TÍTULO TFG

Herramientas de IA en la Planificación Quirúrgica:

Segmentación y Clasificación Automatizada en

Imágenes Médicas RM

Alejandro Martínez Guillermo

Grado

Ingeniería Biomédica

E.T.S.I.I

UPCT

Martínez Cabeza de Vaca Alajarín, Juan de la Cruz

Zapata Pérez, Juan Francisco

Arévalo García, Alicia

2025

**Contenido**

[*Capítulo 1* 7](#_Toc192051938)

[**Introducción** 7](#_Toc192051939)

[**1.1.** **Evolución de la Inteligencia Artificial y su Aplicación en Medicina** 7](#_Toc192051940)

[**1.2.** **Principios de las Imágenes Médicas y la Resonancia Magnética** 8](#_Toc192051941)

[**1.2.1.** **Imágenes Médicas** 8](#_Toc192051942)

[**1.2.2.** **Principios de RM** 9](#_Toc192051943)

[**1.2.3.** **Ponderaciones en T1 y T2 en Resonancia Magnética** 11](#_Toc192051944)

[**Imágenes ponderadas en T1** 11](#_Toc192051945)

[**Imágenes ponderadas en T2** 12](#_Toc192051946)

[**1.2.4.** **Importancia de la clasificación de imágenes en T1 y T2** 13](#_Toc192051947)

[**1.3.** **Objetivos del Trabajo** 14](#_Toc192051948)

[*Capítulo 2* 16](#_Toc192051949)

[**Estado del arte** 16](#_Toc192051950)

[**2.1.** **Introducción** 16](#_Toc192051951)

[**2.3.** **Inteligencia Artificial en Medicina** 17](#_Toc192051952)

[**2.3.1.** **Evolución de la IA en Imágenes Médicas** 17](#_Toc192051953)

[**Introducción del Machine Learning en imágenes médicas** 17](#_Toc192051954)

[**Revolución del Deep Learning y su impacto en imágenes médicas** 17](#_Toc192051955)

[**2.3.2.** **Aplicaciones de la IA en Radiología y Resonancia Magnética** 18](#_Toc192051956)

[**Detección de tumores y patologías oncológicas** 18](#_Toc192051957)

[**2.4.** **Modelos de Segmentación de Imágenes Médicas** 18](#_Toc192051958)

[**2.4.1.** **Modelos Convencionales** 19](#_Toc192051959)

[**Métodos tradicionales de segmentación** 19](#_Toc192051960)

[**Limitaciones de los métodos tradicionales frente a Deep Learning** 19](#_Toc192051961)

[**2.4.2.** **Redes Neuronales Convolucionales (CNNs) en Segmentación Médica** 20](#_Toc192051962)

[**Explicación básica de las CNNs** 20](#_Toc192051963)

[**2.5.** **Modelos clave en Segmentación Médica** 21](#_Toc192051964)

[**2.5.1.** **U-Net: Arquitectura y aplicaciones en segmentación médica** 21](#_Toc192051965)

[**2.5.2.** **U-Net: Arquitectura y aplicaciones en segmentación médica** 22](#_Toc192051966)

[**2.5.3.** **TransDeepLab**: **Aplicación de Transformers en Segmentación Médica** 22](#_Toc192051967)

[**2.6.** **Modelos de Clasificación en Imágenes Médicas** 23](#_Toc192051968)

[**2.6.1.** **Modelos de Aprendizaje Profundo para Clasificación** 23](#_Toc192051969)

[**2.6.2.** **Redes CNN estándar y su uso en clasificación de imágenes médicas** 23](#_Toc192051970)

[**Modelos avanzados (ResNet, DenseNet, VGG) y su desempeño en RM** 23](#_Toc192051971)

[**Enfoques híbridos: uso combinado de segmentación y clasificación** 24](#_Toc192051972)

[**2.7.** **Empresas que utilizan la IA para Clasificación y Segmentación de Imagen Médica** 24](#_Toc192051973)

[**2.7.1.** **Avamed** 24](#_Toc192051974)

[**2.7.2.** **CELLA MS** 24](#_Toc192051975)

[*Capítulo 3* 25](#_Toc192051976)

[**Metodología** 25](#_Toc192051977)

[**3.1.** **Descripción general del proceso** 25](#_Toc192051978)

[**3.2.** **Fase 1: Clasificación de Imágenes RM en T1 y T2** 25](#_Toc192051979)

[**3.2.1.** **Búsqueda de información y selección del enfoque** 25](#_Toc192051980)

[**3.2.2.** **Primer entrenamiento** 26](#_Toc192051981)

[**3.2.2.1.** **División de conjunto de datos y configuración del entrenamiento** 26](#_Toc192051982)

[**3.2.2.2.** **Entrenamiento y Métricas** 28](#_Toc192051983)

[**3.2.2.3.** **Definición de Métricas de la Red Clasificadora** 29](#_Toc192051984)

[**3.2.3.** **Ampliación del conjunto de datos** 31](#_Toc192051985)

[**3.2.4.** **Segundo entrenamiento** 31](#_Toc192051986)

[**3.2.4.1.** **División de conjunto de datos y configuración del entrenamiento** 31](#_Toc192051987)

[**3.2.4.2.** **Entrenamiento y Métricas de validación** 32](#_Toc192051988)

[**3.2.5.** **Tercer entrenamiento** 35](#_Toc192051989)

[**3.2.5.1.** **División de conjunto de datos y configuración del entrenamiento** 35](#_Toc192051990)

[**3.2.5.2.** **Entrenamiento y Métricas de validación** 36](#_Toc192051991)

[**3.2.6.** **Separación del conjunto de datos en subtipos de T1 y T2** 39](#_Toc192051992)

[**3.2.7.** **Primer entrenamiento subclases T1 y T2** 41](#_Toc192051993)

[**3.2.7.1.** **División de conjunto de datos y configuración del entrenamiento** 41](#_Toc192051994)

[**3.2.7.2.** **Entrenamiento y Métricas de validación** 41](#_Toc192051995)

[**3.2.8.** **Prueba de otras redes para la clasificación de T1, T2 y subclases de T1** 45](#_Toc192051996)

[**3.3.** **Fase 2: Redes de Segmentación para T1, T2 y RM** 46](#_Toc192051997)

[**3.3.1.** **Preparación del conjunto de datos para la segmentación** 47](#_Toc192051998)

[**3.3.2.** **Selección de las estructuras anatómicas a segmentar** 47](#_Toc192051999)

[**3.3.3.** **Bloques de segmentación:** 49](#_Toc192052000)

[**3.3.3.1.** **Vasculaturas** 49](#_Toc192052001)

# *Capítulo 1*

# **Introducción**

* 1. **Evolución de la Inteligencia Artificial y su Aplicación en Medicina**

La inteligencia artificial (IA) ha estado experimentando un crecimiento exponencial en las últimas décadas, se ha consolidado como una de las tecnologías más importantes en muchos campos, de entre ellos, la medicina. Desde sus primeras menciones en la década de 1950 con los trabajos pioneros de Alan Turing en el concepto de máquinas inteligentes y John McCarthy, quien acuñó el término "inteligencia artificial" en 1956 [1], la IA ha evolucionado significativamente. Inicialmente la mayoría de los avances en IA se basaban en **sistemas expertos** y en programas con reglas fijas, es decir, sistemas que seguían instrucciones muy concretas, escritas directamente por personas expertas. Estos sistemas eran útiles para tareas específicas, pero eran rígidos y no podían adaptarse a situaciones nuevas., proporcionando soluciones limitadas y altamente dependientes del conocimiento humano explícito. Sin embargo, con la llegada del aprendizaje automático (Machine Learning) en las décadas de 1980 y 1990, se inició una transición hacia modelos capaces de aprender patrones a partir de grandes volúmenes de datos, mejorando su capacidad de generalización y adaptabilidad [2].

