk, H. Woodruff, J. van Soest, T. Lustberg, E. Roelofs, W. van Elmpt, A. Dekker, F. Mottaghy, J. Wildberger y S. Walsh, "Radiomics: the bridge between medical imaging and personalized medicine," *Nature reviews. Clinical oncology*, vol. 14, n.º 12, págs. 749-762, 2017. DOI: 10.1038/nrclinonc.2017.141.

[38] J. E. van Timmeren, D. Cester, S. Tanadini-Lang, H. Alkadhi y B. Baessler, "Radiomics in medical imaging—"how-to" guide and critical reflection," *Insights into Imaging*, vol. 11, n.º 91, págs. 1-16, 2020. DOI: 10.1186/s13244-020-00887-2.

[39] G. D. Tourassi, "Journey toward Computer-aided Diagnosis: Role of Image Texture Analysis," *Radiology*, vol. 213, n.º 2, págs. 317-320, 1999. DOI: 10.1148/radiology.213.2.r99nv49317.

[40] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, n.º 3, págs. 379-423, 1948. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[41] R. M. Haralick, K. Shanmugam e I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, n.º 6, págs. 610-621, 1973. DOI: 10.1109/TSMC.1973.4309314.

[42] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer Graphics and Image Processing*, vol. 4, n.º 2, págs. 172-179, 1975. DOI: 10.1016/S0146-664X(75)80008-6.

[43] V. Parekh y M. A. Jacobs, "Radiomics: a new application from established techniques," *Expert Review of Precision Medicine and Drug Development*, vol. 1, n.º 2, págs. 207-226, 2016. DOI: 10.1080/23808993.2016.1164013.

[44] T. Randen y J. H. Husoy, "Filtering for texture classification: a comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, n.º 4, págs. 291-310, 1999. DOI: 10.1109/34.761261.

[45] H. Aerts, E. R. Velazquez y R. Leijenaar, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, n.º 4006, págs. 1-9, 2014. DOI: 10.1038/ncomms5006.

[46] A. Zwanenburg, M. Vallières, M. A. Abdalah, H. Aerts, V. Andrearczyk, A. Apte, S. Ashrafinia, S. Bakas, R. J. Beukinga y R. Boellaard, "The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping," *Radiology*, vol. 295, n.º 2, págs. 328-338, 2020. DOI: 10.1148/radiol.2020191145.

[47] S. Mukherjee, A. Kamal y K. Kumar, "Optimization of Material Removal Rate During Turning of SAE 1020 Material in CNC Lathe using Taguchi Technique," *Procedia Engineering*, vol. 97, págs. 29-35, 2014. DOI: 10.1016/j.proeng.2014.12.221.

[48] N. Aggarwal y R. Agrawal, "First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images," *Journal of Signal and Information Processing*, vol. 3, n.º 2, págs. 146-153, 2012. DOI: 10.4236/jsip.2012.32019.

[49] R. Haralick y L. Watson, "A facet model for image data," *Computer Graphics and Image Processing*, vol. 15, págs. 113-129, 1981.

[50] G. J. Djuričić, M. Radulovic, J. P. Sopta, M. Nikitović y N. T. Milošević, "Fractal and Gray Level Cooccurrence Matrix Computational Analysis of Primary Osteosarcoma Magnetic Resonance Images Predicts the Chemotherapy Response," *Frontiers in Oncology*, vol. 7, pág. 246, 2017. DOI: 10.3389/fonc.2017.00246.

[51] N. B. Bahadure, A. K. Ray y H. P. Thethi, "Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM," *International Journal of Biomedical Imaging*, págs. 1-12, 2017. DOI: 10.1155/2017/9749108.

[52] S. Rizzo, F. Botta, S. Raimondi, D. Origgi, C. Fanciullo, A. G. Morganti y M. Bellomi, "Radiomics: the facts and the challenges of image analysis," *European radiology experimental*, vol. 2, n.º 3, págs. 1-8, 2018. DOI: 10.1186/s41747-018-0068-z.

[53] H. J. Park, B. Park y S. S. Lee, "Radiomics and Deep Learning: Hepatic Applications," *Korean Journal of Radiology*, vol. 21, n.º 4, págs. 387-401, 2020. DOI: 10.3348/kjr.2019.0752.

[54] F. Pesapane, M. Codari y F. Sardanelli, "Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine," *European Radiology Experimental*, vol. 2, n.º 35, 2018. DOI: 10.1186/s41747-018-0061-6.

[55] G. Currie, K. E. Hawk, E. Rohren, A. Vial y R. Klein, "Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging," *Journal of Medical Imaging and Radiation Sciences*, vol. 50, n.º 4, págs. 477-487, 2019. DOI: 10.1016/j.jmir.2019.09.005.

[56] J. Jiang, P. Trundle y J. Ren, "Medical image analysis with artificial neural networks," *Computerized Medical Imaging and Graphics*, vol. 34, n.º 8, págs. 617-631, 2010. DOI: 10.1016/j.compmedimag.2010.07.003.

[57] M. Zakaria, M. AL-Shebany y S. Sarhan, "Artificial Neural Network : A Brief Overview," *International Journal of Engineering Research and Applications*, vol. 4, n.º 2, págs. 7-12, 2014.

[58] M. M. Raju, R. K. Srivastava, D. Bisht, H. C. Sharma y A. Kumar, "Development of Artificial Neural-Network-Based Models for the Simulation of Spring Discharge," *Advances in Artificial Intelligence*, vol. 2011, págs. 1-3, 2011. DOI: 10.1155/2011/686258.

[59] N. Manjula y B. K. Reddy, "Deep Learning Applications in Medical Image Analysis," *International Journal of Advanced Science and Technology*, vol. 29, n.º 12, págs. 2391-2401, 2020.

[60] G. S. Fu, Y. L. Schwartz, Q. H. Lin y D. Zhang, "Machine Learning for Medical Imaging," *International Journal of Advanced Science and Technology*, vol. 2019, 2019. DOI: 10.1155/2019/9874591.

[61] G. Chartrand, P. M. Cheng, E. Vorontsov, M. Drozdzal, S. Turcotte, C. J. Pal, S. Kadoury y A. Tang, "Deep Learning: A Primer for Radiologists," *Radiographics*, vol. 37, n.º 7, págs. 2113-2131, 2017. DOI: 10.1148/rg.2017170077.

[62] A. S. Lundervold y A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift fur medizinische Physik*, vol. 29, n.º 2, págs. 102-127, 2019. DOI: 10.1016/j.zemedi.2018.11.002.

[63] J. Ker, L. Wang, J. Rao y T. Lim, "Deep Learning Applications in Medical Image Analysis," *IEEE Access*, vol. 6, págs. 9375-9389, 2018. DOI: 10.1109/ACCESS.2017.2788044.

[64] A. Krizhevsky, I. Sutskever y G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," en *Advances in Neural Information Processing Systems*, F. Pereira, C. J. Burges, L. Bottou y K. Q. Weinberger, eds., vol. 25, Curran Associates, Inc., 2012, págs. 1097-1105.

[65] Y. Lecun, L. Bottou, Y. Bengio y P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, n.º 11, págs. 2278-2324, 1998. DOI: 10.1109/5.726791.

[66] S. Reede y C. Sirlin, "Quantification of Liver Fat with Magnetic Resonance Imaging," *Magn Reson Imaging Clin N Am*, vol. 15, n.º 3, págs. 337-357, 2010. DOI: 10.1016/j.mric.2010.08.013.

[67] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. Miller, S. Pieper y R. Kikinis, "3D Slicer as an Image Computing Platform for the Quantitative Imaging Network," *Magnetic Resonance Imaging*, vol. 30, n.º 9, págs. 1323-1341, 2012. DOI: 10.1016/j.mri.2012.05.001.

[68] C. C. Dalavia, S. M. Goldman, H. J. de Farias e Melo, C. E. Kater, J. Szejnfeld, W. Iared y S. A. Ajzen, "The value of signal intensity on T1-weighted chemical shift magnetic resonance imaging combined with proton magnetic resonance spectroscopy for the diagnosis of adrenal adenomas," *Radiologia Brasileira*, vol. 53, n.º 2, págs. 86-94, 2020. DOI: 10.1590/0100-3984.2019.0095.

[69] D. W. Scott, "Sturges' rule," *WIREs Computational Statistics*, vol. 1, n.º 3, págs. 303-306, 2009. DOI: 10.1002/wics.35.

[70] J. J. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper y H. J. Aerts, "Computational Radiomics System to Decode the Radiographic Phenotype," *Cancer Research*, vol. 77, n.º 21, págs. 104-107, 2017. DOI: 10.1158/0008-5472.CAN-17-0339.

[71] S. Naganawa, K. Enooku, R. Tateishi, H. Akai, K. Yasaka, J. Shibahara, T. Ushiku, O. Abe, K. Ohtomo y S. Kiryu, "Imaging prediction of nonalcoholic steatohepatitis using computed tomography texture analysis," *European Radiology*, vol. 28, n.º 7, págs. 3050-3058, 2018. DOI: 10.1007/s00330-017-5270-5.

[72] M. Fishbein, F. Castro, S. Cheruku, S. Jain, B. Webb, T. Gleasone, Stevens y W. Ross, "Hepatic MRI for fat quantitation: its relationship to fat morphology, diagnosis, and ultrasound," *Journal of Clinical Gastroenterology*, vol. 39, n.º 7, págs. 619-625, 2005. DOI: 10.1097/00004836-200508000-00012.

[73] F. H. Cassidy, T. Yokoo, L. Aganovic, R. F. Hanna, M. Bydder, M. S. Middleton, G. Hamilton, A. D. Chavez, J. B. Schwimmer y C. B. Sirlin, "Fatty Liver Disease: MR Imaging Techniques for the Detection and Quantification of Liver Steatosis," *RadioGraphics*, vol. 29, n.º 1, págs. 231-260, 2009. DOI: 10.1148/rg.291075123.

[74] Q. Li, M. Dhyani, J. R. Grajo, C. Sirlin y A. E. Samir, "Current status of imaging in nonalcoholic fatty liver disease," *World Journal of Hepatology*, vol. 10, n.º 8, págs. 530-542, 2018. DOI: 10.4254/wjh.v10.i8.530.

[75] P. Schober, C. Boer y L. Schwarte, "Correlation Coefficients: Appropriate Use and Interpretation," *Anesthesia & Analgesia*, vol. 126, n.º 5, págs. 1763-1768, 2018. DOI: 10.1213/ANE.0000000000002864.

[76] I. T. Jolliffe1 y J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions*, vol. 374, n.º 2065, 2016. DOI: 10.1098/rsta.2015.0202.

[77] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot y E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, págs. 2825-2830, 2011.

[78] M. Byra, G. Styczynski, C. Szmigielski, P. Kalinowski, Ł. Michałowski, R. Paluszkiewicz, B. Ziarkiewicz-Wróblewska, K. Zieniewicz, P. Sobieraj y A. Nowicki, "Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images," *International Journal of Computer*

Assisted Radiology and Surgery, vol. 13, n.º 12, págs. 1895-1903, 2018. DOI: 10.1007/s11548-018-1843-2.

[79] S. Reddy, R. Bharath y P. Rajalakshmi, "Classification of Nonalcoholic Fatty Liver Texture Using Convolution Neural Networks," en *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services*, 2018, págs. 1-5. DOI: 10.1109/HealthCom.2018.8531193.

[80] M. Graffigna, N. Catoira, J. Soutelo, A. Azpelicueta, G. Berg, C. Perel, M. E. Migliano, M. Aranguren, C. Musso y J. Farias, "Diagnóstico de esteatosis hepática por métodos clínicos, bioquímicos y por imágenes," *Revista Argentina de Endocrinología y Metabolismo*, vol. 54, n.º 1, págs. 37-46, 2017. DOI: 10.1016/j.raem.2016.12.001.

Apéndice A

Repositorio: Datos y Programación

Para descargar los datos, gráficas y códigos de programación de este trabajo de titulación se requiere debes seguir las guías de este Apéndice. Los recursos se encuentran en el repositorio GitHub: Tesis_Alejandra_Ichina, disponible en el siguiente link `https://github.com/ai232/Tesis_Alejandra_Ichina.git`.
A continuación se detallan las instrucciones generales para su uso. Este archivo también forma parte del repositorio y se lo encuentra como README.md.

Tesis_Alejandra_Ichina

Este repositorio contiene todos los códigos utilizados para llevar a cabo la tesis titulada: "Desarrollo de algoritmos para el análisis y procesamiento de imágenes de resonancia magnética para diferenciar los tejidos presentes en el hígado" para la obtención del título de Físico de Alejandra Estefanía Ichina López.