El avance más significativo se produjo con la introducción del aprendizaje profundo (Deep Learning), impulsado por tres factores clave:

1. El incremento en la capacidad de cómputo, especialmente con las unidades de procesamiento gráfico (GPUs), que han permitido entrenar modelos más complejos en tiempos reducidos [3].
2. La enorme cantidad de datos disponibles hoy en día, gracias a internet, redes sociales, sensores y todo tipo de dispositivos conectados, lo que ha dado a los modelos muchísimo material para aprender y mejorar [4].
3. Las constantes mejoras en los algoritmos y en las estructuras de las redes neuronales, que han hecho que estos modelos no solo sean más precisos, sino también más capaces de adaptarse a situaciones nuevas y resolver problemas cada vez más complejos [5].

En el campo de la medicina, la IA ha transformado por completo la forma en la que se aborda el diagnóstico, el tratamiento e incluso la investigación. Uno de los ámbitos donde más impacto ha tenido la IA es en el análisis de imágenes médicas , donde las redes neuronales convolucionales (CNNs) han demostrado ser muy eficaces. Este tipo de redes son capaces de analizar imágenes médicas de forma automática, además de realizar tareas de detección, clasificación y segmentación de estructuras anatómicas con niveles de exactitud y precisión similares a los de especialistas humanos en determinadas aplicaciones

[6].

Algunos modelos de aprendizaje automático han sido implementados en la identificación temprana de patologías como el cáncer, la retinopatía diabética y enfermedades neurodegenerativas, mejorando así las tasas de detección precoz. Además, la integración de la IA con sistemas de historial clínico permite proporcionar recomendaciones terapéuticas basadas en datos de pacientes, optimizando la toma de decisiones y reduciendo errores médicos. La capacidad de estos sistemas para analizar grandes volúmenes de información ha facilitado la identificación de patrones en datos genómicos y moleculares, impulsando avances en la medicina personalizada y el descubrimiento de nuevos fármacos.

La capacidad de los algoritmos para identificar patrones invisibles al ojo humano ha revolucionado la radiología y otras disciplinas médicas, permitiendo diagnósticos más precisos y la detección temprana de patologías [7]. Tecnologías como U-Net [8], nnU-Net [9] y Transformers en visión médica [10] han mejorado la segmentación automática de tejidos y órganos, mientras que otros modelos basados en redes profundas han permitido la clasificación automática de imágenes en modalidades como radiografías (RX), tomografías computarizadas (TC) y resonancias magnéticas (RM).

Además, la integración de la IA en el ámbito clínico ha abierto nuevas posibilidades en la planificación quirúrgica personalizada, permitiendo a los cirujanos acceder a reconstrucciones tridimensionales precisas de la anatomía del paciente. Estos avances han mejorado la seguridad y la eficiencia de los procedimientos quirúrgicos, reduciendo el margen de error y optimizando la preparación preoperatoria [11].

A pesar de los logros alcanzados, persisten desafíos importantes en la implementación clínica de la IA, incluyendo la necesidad de bases de datos de alta calidad, la validación regulatoria de los modelos y la interpretabilidad de los resultados [12]. El presente trabajo busca abordar estos desafíos mediante el desarrollo de una herramienta basada en IA para la clasificación y segmentación automatizada de imágenes de RM en pacientes oncológicos, con el fin de mejorar la precisión diagnóstica y optimizar la planificación quirúrgica.

* 1. **Principios de las Imágenes Médicas y la Resonancia Magnética**
     1. **Imágenes Médicas**

Las imágenes médicas desempeñan un papel fundamental en la práctica clínica y la investigación biomédica, permitiendo la evaluación no invasiva de la anatomía y fisiología del cuerpo humano. Existen diversas modalidades de adquisición de imágenes médicas, cada una con aplicaciones específicas según el tipo de tejido o patología a evaluar. Entre las principales técnicas se encuentran:

* Radiografía (RX): Basada en la atenuación diferencial de los rayos X, es ampliamente utilizada para el diagnóstico de fracturas óseas y patologías pulmonares [13].
* Tomografía Computarizada (TC): Emplea múltiples proyecciones de rayos X para generar imágenes seccionales del cuerpo, permitiendo una mejor caracterización de estructuras óseas y tejidos blandos [14].
* Ultrasonido (US): Utiliza ondas sonoras de alta frecuencia para obtener imágenes en tiempo real, siendo de gran utilidad en obstetricia, cardiología y evaluación de tejidos superficiales [15].
* Medicina Nuclear (PET y SPECT): Requiere la administración de radiofármacos que emiten radiación gamma, permitiendo el estudio de procesos metabólicos y la detección de enfermedades como el cáncer [16].
* Resonancia Magnética (RM): Basada en la interacción entre un campo magnético externo y los protones presentes en los tejidos, es una técnica avanzada que ofrece imágenes de alta resolución sin el uso de radiación ionizante [17].

Entre estas modalidades, la **Resonancia Magnética (RM)** ha demostrado ser especialmente útil en la evaluación de tejidos blandos, permitiendo la identificación de estructuras anatómicas con un alto nivel de detalle [18]. A diferencia de otras técnicas de imagen como la tomografía computarizada (TC) o las radiografías, la RM **no emplea radiación ionizante**, lo que la convierte en una opción más segura para el seguimiento a largo plazo de pacientes, especialmente aquellos con enfermedades crónicas o en poblaciones vulnerables como niños y embarazadas [19]. Su capacidad para generar distintos contrastes mediante variaciones en los parámetros de adquisición permite resaltar diferencias sutiles entre tejidos sanos y patológicos, lo que resulta fundamental en el diagnóstico y seguimiento de enfermedades neurológicas, musculoesqueléticas y oncológicas. En el ámbito de la oncología, la RM juega un papel clave en la detección y caracterización de tumores, facilitando la evaluación de su extensión y respuesta al tratamiento sin exponer al paciente a dosis acumulativas de radiación. Además, su aplicación en técnicas avanzadas como la RM funcional y la espectroscopía por RM ha permitido el análisis de la actividad metabólica de los tejidos, proporcionando información valiosa para la planificación quirúrgica y la personalización de terapias en pacientes oncológicos.