Segmentos, histogramas y extracción de características

- El archivo **segmentos e histogramas.txt** contiene el código para crear la máscara de la segmentación, para la segmentación semiautomática y para la graficación de histogramas. Este debe ser implementado en el interfaz der Python del software 3D-Slicer.

- El archivo **graf.pro** permite graficar los histogramas normalizados con su respectiva función de distribución. Los datos necesarios para realizar esta tarea están disponibles en las carpetas: **Datos procesados (1-10).zip**, **Datos procesados (11-20).zip** y **Datos procesados (21-26).zip**. En las carpetas mencionadas se encuentran los histogramas no normalizados extraídos directamente del 3D-Slicer.

- La carpeta **PyRadiomics personalizado.zip** contiene los archivos *.py* que deben ser reemplazados para extraer únicamente las 12 características utilizadas en este trabajo de titulación.

Instalación y modificación de PyRadiomics

1. Utilizamos el proceso de instalación 3 (Install from source) disponible en: https://pyradiomics.readthedocs.io/en/latest/installation.html#installation. Es decir, primero clonamos el repositorio

   ```
   git clone git://github.com/Radiomics/pyradiomics
   ```

   Luego escribimos en la terminal lo siguiente:

   ```
   cd pyradiomics
   python -m pip install -r requirements.txt
   python setup.py install
   ```

2. Se creará una carpeta pyradiomics la cual contiene la carpeta radiomics. En esta última se reemplazarán 8 archivos .py. Estos nuevos archivos son los disponibles en PyRadiomics personalizado.zip

3. Finalmente se volverá a correr la línea:

   ```
   python setup.py install
   ```

Análisis de características y redes neuronales

El repositorio contiene las características de todos los segmentos con sus respectivas etiquetas (archivos de datos .csv). Los archivos

- **Análisis_de_características.ipynb**
- **Red_PCA2_(0_2).ipynb**
- **Red_PCA2_(0_3).ipynb**
- **Red_PCA3_(0_2).ipynb**
- **Red_PCA3_(0_3).ipynb**

son importantes para la construcción de las diferentes redes neuronales. En cada .pynb existe una pequeña guía que describe las diferentes actividades realizadas.

Apéndice B

Características radiómicas

	Mean	Median	Min	Max	Variance	Skewness	Kurtosis	Energy	Entropy	IDM	Contrast	Correlation	Percentage
1	114.0409463	115	37	160	219.358651	-0.814405575	5.5976505	0.307959614	2.424936384	0.840067601	0.354154997	0.559278004	23.16290847
1	120.7076503	120	81	158	170.157701	0.074498225	2.775905429	0.272841293	2.307922482	0.856137194	0.291179403	0.601057465	21.61815189
1	105.0273723	105	71	143	141.4353821	0.146668827	2.988501925	0.301215515	2.161353916	0.860138371	0.280875995	0.560229884	18.18000036
1	110.6806569	110	78	142	130.3815966	0.11814594	2.652145952	0.353975935	2.09012251	0.8630116	0.273976801	0.536717081	22.90603856
1	113.8568824	113	78	160	211.5137015	0.395898039	3.006044808	0.300917086	2.331390598	0.875840189	0.248895437	0.701315353	24.66398747
1	115.0518653	114	74	176	278.9918505	0.320523945	2.717618064	0.225662347	2.6044857	0.867017148	0.266540418	0.758973938	17.51283369
1	150.9372157	155	97	199	500.9032465	-0.408924334	2.361602489	0.186890409	2.874542318	0.865315555	0.272818267	0.837336751	15.4191754
1	157.2848044	159	29	240	1343.224619	-0.301295755	2.786679255	0.08305752	3.982878725	0.785063114	0.486637318	0.889134259	12.87260762
2	124.7692308	125	86	157	212.7605107	0.063587331	2.619782871	0.212274957	2.644232401	0.81458926	0.389669275	0.572097984	23.12229379
2	146.0362903	146	107	177	199.1801346	-0.132634369	2.674043641	0.256485284	2.359950545	0.858041511	0.28663191	0.6397603	19.55541009
2	107.6842105	106	79	144	138.1755807	0.56757176	3.199625907	0.297990993	2.246543667	0.834572105	0.338905742	0.472534769	19.51586147
2	110.0441767	111	78	148	124.2028677	-0.127589853	2.985861172	0.365327248	2.010036775	0.843322834	0.313354332	0.389577855	20.56828637
2	124.8380567	117	79	209	703.7875395	1.264111463	3.81858361	0.241435352	2.908851957	0.819865713	0.373843234	0.820287315	22.32145233
2	138.8016194	136	91	201	467.3250176	0.426613636	3.203648429	0.17039781	3.168134796	0.806406469	0.408716829	0.765036543	11.74925732
2	206.7379032	211	28	327	2658.47566	-0.333597883	3.118580126	0.035752123	5.195134382	0.577897947	1.939361305	0.767732865	4.250168183
2	118.0443548	118	93	143	114.2359359	0.015347402	2.410281274	0.285140496	2.184374383	0.81833772	0.374053593	0.359216104	24.41002406
3	182.563822	185	98	227	427.2816671	-0.968700443	4.266113077	0.222671176	2.844538989	0.853825311	0.303778389	0.806138174	20.47455943
3	263.7747813	265	130	327	756.5185185	-0.515350129	3.938177562	0.148841272	3.249677811	0.855155345	0.2960829	0.881193765	15.51036216
3	212.9774052	219	127	282	913.7043	-0.671195372	2.997058517	0.131676327	3.464856982	0.835219627	0.344198611	0.888933026	0.724178301
3	166.6754194	166	130	217	211.0959603	0.139229171	2.854479145	0.296086558	2.30559348	0.870801949	0.26068137	0.664108264	0.368643286
3	185.1058394	189	89	254	783.8844185	-0.777430003	3.816484367	0.151486928	3.369137931	0.829806451	0.354587578	0.866913684	0.054081533
3	163.1653314	163	96	220	614.28075	-0.154255047	2.469769503	0.136736571	3.29611496	0.828884435	0.354575712	0.834549479	0.256999659
3	207.2784257	205	153	282	470.4079019	0.589575101	3.611267982	0.171219463	3.086860961	0.828042177	0.356707334	0.791161666	1.695688825
3	209.9233577	211.5	116	270	568.4985055	-0.254770406	3.232873684	0.145418958	3.277716594	0.825057971	0.370939341	0.816096935	2.0627432
4	71.58912387	69	19	245	1102.703286	0.88044152	4.802678434	0.098924713	3.867456326	0.76559681	0.64309846	0.817991732	2.718318101
4	200.3356855	197	145	270	583.3157427	0.5320904	2.726887055	0.177674714	3.028555963	0.858782458	0.283074062	0.854571057	2.136347312
4	197.5909091	195	148	264	384.4922406	0.450914002	3.339259928	0.210823803	2.841863302	0.843864654	0.328246978	0.749676076	0.431486471
4	178.2240161	180	121	232	535.7762547	-0.229278749	2.201778505	0.158885787	3.043424637	0.851356017	0.300482853	0.838400224	0.845606487
4	68.99393939	66	19	134	572.7959229	0.332331381	2.367867951	0.128774097	3.361386684	0.793368186	0.485268314	0.75375165	20.5184511
4	269.159919	275	84	406	2700.579689	-0.493401948	3.482462085	0.039895993	5.225709932	0.617198164	1.518966336	0.824877736	7.516927885

Cuadro B.1 continuación de la página previa

4	93.31350806	93	56	154	196.5640111	0.425924971	3.700090965	0.297122878	2.300947702	0.861105635	0.280343961	0.629566858	3.785980309
4	143.136226	142	83	268	688.1398683	1.395002873	7.317791664	0.183224477	3.133059606	0.83949345	0.344593802	0.845144923	0.417757825
5	308.2421796	310	154	405	1535.755678	-0.508221884	3.79119483	0.080995863	4.136166222	0.777943506	0.521794034	0.897658511	3.400389447
5	287.8878788	289	211	342	552.2349036	-0.457129686	3.110719783	0.170661379	3.106053378	0.848686441	0.309023707	0.838252228	17.84211146
5	207.4424242	213	104	282	1958.788099	-0.451539181	2.20023586	0.076155071	4.106208516	0.796082304	0.46841664	0.92799813	16.76035124
5	307.2158635	318	134	383	1850.65521	-1.283602803	4.749015141	0.10580856	4.042895499	0.784216936	0.567865461	0.908314799	3.028618773
5	288.8709677	294	179	376	1645.678915	-0.318456891	2.686027529	0.072738616	4.227809655	0.76922276	0.556417978	0.895142944	20.28901931
5	204.077621	215	88	300	1774.309499	-0.689653888	2.732669966	0.089717367	4.158345774	0.762935701	0.599611093	0.898306158	16.98301246
5	306.7911201	310	204	384	1309.487146	-0.25254005	2.337152424	0.082840225	4.027321139	0.785942822	0.498239746	0.887107186	15.31498158
5	239.7477296	235	135	350	1154.592263	0.473845549	3.409601355	0.097888592	3.993293466	0.768037325	0.588057749	0.85035103	10.76188855
6	223.7654723	224	163	285	569.3326189	-0.05903814	2.516858218	0.090376817	3.855056947	0.667741611	1.055918901	0.48010697	6.005984751
6	212.5700326	214	151	267	390.4144765	-0.243953917	2.897127431	0.122307574	3.477529252	0.67599267	0.972141628	0.316245593	7.253802656
6	183.8187702	182	88	323	1160.329615	0.4368968	4.161030802	0.073417773	4.601197984	0.584270477	2.202827381	0.436149324	5.20433776
6	234.9084967	236	126	301	1001.775941	-0.412488009	3.08814192	0.078400027	4.089563545	0.70488122	0.839047234	0.747175873	6.974507944
6	231.3147541	230	160	287	538.2484708	0.064914663	2.788743824	0.1196831	3.506599389	0.735882914	0.657570557	0.628909444	7.530541877
6	231.5179153	229	171	302	737.8197116	0.266430321	2.515915616	0.086620137	3.899929809	0.690681441	0.868981738	0.647426422	5.336273386
6	276.9868852	278	191	363	1419.580156	0.050247156	2.31921664	0.056529738	4.506830743	0.668220975	1.015985271	0.778852071	5.564343284
6	178.224026	177	119	260	452.0504617	0.420278473	3.912834327	0.124735208	3.497097555	0.680423691	0.943535898	0.398846285	5.373043139
7	204.1837945	203	132	289	453.7083145	0.343298076	3.565659564	0.14469479	3.29435144	0.775454032	0.501968071	0.695716748	1.404894163
7	250.6181998	251	165	347	835.1203018	0.103309765	2.852783779	0.103187182	3.725052904	0.776211146	0.504779592	0.815084374	1.915011186
7	240.1997034	237	159	533	1224.231003	3.128717547	21.08133935	0.126433511	3.738274009	0.744859121	0.824305326	0.798588545	0.715871459
7	281.0247402	265	144	645	4837.734173	2.274716214	9.888439393	0.060324318	5.025580098	0.676290146	1.363196681	0.914226837	1.452327317
7	263.7791502	258.5	174	414	1768.147372	0.321479198	2.372139172	0.063700868	4.369245708	0.750676198	0.625908211	0.892377882	1.513708853
7	360.553907	356	255	556	1577.842544	0.555965929	3.727890271	0.069309124	4.355774411	0.736615483	0.73071509	0.857486223	1.379902747
7	239.0786739	247	47	467	2678.384211	-0.517845624	4.575959955	0.050370421	4.888321788	0.682419901	0.992049436	0.883463916	2.065514213
7	254.7388724	256	56	454	4074.843781	-0.087520097	4.068209793	0.040580658	5.308340187	0.652228575	1.18174333	0.90874716	2.307320745
8	96.12408759	94.5	68	165	162.7145293	1.238335695	6.964380156	0.342974155	2.156607768	0.849591511	0.341761105	0.473866789	2.33783556
8	85.81751825	85	51	113	126.3097661	0.113682708	2.684431117	0.360979237	2.076358087	0.856638345	0.286723311	0.492408433	0.480152245
8	82.84306569	79	44	172	446.8038388	1.701776972	6.709455119	0.201974671	2.941433779	0.771121902	0.619070532	0.603392564	0.249822542
8	117.1167883	115	65	174	283.7746817	0.35514332	3.269129989	0.258212404	2.566125019	0.841646903	0.365063352	0.58748159	1.509568328
8	185.1642336	186	129	243	445.1372609	0.156356145	2.44394313	0.167150519	2.99323939	0.828515912	0.362480371	0.766371809	0.167166088