* + 1. **Principios de RM**

La RM se basa en la propiedad física del spin nuclear, un fenómeno cuántico asociado a los protones de los átomos de hidrógeno. Al colocar un paciente dentro de un campo magnético estático (B₀) generado por el equipo de RM, los protones de hidrógeno se alinean en paralelo o antiparalelo a dicho campo, creando una magnetización neta en la dirección del campo externo [20].

Interfaz de usuario gráfica

Descripción generada automáticamente

**Figura 1**. https://radiodiagnosticando.com/2016/06/13/rm-introduccion-a-conceptos-basicos-definicion-de-spin/ 

A continuación, se emite un pulso de radiofrecuencia (RF) con una frecuencia específica (frecuencia de Larmor), lo que provoca que los protones absorban energía y cambien su alineación. Cuando cesa la excitación, los protones regresan a su estado de equilibrio, emitiendo señales electromagnéticas que son captadas por las antenas del equipo de RM [21].

donde:

* es la frecuencia de Larmor (en Hz).
* es la razón giromagnética del protón ( 42.58 MHz/T para el hidrógeno)
* es la intensidad del campo magnético (en Tesla)

El retorno a la alineación original se describe mediante dos tiempos de relajación:

* Tiempo de Relajación Longitudinal (T1): Representa el tiempo que tardan los protones en recuperar su alineación con el campo magnético externo.
* Tiempo de Relajación Transversal (T2): Describe la pérdida de coherencia entre los protones debido a interacciones entre ellos.

Estos parámetros determinan el contraste de la imagen y permiten la diferenciación entre distintos tipos de tejidos [22].

Diagrama

Descripción generada automáticamente

**Figura 2**. Recovery of longitudinal magnetization (T1) and decay of transverse magnetization (T2). Source: [ResearchGate](https://www.researchgate.net/figure/Relaxation-rates-of-longitudinal-magnetization-1-T1-and-transverse-magnetization-1-T2_fig1_281727332). 

* + 1. **Ponderaciones en T1 y T2 en Resonancia Magnética**

La manipulación de los tiempos de excitación y detección permite obtener imágenes con distintos contrastes, siendo las secuencias ponderadas en T1 y T2 las más utilizadas en el análisis de tejidos blandos [23].

En una adquisición de Resonancia Magnética (RM), dos parámetros clave determinan el contraste de la imagen:

* Tiempo de repetición (TR), es el intervalo de tiempo entre pulsos sucesivos de radiofrecuencia que excitan los protones en el tejido. Un TR corto favorece la recuperación de la magnetización longitudinal, lo que resalta estructuras con tiempos de relajación T1 cortos, como la grasa.
* Tiempo de eco (TE), es el tiempo transcurrido entre la aplicación del pulso de radiofrecuencia y la adquisición de la señal emitida por los protones. Un TE corto minimiza la atenuación de la señal y destaca la intensidad de tejidos con relajación T1 corta, mientras que un TE largo potencia la señal de tejidos con relajación T2 prolongada, como los líquidos.

**Imágenes ponderadas en T1**

Las imágenes ponderadas en T1 se caracterizan por un tiempo de repetición (TR) corto y un tiempo de eco (TE) corto, lo que favorece la recuperación de la magnetización longitudinal y la atenuación de la señal proveniente de tejidos con tiempos de relajación T1 prolongados [24].

* Tejidos con T1 corto (grasa, médula ósea amarilla, hígado): Presentan una señal intensa, apareciendo brillantes en la imagen.
* Tejidos con T1 largo (líquidos como la bilis en la vesícula biliar, el contenido intestinal, el líquido pleural o peritoneal en casos patológicos, médula espinal): Generan una señal débil, apareciendo oscuros.

Imagen que contiene interior, objeto, tabla, bolsa

Descripción generada automáticamente

**Figura 3**. Imagen Abdominal RM ponderada en T1 (poner la información anterior a modo resumen en el pie de página/pie de foto)

**Imágenes ponderadas en T2**

Las imágenes ponderadas en T2 emplean un TR largo y un TE largo, lo que permite la detección de tejidos con tiempos de relajación T2 prolongados [25].

* Tejidos con T2 largo (líquidos como el contenido de la vejiga, bilis en la vesícula biliar, inflamaciones, edemas, médula espinal): Se visualizan con una señal intensa, apareciendo brillantes en la imagen.
* Tejidos con T2 corto (grasa, médula ósea amarilla, hígado, bazo, diafragma): Presentan menor intensidad de señal, apareciendo oscuros.

Imagen que contiene zapatos, pequeño, par, perro

Descripción generada automáticamente

***Figura 3****. Imagen Abdominal RM ponderada en T2*

La diferenciación entre imágenes ponderadas en T1 y T2 es esencial para el diagnóstico y planificación clínica, ya que cada modalidad proporciona información complementaria sobre las características tisulares [26].

* + 1. **Importancia de la clasificación de imágenes en T1 y T2**

La correcta clasificación de imágenes T1 y T2 es un desafío en la práctica clínica, dado que la variabilidad en los protocolos de adquisición puede generar contrastes similares entre ambas modalidades [27]. Tradicionalmente, esta clasificación ha sido realizada de manera manual por especialistas en radiología, lo que implica un proceso sujeto a variabilidad del observador y consumo de tiempo [28].

Uno de los principales retos en la segmentación de órganos con imágenes de resonancia magnética es la diferenciación entre imágenes ponderadas en T1 y T2, ya que los modelos de segmentación no siempre responden de la misma manera a ambas modalidades. Esta distinción es crucial porque:

* El contraste entre tejidos varía significativamente entre imágenes T1 y T2, pudiendo afectar el rendimiento de los algoritmos de segmentación.
* El mismo modelo de segmentación podría no ser aplicado indistintamente a ambas ponderaciones sin pérdida de precisión.
* En la práctica clínica, los radiólogos eligen distintas secuencias para diferentes patologías, por lo que un sistema automatizado debe ser capaz de adaptar su segmentación en función del tipo de imagen.

Dado que muchas bases de datos de imágenes médicas no siempre incluyen una etiqueta clara sobre si una imagen es T1 o T2, surgió la necesidad de desarrollar un modelo de clasificación automática. Este modelo permitiría:

1. Facilitar la organización de imágenes en proyectos de segmentación, asegurando que cada modelo trabaje sobre imágenes homogéneas.
2. Evitar posibles errores de segmentación causados por aplicar modelos entrenados en un tipo de imagen sobre otro.
3. Automatizar el preprocesamiento de imágenes médicas, reduciendo la intervención manual y mejorando la eficiencia de los flujos de trabajo en empresas de segmentación de estructuras anatómicas como por ejemplo CELLA MS.

En este contexto, el desarrollo de herramientas basadas en inteligencia artificial para la clasificación automática de imágenes T1 y T2 representa una solución innovadora con el potencial de optimizar los flujos de trabajo en radiología. Al aplicar modelos de aprendizaje profundo, es posible diferenciar ambas modalidades con alta precisión, permitiendo la segmentación automática de estructuras anatómicas de manera más eficiente y personalizada [29].