Cuadro B.1 continuación de la página previa

8	146.9343066	144	111	193	248.7402099	0.334470938	2.417484972	0.254254098	2.527688492	0.835224908	0.356110106	0.585036863	0.24246529
8	86.12	82	42	153	506.7892364	0.600693432	2.892184173	0.144502957	3.244124677	0.781925942	0.5151623	0.72752107	14.64407477
8	58.14492754	59	26	83	111.305083	-0.254301039	2.674986804	0.40622518	1.84444743	0.862333076	0.280172558	0.380224708	0.446378994
9	139.010989	139	105	193	186.1232004	0.164017049	3.028695683	0.337877232	2.13605298	0.902323271	0.195353458	0.740744958	6.073121406
9	103.497561	103	81	146	76.82804283	0.738511747	4.396827082	0.358979756	1.811674621	0.891884351	0.216231298	0.576018694	6.251690789
9	110.2066015	105	68	241	554.6993696	2.850286817	12.78079735	0.293804218	2.59714803	0.843559774	0.349342968	0.824364318	6.443364161
9	104.6593407	105	69	142	154.661729	-0.036301291	3.268201986	0.330264336	2.124603489	0.881485836	0.237028329	0.648629612	6.100885876
9	240.3321123	247	147	306	1301.066747	-0.158201096	1.800213839	0.105869582	3.656515195	0.829253554	0.375553711	0.912384651	2.917621706
9	147.6581197	146	109	203	224.5961825	0.374143137	3.127716925	0.312582063	2.19025441	0.898522936	0.205288759	0.741114051	5.900342863
9	129.6344743	130	99	168	141.1438971	-0.049932911	2.579615926	0.331025637	2.014289731	0.87773765	0.246083142	0.575234339	6.596824632
9	23.00731707	22	6	59	68.16580012	0.882770454	3.888480209	0.326391136	1.845708892	0.85317778	0.295196833	0.390776949	9.118894057
10	110.6569122	111	81	145	117.3716954	-0.049960211	2.819821921	0.426696499	1.872736731	0.884809321	0.232299654	0.528432406	0.829600796
10	145.3629032	145	107	187	186.0275754	0.122178284	2.62791036	0.297452871	2.223639942	0.876852812	0.248211309	0.656421456	0.323342265
10	131.4561049	134	74	179	447.2228462	-0.335635524	2.532530699	0.188905988	2.850442665	0.867431368	0.265137263	0.83200156	2.063246641
10	102.2383838	102	74	133	78.98963779	-0.052209509	3.026323919	0.344201185	1.799277554	0.877527081	0.244945839	0.495409928	0.430060606
10	104.5060606	106	71	130	132.5226905	-0.365562854	2.616688583	0.364203996	1.899461003	0.894125079	0.2123895	0.614774005	0.031118048
10	115.84375	114	76	158	207.6540134	0.439940262	2.922216875	0.32496404	2.248092026	0.886703299	0.227871357	0.71320367	3.370311632
10	122.0656566	122	79	155	124.6512448	0.001718508	3.675725707	0.317806121	2.026792197	0.879849074	0.24094151	0.600028486	0.540518639
10	132.9230769	133	96	175	139.4961071	0.002286258	2.722596804	0.352207395	2.04492081	0.886823743	0.226352513	0.634367593	0.618050267
11	137.5196182	138	97	182	305.7957445	0.024239229	2.165150735	0.239085942	2.39998112	0.905780155	0.188439689	0.836682729	0.908156271
11	124.5815678	123	85	171	302.3958891	0.157263903	2.337487616	0.270312193	2.399927465	0.897287271	0.205425458	0.799519759	2.247030799
11	136.2566278	135	97	190	235.6944815	0.239230221	2.898363625	0.302884105	2.27478463	0.897315381	0.205369238	0.762743054	2.130647037
11	103.9660297	104	84	127	48.34916449	0.021451756	2.963082862	0.455697222	1.504067175	0.900787285	0.198425431	0.487688295	0.185360391
11	119.4888889	118	91	150	113.1832099	0.313441476	2.821951596	0.398915552	1.741906278	0.902755763	0.194488475	0.598414484	0.025094313
11	93.99363057	94	69	115	52.55834584	-0.00212839	2.661414354	0.479310818	1.481001747	0.90375053	0.192498941	0.481421739	0.271968474
11	98.85608466	99	64	133	130.5888122	0.089912152	3.199445128	0.287976278	2.128945621	0.867828044	0.265015054	0.584546445	1.933669153
11	92.00743889	92	64	124	110.7108905	0.028504949	2.830486822	0.42152672	1.792197829	0.90406217	0.19187566	0.617656138	1.211553023
12	172.6582661	174	121	238	284.4245486	-0.150318476	3.345924827	0.237376471	2.542801993	0.863388143	0.273862692	0.733256006	4.869856604
12	145.6040404	145	95	218	327.9139231	0.35263109	3.632171052	0.224446897	2.718740697	0.850116567	0.306155434	0.73544318	0.706208951
12	209.9305136	210	158	253	189.3254838	0.024914066	3.449214922	0.344056432	2.19940645	0.881352518	0.242389287	0.65940833	0.806107672
12	241.421371	242	192	305	322.7881723	0.207262761	2.682515258	0.177062987	2.873767636	0.824632616	0.366666388	0.711702688	0.394711969

Cuadro B.1 continuación de la página previa

12	180.2454545	182	115	216	247.5084389	-0.767556345	3.831207984	0.286900445	2.441476345	0.853259253	0.299238423	0.661959767	0.790916338
12	211.1099899	212	170	248	186.9757932	-0.376378814	3.072402638	0.289091566	2.387624359	0.846635471	0.31184088	0.598937261	0.732230142
12	216.0322906	213	179	264	329.2280189	0.457479387	2.408760021	0.223332271	2.641062181	0.866243427	0.267513146	0.778212013	0.163329335
12	238.4117053	236	188	305	387.3602666	0.289743274	2.610930511	0.188685455	2.83370314	0.849477593	0.304242427	0.775108859	0.230672562
13	187.4627364	182	82	285	1998.960514	0.150413753	1.978386761	0.064093868	4.313519672	0.765985934	0.539955046	0.916716858	3.466730213
13	262.0812918	262	148	339	1046.054639	-0.228688728	2.804786393	0.095151937	3.820500932	0.78436648	0.476381918	0.858134289	1.492172664
13	216.7588889	218	133	296	514.3296432	-0.267126398	3.477874773	0.14687019	3.286620392	0.793318827	0.456293746	0.743144789	1.889920719
13	254.674833	257	128	323	600.8809034	-1.026350413	6.203711584	0.163224284	3.217944667	0.805518369	0.443903075	0.765672499	2.326978802
13	207.7060134	209	140	251	338.1652422	-0.395044499	3.066726854	0.195487256	2.835501287	0.820507018	0.377345297	0.676482025	2.562943068
13	250.2157953	250	190	317	531.003488	0.092988742	2.525945518	0.133067395	3.302585948	0.797524986	0.432410812	0.766897249	1.818903035
13	198.188196	200	122	238	391.0703729	-0.66716402	3.638609193	0.16980443	3.07395308	0.78860764	0.465823729	0.671950325	1.786583111
13	164.7022222	174	62	223	1040.851328	-1.093376158	3.552195563	0.13462395	3.562544657	0.796834613	0.44504053	0.87451378	2.601831441
14	168.9798387	170	118	198	247.8907226	-0.547223358	3.187850571	0.267458116	2.33180707	0.871488087	0.258331016	0.709005686	0.390464447
14	210.8319838	209	167	260	243.2167098	0.208866403	2.579746631	0.24872781	2.418901008	0.868430773	0.263138454	0.709957522	1.937263484
14	140.1814516	140	115	166	77.71707531	-0.091860516	2.897841813	0.576837385	1.39161138	0.917242039	0.165515923	0.474680628	0.706449117
14	173.0623742	171	147	210	166.2999324	0.199112224	2.188302236	0.363543643	1.831046959	0.922816856	0.154366288	0.741360066	2.715300546
14	129.8701826	130	108	152	64.2630663	-0.125542085	2.558240316	0.431408407	1.599296347	0.892564404	0.214871192	0.474496243	1.10296143
14	124.9979879	125	103	150	72.53721929	0.204931631	2.702925711	0.320828172	1.810501033	0.881659628	0.236680744	0.534507664	0.688011867
14	154.6370968	157	107	188	246.6424948	-0.598808206	2.961515295	0.319392368	2.272039728	0.894304223	0.211391554	0.766308784	1.414486707
14	118.6363636	119	94	139	61.60514233	-0.155536496	3.405010902	0.501133316	1.517116396	0.909218722	0.181562555	0.51633977	1.199323304
15	172.1967213	172	125	214	205.0104811	-0.034853728	2.876996327	0.237913015	2.485341712	0.826606464	0.372751555	0.524645447	21.7332078
15	207.9771242	207	119	256	381.5648362	-0.379219811	4.602706746	0.173902587	2.934685301	0.800111787	0.423481538	0.661441555	17.61544664
15	223.8585526	221	182	281	444.8648611	0.421378043	2.549962117	0.171753442	3.023258467	0.821081272	0.375168648	0.758001413	15.60585901
15	145.1348684	147	86	174	152.728521	-1.357918173	6.880268911	0.251886684	2.339763682	0.799404958	0.420695629	0.412170796	19.7348696
15	106.5307443	107	45	156	637.1552037	-0.202864856	2.015698202	0.14488143	3.32260957	0.797172619	0.505654762	0.779623883	3.5410221
15	169.7035831	170	140	202	131.5701175	-0.174548098	2.816883867	0.282651245	2.229932676	0.823262439	0.357731297	0.41616327	18.89238913
15	209.1331169	210	176	244	203.4140981	0.030160737	2.503354324	0.235039352	2.468415934	0.815937957	0.37881683	0.513772257	16.10810811
15	226.124183	226	151	273	437.6381734	-0.377248114	3.428239123	0.149137219	3.158467575	0.808260498	0.4199981	0.743359078	17.22038574
16	333.6807818	327	221	450	2094.243377	0.383212295	2.571962645	0.040494748	5.008907263	0.59042601	1.598197666	0.753764746	0.811476708
16	433.7864078	423	308	579	3497.883181	0.419709877	2.283375799	0.030857836	5.419406786	0.565428046	1.862053571	0.830787359	1.253736634
16	277.5915033	279	206	363	977.9017579	0.089030245	2.552535961	0.064380048	4.361232941	0.648544893	1.242558885	0.624415695	0.882929557