La integración de un sistema de clasificación automático previo a la segmentación podría suponer una mejora en la precisión del modelo, evitando errores que podrían surgir al aplicar un mismo algoritmo de segmentación sin considerar las diferencias en contraste entre imágenes ponderadas en T1 y T2. Este enfoque representa un avance en la planificación quirúrgica asistida por IA, proporcionando herramientas más robustas para la toma de decisiones clínicas [30].

* 1. **Objetivos del Trabajo**

El objetivo de este Trabajo de Fin de Grado es desarrollar una herramienta basada en inteligencia artificial para la clasificación y segmentación automatizada de imágenes de resonancia magnética en planificación quirúrgica.

Este proyecto busca contribuir al desarrollo de herramientas basadas en IA para mejorar la **planificación quirúrgica en oncología**, utilizando imágenes médicas de alta resolución para asistir en la toma de decisiones clínicas. Para lograrlo, se han definido los siguientes subobjetivos:

* Crear una base de datos de imágenes médicas que incluya resonancia magnética (RM) de pacientes oncológicos, garantizando la diversidad y calidad de las muestras utilizadas para el entrenamiento de los modelos.
* Diseñar y entrenar modelos de clasificación automáticos utilizando redes neuronales convolucionales (CNN) capaces de distinguir entre diferentes modalidades de imagen médica.
* Implementar modelos de segmentación automática, optimizando la precisión en la identificación de estructuras anatómicas y facilitando la planificación quirúrgica asistida por IA.
* Validar estadísticamente los modelos desarrollados para asegurar su viabilidad clínica mediante métricas de rendimiento y comparación con estándares médicos.
* Integrar los modelos de clasificación y segmentación en un planificador quirúrgico web, proporcionando una interfaz accesible para los profesionales de la salud.
* Explorar el potencial de las tecnologías de realidad mixta en la visualización y análisis de estructuras anatómicas para mejorar la toma de decisiones clínicas.
* Evaluar y analizar el impacto clínico y tecnológico del proyecto, asegurando que las soluciones desarrolladas aporten un beneficio real en entornos hospitalarios.

# *Capítulo 2*

# **Estado del arte**

* 1. **Introducción**

A continuación, se va a realizar un repaso del estado del arte de la inteligencia artificial y su aplicación en la Medicina, desde modelos de segmentación a modelos de clasificación.

* 1. **Primeros Enfoques: Métodos tradicionales de segmentación**

Los primeros enfoques de segmentación de imágenes médicas se basaban en técnicas matemáticas y estadísticas sin la intervención de modelos de aprendizaje automático. Algunos de los métodos más representativos fueron:

* Umbralización (Thresholding): Técnica basada en la separación de regiones de interés mediante la definición de un valor de intensidad umbral. Aunque es eficiente y computacionalmente simple, su principal limitación es su sensibilidad a la iluminación y variaciones de contraste [33].
* Crecimiento de regiones (Region Growing): Algoritmo que expande regiones a partir de píxeles semilla con características similares. Este método mejora la segmentación respecto a la umbralización, pero es susceptible al ruido y la selección del punto de inicio [34].
* Contornos activos (Active Contours): Modelos basados en energía que ajustan una curva a los bordes de un objeto en la imagen. Estos métodos son efectivos para segmentaciones suaves y continuas, pero pueden requerir ajustes manuales significativos [35].

Aunque estas técnicas fueron ampliamente utilizadas, presentaban limitaciones en la segmentación de estructuras complejas y en su capacidad de generalización para diferentes modalidades de imagen médica.

Imagen que contiene foto, cámara, hombre, posando

Descripción generada automáticamente

Apuntes PIM (Thresholding)

* 1. **Inteligencia Artificial en Medicina**
     1. **Evolución de la IA en Imágenes Médicas**

El procesamiento de imágenes médicas ha evolucionado significativamente con la incorporación de técnicas de inteligencia artificial. Desde los primeros métodos basados en algoritmos tradicionales hasta el auge del aprendizaje profundo (Deep Learning), la segmentación y clasificación de imágenes han experimentado avances sustanciales en precisión y automatización.

**Introducción del Machine Learning en imágenes médicas**

Con el auge del aprendizaje automático (Machine Learning, ML), se comenzaron a desarrollar enfoques más robustos para la segmentación y clasificación de imágenes médicas. Algunos de los algoritmos más utilizados fueron:

* Support Vector Machines: Técnica basada en la búsqueda de un hiperplano óptimo para clasificar datos. Se utilizó en la clasificación de tejidos y detección de tumores, pero su rendimiento dependía en gran medida de la selección de características adecuadas [36].
* Random Forests: Algoritmo basado en la combinación de múltiples árboles de decisión, utilizado en la detección de patrones en imágenes médicas. Si bien ofrece buenos resultados, su desempeño se ve afectado por la calidad de los atributos extraídos manualmente [37].

Estos enfoques mejoraron la segmentación y clasificación en comparación con los métodos tradicionales, pero aún requerían la extracción manual de características y dependían de la parametrización adecuada para cada tipo de imagen.

**Revolución del Deep Learning y su impacto en imágenes médicas**

La llegada del aprendizaje profundo (Deep Learning, DL) supuso un cambio radical en la segmentación y clasificación de imágenes médicas. A diferencia de los métodos anteriores, las redes neuronales profundas pueden aprender automáticamente las características más relevantes sin necesidad de ingeniería manual de características [38].

* Redes Neuronales Convolucionales (CNNs): Algoritmos capaces de extraer características jerárquicas de las imágenes, permitiendo una clasificación y segmentación más precisa [39].
* Modelos U-Net y variantes: Arquitecturas especializadas en la segmentación de imágenes biomédicas, como U-Net y nnU-Net, que han demostrado un rendimiento superior en segmentación automática de tejidos y órganos [40].
* Transformers en visión médica: Técnicas más recientes basadas en autoatención, como los Vision Transformers (ViTs), que han mejorado la segmentación en imágenes médicas con menos dependencia de grandes volúmenes de datos etiquetados [41].

El impacto del Deep Learning ha sido tan significativo que ha permitido el desarrollo de modelos autoajustables, reduciendo la necesidad de intervención manual y mejorando la reproducibilidad de los resultados en distintos centros hospitalarios.

* + 1. **Aplicaciones de la IA en Radiología y Resonancia Magnética**

La inteligencia artificial ha revolucionado el campo de la radiología y la resonancia magnética (RM), permitiendo la automatización de tareas complejas que anteriormente requerían intervención manual. Sus aplicaciones abarcan desde la detección temprana de patologías hasta la planificación quirúrgica avanzada, facilitando el análisis de grandes volúmenes de datos con alta precisión y reduciendo la variabilidad del observador [42].

**Detección de tumores y patologías oncológicas**

Los modelos basados en redes neuronales convolucionales (CNNs) han demostrado un gran potencial en la identificación de tumores en diversas modalidades de imágenes médicas, incluidas las imágenes de resonancia magnética. Algunos ejemplos incluyen:

* **Cáncer de hígado**: Modelos como **nnU-Net y DeepLabV3+** han sido utilizados en la **segmentación de hepatocarcinomas (HCC) y metástasis hepáticas** en imágenes de **RM ponderadas en T1 y T2**, permitiendo una identificación más precisa de los márgenes tumorales y facilitando la planificación de cirugías y terapias dirigidas [44].
* **Cáncer de páncreas**: Algoritmos de IA han sido aplicados en la **detección y clasificación de tumores pancreáticos** en imágenes de RM, diferenciando entre **adenocarcinomas ductales pancreáticos (¿explicar?) y lesiones benignas** con alta precisión. Estudios recientes han demostrado que estos modelos mejoran la identificación de tumores en etapas tempranas, donde el diagnóstico suele ser más desafiante debido a la heterogeneidad del tejido pancreático [45].
* **Cáncer de pulmón (metástasis y tumores mediastínicos)**: La integración de CNNs con **transformers en visión médica** ha permitido el desarrollo de modelos híbridos capaces de segmentar masas tumorales en **RM torácicas**, especialmente en casos de metástasis pulmonares derivadas de tumores primarios en hígado o páncreas. Estos modelos han mostrado gran eficacia en la identificación de **patrones de diseminación tumoral** [46].
* **Cáncer colorrectal (metástasis hepáticas y peritoneales)**: Modelos basados en IA han sido aplicados en imágenes de **RM con contraste dinámico**, mejorando la segmentación de lesiones metastásicas en el peritoneo y el hígado, lo que facilita la evaluación de la **carga tumoral y la planificación de cirugías oncológicas** [47].
  1. **Modelos de Segmentación de Imágenes Médicas**

La segmentación de imágenes médicas es un proceso fundamental en la interpretación y análisis de imágenes biomédicas, facilitando la identificación de estructuras anatómicas y patológicas. A lo largo del tiempo, se han desarrollado diversos enfoques, desde métodos tradicionales basados en reglas matemáticas hasta modelos avanzados basados en aprendizaje profundo (Deep Learning).

* + 1. **Modelos Convencionales**

Los métodos tradicionales de segmentación han sido utilizados ampliamente en el análisis de imágenes médicas, empleando técnicas matemáticas y estadística para identificar regiones homogéneas dentro de una imagen.

**Métodos tradicionales de segmentación**

* Segmentación basada en umbralización: Separa regiones de una imagen en función de un valor umbral, lo que la hace eficaz en imágenes con contraste bien definido. Sin embargo, presenta dificultades en la segmentación de estructuras complejas con variabilidad en la intensidad [49].
* Método Watershed: Basado en el concepto de cuencas hidrográficas, este método es útil para segmentar imágenes con bordes bien definidos, aunque es propenso a la sobresegmentación en imágenes con ruido o estructuras anatómicas adyacentes [50].
* Clustering K-Means: Técnica de agrupamiento no supervisado que divide los píxeles en diferentes regiones en función de su similitud en intensidad. Sin embargo, su rendimiento depende de la selección del número de clases y es sensible a variaciones en contraste [51].
* Modelos de contornos activos (Active Contours o Snakes): Utilizan curvas que se ajustan a los bordes de los objetos en la imagen. Aunque son eficaces para la segmentación de estructuras anatómicas con límites suaves, requieren inicialización manual y pueden verse afectados por ruido en la imagen [52].

**Limitaciones de los métodos tradicionales frente a Deep Learning**

Los enfoques convencionales presentan varias limitaciones que han impulsado la adopción de técnicas basadas en aprendizaje profundo, especialmente en el ámbito de la oncología torácico-abdominal:

* Dependencia de parámetros manuales: Requieren el ajuste manual de umbrales y semillas iniciales, lo que incrementa la variabilidad del observador y la posibilidad de errores en la segmentación de tejidos y órganos.
* Baja adaptabilidad: Son altamente sensibles a variaciones en el contraste, artefactosy ruido presentes en las imágenes de resonancia magnética, lo que limita su aplicabilidad en estudios clínicos.
* Capacidad limitada de generalización: No se adaptan bien a conjuntos de datos diversos sin ajustes específicos, lo que dificulta su implementación en entornos clínicos con pacientes de distintas características.
* Procesos manuales lentos y demandantes: La segmentación convencional de tumores requiere la intervención de radiólogos expertos, lo que puede ralentizar el diagnóstico y la planificación del tratamiento, especialmente en hospitales con alta carga asistencial.

El desarrollo de redes neuronales convolucionales (CNNs) ha permitido abordar estas limitaciones mediante la automatización del proceso de segmentación, reduciendo significativamente el tiempo necesario para analizar grandes volúmenes de datos. Modelos avanzados como nnU-Net y DeepLabV3+ han demostrado una mayor precisión y robustez, permitiendo la identificación de estructuras tumorales con velocidades muy superiores a las de los métodos tradicionales, lo que agiliza el flujo de trabajo clínico y facilita la toma de decisiones en oncología.

* + 1. **Redes Neuronales Convolucionales (CNNs) en Segmentación Médica**

Las redes neuronales convolucionales (CNNs) han revolucionado la segmentación de imágenes médicas al permitir el aprendizaje automático de características relevantes directamente desde los datos de entrada, sin necesidad de una extracción manual de características.

**Explicación básica de las CNNs**

Las CNNs se componen de múltiples capas de procesamiento, incluyendo:

* Capas de convolución: Aplican filtros sobre la imagen para extraer características como bordes y texturas.
* Capas de pooling: Reducen la dimensionalidad de la imagen manteniendo las características más relevantes.
* Capas totalmente conectadas: Integran la información extraída y generan la predicción final de segmentación [53].

Diagrama

Descripción generada automáticamente

**Figura 4.** <https://es.mathworks.com/discovery/convolutional-neural-network.html>

* 1. **Modelos clave en Segmentación Médica**
     1. **U-Net: Arquitectura y aplicaciones en segmentación médica**

El modelo U-Net es una arquitectura de red neuronal convolucional desarrollada en 2015 por Olaf Ronneberger y su equipo, diseñada específicamente para la segmentación de imágenes biomédicas. Su estructura en forma de "U" consta de dos caminos principales:

1. Camino de Contracción (Encoder): Este segmento captura el contexto de la imagen mediante la aplicación repetida de convoluciones seguidas de funciones de activación y operaciones de pooling. A medida que se avanza en este camino, la resolución espacial disminuye mientras que la profundidad de las características aumenta, permitiendo una representación más abstracta de la información.
2. Camino de Expansión (Decoder): En esta sección, la resolución de la imagen se restaura progresivamente. Se emplean operaciones de upsampling y convoluciones para aumentar la resolución, incorporando conexiones de salto que transfieren información detallada desde las capas correspondientes del encoder. Estas conexiones son cruciales para recuperar detalles espaciales que podrían haberse perdido durante la fase de contracción.

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente  
**Figura 5.** Arquitectura de la UNet --> <https://github.com/zhixuhao/unet>

El U-Net ha sido ampliamente utilizado en la segmentación de imágenes médicas, logrando resultados de vanguardia en tareas como la detección de tumores, órganos y anomalías en resonancia magnética (RM), tomografía computarizada (TC) y ecografías. Sin embargo, su implementación puede requerir ajustes manuales significativos para adaptarse a diferentes conjuntos de datos y tareas específicas, lo que ha llevado al desarrollo de enfoques más automatizados, como nnU-Net.