Cuadro B.1 continuación de la página previa

16	192.7745098	191	149	251	338.5798731	0.389522699	3.072796605	0.158298244	3.096809273	0.743612885	0.590139103	0.510698135	0.419227901
16	249.4788274	262	103	342	2363.102972	-0.740003291	2.951246769	0.047802721	4.816355646	0.639429878	1.254411802	0.831131314	0.919524853
16	197.8892508	199	136	246	342.5479952	-0.402098514	3.343027805	0.144083783	3.221487397	0.705881798	0.750635156	0.395914546	1.001807512
16	219.8078176	218	166	279	517.2953135	0.131596386	2.600014565	0.115481492	3.460558161	0.731528274	0.635439201	0.643484689	1.060903824
16	215.4318182	220	117	278	857.6349616	-0.788370098	3.914476606	0.092744465	3.986636403	0.666567096	1.035214937	0.652902092	0.10792304
17	296.9403595	299	181	370	718.0086979	-0.385898317	4.244201144	0.1353187	3.476004188	0.80555658	0.427877017	0.825206579	0.02238507
17	113.1013072	103	29	279	2055.133528	0.889271403	3.608287179	0.066634859	4.50787226	0.717653866	0.772379265	0.881661131	1.544968144
17	263.5584628	265	171	347	983.319354	-0.221389722	2.800489737	0.083550584	4.025902612	0.736523814	0.668158881	0.798616974	2.88081862
17	337.589198	339	82	459	2638.844335	-0.447054316	4.128948203	0.041935327	5.12789749	0.652679836	1.342021196	0.83938856	3.684573769
17	142.2248569	139	23	337	3233.821067	0.407349419	2.723358275	0.026342565	5.60189346	0.550348831	2.115377991	0.794853254	21.83973362
17	422.4403595	425	344	501	867.3624561	-0.246359458	2.630691162	0.115253976	3.577322603	0.808751001	0.409724896	0.855355881	0.091914805
17	267.0834015	271	133	329	778.7207629	-1.496898912	6.799949967	0.1589777	3.314356583	0.803885966	0.436409568	0.822129439	0.234165168
17	351.3079443	353	244	427	1019.580846	-0.392143015	3.115711883	0.118369622	3.600784962	0.827219805	0.357937064	0.889128605	1.841642367
18	185.5553936	186	149	219	232.5486808	-0.073785107	2.192550691	0.264634017	2.274014731	0.902681872	0.194636255	0.795054241	2.18369419
18	187.0562044	190	111	223	266.7785929	-2.052321139	8.181557088	0.45191666	1.989787435	0.90454063	0.19915694	0.790344015	2.178181372
18	173.6666667	176	111	194	126.473134	-1.1522842	5.053278376	0.337487979	1.914830215	0.905825769	0.188348462	0.710360388	3.648257676
18	176.7441691	178	130	203	158.1014602	-0.65982529	3.145095063	0.359991918	1.902722379	0.902687273	0.194625455	0.682052083	3.290520363
18	185.6946064	187	133	221	197.5838485	-0.391546886	3.152219536	0.286000363	2.372749732	0.867258782	0.267314496	0.679368362	3.775247559
18	188.3522976	195	100	232	559.4769076	-0.957185851	3.432613368	0.184439971	2.935781352	0.849104542	0.307278863	0.836099722	3.830088686
18	222.5353756	226	161	261	452.571141	-0.656699061	2.689306452	0.239395855	2.651472614	0.899560605	0.200878791	0.869776054	1.207880935
18	196.8942378	195	160	247	170.6693395	0.661100695	3.643176308	0.377731795	1.902897151	0.912362777	0.175274446	0.726548724	1.437676041
19	163.6742671	163	112	205	227.5746798	-0.177952596	3.507538974	0.250204637	2.611143927	0.798602198	0.454255759	0.455125473	0.144989339
19	170.3344262	171	133	208	131.9799624	-0.15285059	3.195601982	0.236632475	2.361126598	0.768573608	0.497292143	0.215765498	2.147956747
19	140.1954397	144	62	172	369.414572	-1.209548977	4.638730031	0.167314661	3.012777339	0.734798385	0.656896161	0.526316541	0.051304342
19	129.9934641	131	69	188	446.7450553	-0.252034971	3.267132383	0.140167501	3.404780154	0.733443362	0.702664664	0.557939488	2.997816089
19	245.8464052	244.5	153	330	927.1365394	-0.068691266	2.668474504	0.091212726	3.912153212	0.739181988	0.739252011	0.749131122	1.559538306
19	279.3496732	280	183	345	854.9659639	-0.325302778	3.158549319	0.09976121	3.791445449	0.723305079	0.718540936	0.716792069	3.793452515
19	201.0359477	202	158	242	215.9496882	-0.063381	2.715708562	0.226100142	2.556909869	0.8066749	0.429545886	0.461532646	1.881596944
19	181.7335526	186	93	247	707.971769	-0.61541209	3.111346165	0.097448488	3.818682213	0.683870697	0.889010883	0.612844137	1.464343516
20	535.930303	556	332	658	4439.878981	-1.051080327	3.280092498	0.043311267	5.233669501	0.620615011	1.486475267	0.895784506	3.120517001
20	534.7971746	543	361	645	1795.941707	-0.911880624	3.852719032	0.062390118	4.631331375	0.67656912	1.11901094	0.806427031	3.968159697

Cuadro B.1 continuación de la página previa

20	537.0514113	550	344	625	2963.087075	-0.753584389	2.877151183	0.056221602	4.808421885	0.695500899	0.991057101	0.895608474	5.43810883
20	536.7336024	553	241	631	3881.35991	-2.518982514	9.862352711	0.097683257	4.343177499	0.726564605	0.890089646	0.927757659	6.843842014
20	499.7749748	502	274	1094	6269.431705	2.368161553	19.68952917	0.046382958	5.240059694	0.608866028	2.933797178	0.846536301	6.362428249
20	534.8178138	535	455	600	555.3493993	-0.280218377	3.219450316	0.121362956	3.53819342	0.747086925	0.601009962	0.69389973	6.485606887
20	501.4727273	504	300	661	3489.845216	-0.497364876	3.367190552	0.038183225	5.313103361	0.653134188	1.269624825	0.888384339	4.663287853
20	515.6441532	515	409	608	1050.001397	0.151773017	2.807523919	0.073686089	4.187400233	0.698935889	0.814619965	0.761223969	5.396278381
21	90.60408163	90	57	135	235.3365431	0.241237512	2.49963183	0.321250078	2.151732799	0.927835332	0.144329336	0.842885458	23.69539719
21	90.93476995	88	61	131	173.661441	0.779342566	3.268734903	0.471525033	1.724245845	0.93553231	0.128935379	0.776982255	22.2192283
21	82.13702624	82	59	103	54.54623839	-0.065608501	2.791923314	0.635998032	1.118442979	0.949436955	0.101126091	0.639830853	25.82926255
21	79.69055944	79	47	106	72.40249829	0.197307613	3.509181899	0.448393182	1.617923598	0.928795616	0.142408768	0.702241214	23.04088661
21	110.4088526	113	70	148	272.4245693	-0.340931309	2.364815965	0.306470564	2.133059618	0.931698149	0.136603702	0.858486169	4.100670974
21	92.28688047	93	60	122	115.2844634	-0.167117578	2.701411227	0.412609565	1.779858767	0.916409717	0.167180566	0.689922417	23.46445803
21	89.997669	89	64	121	75.14801321	0.450863857	3.185781466	0.648411179	1.136979782	0.947224899	0.105550201	0.610803536	24.14780366
21	80.51515152	80	63	111	51.79639048	0.395183826	3.402863121	0.514861189	1.427379744	0.917164154	0.165671693	0.540410759	27.18805957
22	225.8438949	229	179	254	199.8937147	-1.105494884	3.956154067	0.374789044	2.01235775	0.909580328	0.180839344	0.787833169	0.692907518
22	212.5405405	215	116	252	309.5116769	-1.612314708	8.421299932	0.325523582	2.379903126	0.887276261	0.2399762	0.797056017	0.071853974
22	129.8462133	128	113	161	58.29396943	1.091370791	4.445645306	0.4962596	1.498640256	0.931145661	0.137708677	0.671015157	0.106821186
22	141.8230294	140	120	178	117.7978931	0.53445433	2.556759521	0.509162793	1.439819589	0.948246203	0.103507594	0.765689565	0.133975284
22	238.4876543	240	193	265	167.7297859	-0.789790944	3.617614561	0.395048878	1.975954263	0.91390739	0.17218522	0.744660916	0.538896777
22	249.3814672	257	162	281	443.6583438	-1.399868555	4.685647101	0.353549858	2.282881974	0.906630008	0.189657326	0.869400668	0.756727712
22	238.2223938	249	101	270	1015.857877	-2.359462581	8.437214824	0.243578205	2.768836698	0.862550531	0.294285577	0.912851296	0.505645849
22	221.2476852	227	139	260	462.5705965	-1.424104467	5.051840811	0.297768644	2.528493505	0.887251824	0.233736787	0.858387495	0.450373781
23	540.4133404	533	362	738	4756.900536	0.107660354	2.600472732	0.014042054	6.525553805	0.39417167	6.824962688	0.562289776	10.6956356
23	540.1902631	540	402	698	2474.322539	0.040816537	3.138745756	0.029553756	5.670180761	0.467694782	3.961873231	0.495471992	12.26542941
23	552.8705934	553	382	731	2700.724398	-0.043478207	3.321888524	0.024861047	5.847702459	0.446829257	4.527712393	0.488910012	12.22025877
23	564.853411	563	329	764	5113.13664	-0.135380475	3.222914028	0.014196772	6.625378972	0.362831775	8.074015178	0.515035295	6.094639509
23	510.9762401	514	270	738	5792.354292	-0.188762641	3.337155705	0.011694717	6.919633331	0.321356139	12.51012018	0.346103984	11.35428216
23	514.1405455	514	237	757	7425.640418	-0.100284466	3.109362858	0.009152788	7.191463822	0.283722095	15.13541523	0.365581652	9.062860291
23	425.9207609	421	204	711	6945.188929	0.320242133	3.305498326	0.010570902	7.088798855	0.292410603	16.20498585	0.262557536	18.40025166
23	510.2803555	514	184	777	7618.123393	-0.279309459	3.275360582	0.009419549	7.183635623	0.300342235	13.79925228	0.446021167	8.701043076
24	118.5594406	118	84	154	121.8455344	0.051712969	2.718079283	0.385885672	1.824572782	0.906749329	0.186865861	0.652366615	5.236430362

Cuadro B.1 continuación de la página previa

24	166.0944606	166	137	197	117.0278119	0.156739437	2.523719001	0.414410522	1.836689509	0.902154209	0.195691582	0.626235379	13.58638643
24	145.9260338	146	108	174	107.9857928	-0.171816601	3.298572493	0.361286308	1.826788205	0.905065147	0.189869706	0.653359257	11.53735862
24	156.2762238	156	107	219	425.7885023	0.170288952	2.762689153	0.176593836	2.961275619	0.844268514	0.314013063	0.794736382	13.69720635
24	154.0105018	155	98	202	245.1469142	-0.365069774	3.611820117	0.251567125	2.553002457	0.850448757	0.307833298	0.675422996	5.345091528
24	141.0367133	143	82	186	328.9293048	-0.42940086	2.829929665	0.203851017	2.741121898	0.855399521	0.291019183	0.77015745	7.298573193
24	180.7558275	182	121	217	208.9432935	-0.753889576	4.466214503	0.31350823	2.331491904	0.869078798	0.2629353	0.687937898	17.48756702
24	181.004662	182	132	210	134.273871	-0.625299019	3.954497528	0.416402732	1.86782339	0.902841736	0.194680827	0.655240918	13.09036661
25	149.4264877	150	111	179	95.7475131	-0.365143806	3.736876288	0.341542259	1.844146628	0.908974324	0.182051352	0.681581116	4.655361814
25	157.3389633	160	109	193	229.0965191	-0.538328702	2.862221633	0.338862095	2.159480137	0.913736001	0.172527997	0.794908371	0.508319965
25	145.3727432	146	120	173	97.45919882	-0.091548725	2.560434635	0.404317934	1.648622187	0.925213288	0.149573424	0.700581666	3.527632858
25	135.4186589	138	86	161	139.2917801	-1.268727414	5.324282646	0.580263099	1.484837019	0.941989625	0.116020751	0.757645915	4.153999282
25	155.4341492	157	112	191	167.8027732	-0.512757941	3.386306346	0.399274481	1.967439654	0.910359959	0.179280082	0.734068815	4.602169507
25	146.051312	146	113	179	109.4061134	0.107493693	3.057786224	0.364795446	1.843002734	0.897012055	0.20597589	0.61294804	2.947183938
25	172.9452214	176	95	204	260.200962	-2.111433165	8.56012368	0.316529902	2.184742162	0.892327154	0.216438588	0.802238719	4.06702504
25	174.5475219	175	133	207	112.0483248	-0.177198135	3.349668902	0.331590933	1.890015799	0.89798471	0.204030579	0.640998222	3.875018498
26	268.3022714	276	189	302	425.0157957	-1.335628396	4.433447899	0.308660602	2.311365669	0.924101338	0.152161402	0.904822133	1.589395314
26	218.1096849	221.5	126	294	2002.927292	-0.137989232	1.628270981	0.100728721	3.717036919	0.865539094	0.275118483	0.957042412	0.351823294
26	231.8145773	244	139	271	932.5183881	-1.190213503	3.519891776	0.193669454	2.995468723	0.884525113	0.232772374	0.929589147	1.429642678
26	219.8491555	224	141	247	321.7192378	-1.95132602	7.337212055	0.308020657	2.219898245	0.919660776	0.161042525	0.878035669	1.321981829
26	213.1060606	215	127	241	191.6122943	-2.240174927	11.96537893	0.485627821	1.791217561	0.924747118	0.151234804	0.79512806	1.55532546
26	232.7177843	240	121	268	643.9296837	-2.248866492	8.639785874	0.309783318	2.567627888	0.88483772	0.2328762	0.891828929	1.460097738
26	228.5560093	234	136	261	502.1476798	-1.569611808	5.859453451	0.293897821	2.546865821	0.906500637	0.187363689	0.892836564	1.50255985
26	233.5762515	245	131	272	828.2243953	-1.273284151	4.009541966	0.175459087	3.083621427	0.858429242	0.291146367	0.899802557	1.501774965

Cuadro B.1: Características radiómicas extraídas con la librería pyRadiomics de las 208 regiones de interés del hígado obtenido mediante imágenes de resonancia magnética.