* + 1. **U-Net: Arquitectura y aplicaciones en segmentación médica**

Para abordar las limitaciones de U-Net, se desarrolló la nnU-Net ("no-new-Net"), un marco autoajustable que automatiza la configuración de la arquitectura U-Net según las características del conjunto de datos proporcionado.

A diferencia de U-Net, la nnU-Net analiza los datos de entrenamiento y configura automáticamente una tubería de segmentación, eliminando la necesidad de intervención manual en:

* Preprocesamiento: Normalización y transformación de imágenes para mejorar el rendimiento del modelo.
* Configuración del modelo**:** Ajuste automático de hiperparámetros, tamaño del kernel y profundidad de la red.
* Entrenamiento: Optimización de la arquitectura para obtener mejores resultados sin necesidad de prueba y error manual.
* Posprocesamiento**:** Refinamiento de las segmentaciones generadas para mejorar su precisión clínica.

La nnU-Net ha demostrado rendimiento de vanguardia en múltiples desafíos de segmentación médica, destacándose en competiciones como el *Medical Segmentation Decathlon*. Su capacidad para adaptarse dinámicamente a diversas tareas lo ha convertido en una herramienta esencial en la segmentación de tumores hepáticos, pancreáticos y colorrectales en imágenes de resonancia magnética y tomografía computarizada.

* + 1. **TransDeepLab**: **Aplicación de Transformers en Segmentación Médica**

A pesar del éxito de U-Net y nnU-Net, han surgido nuevas arquitecturas basadas en Transformers, como TransDeepLab, que buscan mejorar la segmentación médica aprovechando los mecanismos de autoatención. Estas arquitecturas combinan las ventajas de las CNNs y Transformers, permitiendo una captura más eficiente de dependencias espaciales a largo alcance, lo que puede ser especialmente útil en la segmentación de estructuras anatómicas irregulares.

Entre sus beneficios, TransDeepLab ha mostrado:

* Mejor captura de relaciones espaciales, favoreciendo la segmentación de tejidos complejos como tumores pancreáticos, hepáticos y metástasis torácico-abdominales.
* Mayor capacidad de generalización, al ser menos sensible a variaciones de contraste y ruido en imágenes médicas.
* Potencial para segmentaciones más detalladas en imágenes de resonancia magnética y tomografía computarizada con contraste.

Sin embargo, pese a estos avances, U-Net y nnU-Net siguen siendo las arquitecturas de referencia en la segmentación de imágenes médicas. Su capacidad para obtener resultados robustos en múltiples aplicaciones, junto con su facilidad de entrenamiento y optimización, hace que continúen siendo el estándar en la mayoría de los estudios y aplicaciones clínicas. Aunque modelos como TransDeepLab presentan mejoras teóricas, su adopción aún es limitada debido a la complejidad computacional y la falta de estudios clínicos extensivos que validen su desempeño en diferentes escenarios médicos.

Por tanto, aunque los Transformers en visión médica representan un área prometedora, en la actualidad U-Net y nnU-Net siguen dominando el campo de la segmentación de imágenes médicas, especialmente en tareas oncológicas donde la precisión y confiabilidad del modelo son críticas para la planificación quirúrgica y el tratamiento

* 1. **Modelos de Clasificación en Imágenes Médicas**
     1. **Modelos de Aprendizaje Profundo para Clasificación**

El análisis y clasificación de imágenes médicas han evolucionado considerablemente con el uso del aprendizaje profundo, en particular con las redes neuronales convolucionales (CNN). Estos modelos han demostrado una gran capacidad para extraer patrones complejos y diferenciar entre diferentes tipos de imágenes médicas, como resonancias magnéticas (RM), tomografías computarizadas (TC) y radiografías.

* + 1. **Redes CNN estándar y su uso en clasificación de imágenes médicas**

Las CNN estándar están compuestas por múltiples capas de convolución, seguidas de capas de pooling y capas totalmente conectadas. En la clasificación de imágenes médicas, estas redes se han utilizado para tareas como la detección de tumores, el diagnóstico de enfermedades pulmonares y la identificación de anomalías en imágenes de resonancia magnética. Entre las arquitecturas tradicionales se encuentran AlexNet y LeNet, las cuales fueron las primeras en demostrar el potencial del aprendizaje profundo en la clasificación de imágenes.

**Modelos avanzados (ResNet, DenseNet, VGG) y su desempeño en RM**

A medida que la complejidad de las imágenes médicas ha aumentado, se han desarrollado modelos más sofisticados como:

* ResNet (Redes Residuales): Introducen conexiones residuales que permiten entrenar redes más profundas sin sufrir el problema de la desaparición del gradiente. En imágenes de RM, han mejorado la precisión en la detección de anomalías estructurales.
* DenseNet: Conexiones densas entre capas que facilitan la reutilización de características y mejoran la eficiencia del entrenamiento. Se han aplicado con éxito en la segmentación y clasificación de tejidos en imágenes médicas.
* VGG (Visual Geometry Group): Caracterizada por su simplicidad y uso de múltiples capas convolucionales de pequeño tamaño, ha sido ampliamente utilizada en la clasificación de resonancias magnéticas cerebrales.

**Enfoques híbridos: uso combinado de segmentación y clasificación**

En algunos casos, la clasificación de imágenes médicas se beneficia del uso combinado de segmentación y clasificación. Estos modelos híbridos primero segmentan las regiones de interés en la imagen (por ejemplo, una lesión o un tumor) y luego clasifican la región segmentada para mejorar la precisión del diagnóstico. Este enfoque es particularmente útil en imágenes de resonancia magnética y tomografía computarizada, donde las estructuras anatómicas pueden ser difíciles de diferenciar.

Estado del arte, relacionados con papers que ya he leído, acerca del estado actual (Todo lo que llevo es introducción)

* 1. **Empresas que utilizan la IA para Clasificación y Segmentación de Imagen Médica**
     1. **Avamed**
     2. **CELLA MS**

Seguir con este punto, y tengo dudas de cómo abordar este punto

# *Capítulo 3*

# **Metodología**

* 1. **Descripción general del proceso**

El presente trabajo de fin de grado se desarrolla en torno al diseño e implementación de un flujo de trabajo basado en inteligencia artificial para el análisis automatizado de imágenes de resonancia magnética (RM), con especial foco en la clasificación y la segmentación de estructuras anatómicas relevantes para la planificación quirúrgica.

Dado que las imágenes de RM pueden adquirirse con distintas secuencias de ponderación (principalmente T1 y T2), y que cada tipo de secuencia proporciona un contraste diferente entre los tejidos, se planteó la necesidad de incluir un primer paso de clasificación automática. Este paso permite identificar si cada imagen corresponde a una secuencia T1 o T2, optimizando así el rendimiento de los modelos de segmentación que se aplicarán posteriormente.

El proceso completo se divide, por tanto, en dos fases principales:

* Fase 1: Clasificación automática de imágenes en T1 y T2.  
  En esta fase se entrena un modelo de clasificación basado en redes neuronales convolucionales (CNN) capaz de diferenciar entre imágenes ponderadas en T1 e imágenes ponderadas en T2. Este modelo de clasificación permite adaptar el flujo de trabajo de segmentación en función de la secuencia de adquisición.
* Fase 2: Segmentación automática de estructuras anatómicas.  
  Una vez clasificadas las imágenes, se emplean modelos específicos de segmentación automática, optimizados para cada tipo de imagen (T1 o T2), con el objetivo de identificar estructuras anatómicas de interés de manera precisa y robusta.

Ambas fases se integran dentro de un pipeline automatizado, pensado para su futura aplicación en entornos clínicos, donde se busca facilitar la planificación quirúrgica mediante el uso de herramientas basadas en inteligencia artificial.

* 1. **Fase 1: Clasificación de Imágenes RM en T1 y T2**
     1. **Búsqueda de información y selección del enfoque**

Antes de desarrollar el modelo de clasificación de imágenes de resonancia magnética en T1 y T2, se llevó a cabo una revisión de enfoques previos utilizados en la literatura y en implementaciones existentes. La clasificación de secuencias de RM para imágenes cerebrales ha sido abordada en múltiples estudios mediante técnicas basadas en aprendizaje profundo, principalmente redes neuronales convolucionales (CNN), dada su capacidad para extraer características espaciales y texturales relevantes en imágenes médicas.

Se analizaron distintos métodos, incluyendo modelos preentrenados y arquitecturas personalizadas, con el objetivo de encontrar una solución eficiente y adaptable a las imágenes disponibles. Tras evaluar varias alternativas, se decidió utilizar como base el código disponible en el repositorio de GitHub [Jpvmello/type-identification-mri-sequences](https://github.com/Jpvmello/type-identification-mri-sequences), este estudio clasificaba tipos de imagen de RM cerebrales, por lo que se planteó su posibilidad y utilidad para clasificar tipos de imagen de RM tórax, abdomen y pelvis.

Este repositorio implementa un modelo de clasificación basado en la arquitectura ResNet18, una red neuronal convolucional residual optimizada para la clasificación de imágenes. La elección de este modelo se fundamentó en su equilibrio entre profundidad y eficiencia computacional, lo que lo hace adecuado para trabajar con conjuntos de datos médicos de tamaño limitado.

Además, el código del repositorio proporciona una estructura flexible para la carga de datos, el preprocesamiento de imágenes y el entrenamiento de la red, permitiendo adaptaciones según las necesidades específicas del presente trabajo.

* + 1. **Primer entrenamiento**
       1. **División de conjunto de datos y configuración del entrenamiento**

La división del conjunto de datos se realizó mediante un proceso estratificado en dos pasos, asegurando que la proporción de imágenes T1 y T2 se mantuviera en cada subconjunto. En primer lugar, el conjunto completo de imágenes se dividió en un 70 % destinado al entrenamiento y un 30 % restante que se utilizó temporalmente como conjunto de evaluación. Posteriormente, este 30 % se subdividió nuevamente, asignando:

* Un 20 % del total de imágenes al conjunto de test.
* Un 10 % del total de imágenes al conjunto de validación.

Esta división se aplicó de manera estratificada para asegurar que ambas clases (T1 y T2) estuvieran representadas proporcionalmente en cada subconjunto. En total, el conjunto completo incluye 363 imágenes, de las cuales 174 corresponden a imágenes ponderadas en T1 y el resto (189) a imágenes ponderadas en T2.

La información de cada subconjunto se guardó en tres archivos CSV independientes (train.csv, val.csv y test.csv), cada uno con el mismo formato que el archivo original: una columna con la ruta de la imagen y otra columna con la etiqueta correspondiente (T1 o T2).

Una vez finalizada esta organización, se ejecutó el script encargado de realizar el entrenamiento del modelo de clasificación. La red neuronal utilizada fue ResNet18, con la siguiente configuración de hiperparámetros:

* Batch size: 1
* Número de épocas: 70
* Tasa de aprendizaje (Learning rate): 0.001
* Arquitectura: ResNet18

Además, el código está configurado para trabajar con un subconjunto específico de cortes axiales (Slices) de cada imagen 3D. Concretamente, se seleccionan 10 Slices centrales de cada volumen. Este proceso consiste en identificar el número total de cortes en el eje Z para cada imagen y extraer 10 cortes en torno al plano medio de la imagen. La razón de esta selección es que, en resonancia magnética, los cortes centrales suelen contener la mayor cantidad de información anatómica relevante, evitando cortes iniciales o finales que pueden incluir regiones externas o artefactos de bordes. Esta estrategia de trabajo con Slices centrales es habitual en el procesamiento de imágenes volumétricas, ya que reduce la cantidad de datos a procesar sin perder información crítica para la clasificación.

Adicionalmente, para adaptar la arquitectura de la red a este tipo de entrada multicanal, se realizó una modificación en la primera capa convolucional de la ResNet18. Por defecto, esta capa está diseñada para procesar imágenes RGB, es decir, con tres canales. En este caso, se ajustó para aceptar los 10 Slices como canales de entrada, convirtiendo así la imagen volumétrica en una imagen multicanal de 10 capas. Este ajuste permite que la red pueda capturar información contextual a lo largo del eje axial.

Por último, se modificó la capa fully connected final, configurándola con dos neuronas para que el modelo realice una clasificación binaria entre las clases T1 y T2. Estas adaptaciones permiten que la arquitectura ResNet18 sea compatible con imágenes médicas volumétricas, optimizando su rendimiento para la tarea específica planteada.

Texto

Descripción generada automáticamente

FIgura 6 Código adaptación resnet18

**Gráfico

Descripción generada automáticamente con confianza media**

FIgura 7 Arquitectura original ResNet18

~~¿Explicar las capas de la arquitectura y la definición de los hiperparámetros?~~

¿Poner y explicar parte del código 🡪 Ponerlo a mano?

Interfaz de usuario gráfica, Escala de tiempo

Descripción generada automáticamente

FIgura 8 Modificar esta imagen que es la estructura de mi modelo, pero adaptarla para a lo mío

* + - 1. **Entrenamiento y Métricas**

Una vez completado el primer entrenamiento y evaluación del modelo sobre el conjunto de test, se obtuvieron las siguientes métricas:

Gráfico

Descripción generada automáticamente

FIgura 9 Matriz de confusión

Reporte de Clasificación: Referenciar como tabla

| **Clase** | **Precisión** | **Recall** | **F1-Score** | **Soporte** |
| --- | --- | --- | --- | --- |
| T1 | 0.71 | 0.37 | 0.48 | 41 |
| T2 | 0.63 | 0.88 | 0.73 | 50 |

La precisión global (accuracy) alcanzada fue de 0.65 sobre un total de 91 imágenes del conjunto de test.

El análisis de estas métricas puso de manifiesto que, si bien el modelo es capaz de identificar correctamente una parte relevante de las imágenes T2, su rendimiento al clasificar imágenes T1 es notablemente inferior, con un recall de tan solo 0.37. Este desequilibrio en el desempeño sugiere que el modelo no ha sido capaz de capturar de forma robusta las diferencias entre ambas clases, lo cual podría deberse, en parte, al tamaño relativamente reducido del conjunto de entrenamiento y a la posible heterogeneidad de las imágenes disponibles.

En base a estos resultados preliminares, se decidió que una estrategia viable para mejorar el rendimiento del modelo consistiría en ampliar significativamente el conjunto de datos, incorporando un mayor número de imágenes de ambas clases y aumentando así la diversidad de los casos incluidos en el entrenamiento. Un conjunto de datos más amplio y representativo permitiría al modelo aprender características más generales y robustas, mejorando su capacidad de generalización y su desempeño sobre nuevos datos.

* + - 1. **Definición de Métricas de la Red Clasificadora**

Para interpretar correctamente los resultados obtenidos tras la evaluación del modelo de clasificación, es necesario definir y explicar el significado de las métricas utilizadas, así como su modo de cálculo. Las principales métricas reportadas son:

**Precisión (Precision)**  
La precisión mide la proporción de predicciones correctas de una determinada clase respecto al total de predicciones realizadas para esa misma clase. En términos matemáticos, se calcula como:

¿Explicar antes las abreviaturas 🡪 Glosario? 🡪 ¿Terminología Inglesa (italic)?

Una alta precisión indica que el modelo comete pocos errores al clasificar una imagen como perteneciente a una clase determinada. En este caso, indica qué porcentaje de las imágenes que el modelo ha clasificado como T1 o T2 son realmente de esa clase.

**Recall (Sensibilidad o Exhaustividad) Referenciar ecuaciones y poner las palabras en ingles en cursiva al menos la primera vez que aparecen (ponerla en el glosario)**  
El recall mide la capacidad del modelo para identificar correctamente todos los ejemplos de una clase específica. Su fórmula es:

Un recall elevado indica que el modelo es capaz de detectar la mayoría de las imágenes reales de una clase concreta, aunque puede tener falsos positivos. En el contexto de este trabajo, un bajo recall en T1 indicaría que el modelo está "perdiendo" muchas imágenes que realmente son T1.

**F1-Score**  
El F1-Score es la **media armónica** entre la precisión y el recall, proporcionando un equilibrio entre ambas métricas. Es especialmente útil cuando existe un cierto **desequilibrio entre clases**, ya que penaliza fuertemente los modelos que tienen alta precisión, pero bajo recall, o viceversa. Se calcula como:

En este caso, un F1-Score más bajo para la clase T1 refleja que el modelo tiene dificultades tanto para identificar correctamente las imágenes T1 como para evitar asignar erróneamente imágenes de T2 como T1.

Matriz de Confusión:  
La matriz de confusión es una representación tabular que muestra el rendimiento de un clasificador al comparar sus predicciones con las etiquetas reales. En este caso, es una matriz 2x2, donde:

* Cada fila representa las instancias reales (clase T1 o T2).
* Cada columna representa las predicciones realizadas por el modelo.
* La diagonal principal (de arriba a la izquierda hacia abajo a la derecha) muestra las predicciones correctas (verdaderos positivos para cada clase).
* Las celdas fuera de la diagonal principal representan errores de clasificación (falsos positivos y falsos negativos).



FIgura 10 MATRIZ DE CONFUSION

Finalmente, se incluye la Precisión Global (Accuracy), que mide el porcentaje total de predicciones correctas sobre el conjunto de test. Se calcula como:

Aunque es una métrica global sencilla de interpretar, puede ser engañosa en caso de desbalance entre clases, motivo por el cual suele complementarse con las métricas de precisión, recall y F1-Score.

* + 1. **Ampliación del conjunto de datos**

Tras el primer entrenamiento y evaluación del modelo de clasificación, se consideró necesario ampliar significativamente el conjunto de datos para mejorar la robustez y capacidad de generalización del modelo. Para ello, se mantuvo comunicación directa con el equipo encargado de la gestión y almacenamiento de los conjuntos de datos de resonancia magnética disponibles. En total, se identificaron **3725 casos** de resonancia magnética almacenados en format .nrrd.

En primer lugar, se verificó que los estudios estuvieran adquiridos en **corte axial**, ya que este es el plano habitual utilizado tanto en tareas de segmentación anatómica como en la configuración de la red empleada para la clasificación. Además, se comprobó que cada volumen contara con un mínimo de **30 cortes o slices**, asegurando así que existiera suficiente información volumétrica para poder extraer los **10 cortes centrales** necesarios para el preprocesamiento de cada imagen.

Otro criterio relevante en la selección fue la exclusión de aquellos casos identificados como **multivolumen**, es decir, volúmenes que contienen múltiples adquisiciones o series dentro de un mismo archivo. Este tipo de volúmenes puede generar inconsistencias durante el preprocesamiento y afectar al rendimiento del modelo, al mezclar imágenes con diferentes parámetros de adquisición. Para realizar este control, se emplearon inicialmente herramientas de inspección disponibles en el software **3D Slicer**, lo que permitió revisar manualmente las propiedades de cada volumen, incluyendo el número total de cortes, la orientación de los mismos y la posible existencia de múltiples series dentro de un único archivo.

Con el objetivo de agilizar y estandarizar este proceso, se desarrolló un script de automatización que analiza directamente los metadatos de los archivos **.nrrd**. Este script permite comprobar de manera automática la **matriz de adquisición**, así como otros campos relevantes de los metadatos, detectando tanto la orientación axial como la presencia de múltiples series en un mismo archivo. Gracias a esta combinación de inspección manual inicial y automatización posterior, fue posible realizar un filtrado exhaustivo, asegurando que el conjunto final de datos cumpliera con los requisitos necesarios para su incorporación al flujo de trabajo de clasificación.

Tras este proceso de filtrado, el conjunto final quedó compuesto por **2286 casos**, que cumplen con los criterios establecidos y representan un aumento significativo respecto al conjunto inicial, mejorando así la diversidad y representatividad del dataset utilizado para el entrenamiento.

* + 1. **Segundo entrenamiento**
       1. **División de conjunto de datos y configuración del entrenamiento**

Una vez completado el proceso de filtrado y selección de los casos válidos, el conjunto final quedó compuesto por un total de 2286 volúmenes de resonancia magnética, de los cuales 1588 corresponden a imágenes ponderadas en T1 y 698 a imágenes ponderadas en T2. Este conjunto de datos ampliado permitió entrenar un nuevo modelo de clasificación, utilizando una configuración revisada y adaptada al mayor volumen de datos disponible.

Si bien se es consciente de que existe un cierto desbalance entre ambas clases, durante el proceso de división en subconjuntos de entrenamiento, test y validación, se procuró que la distribución de imágenes de cada clase en cada partición fuera lo más equilibrada posible. Esta estratificación permitió asegurar que tanto el modelo como las métricas obtenidas durante la evaluación reflejaran un comportamiento consistente frente a ambos tipos de imágenes.

Para iniciar el proceso de entrenamiento, se prepararon los correspondientes archivos Excel y CSV que sirvieron de base para la carga de datos y la división en subconjuntos. Tras realizar la división estratificada, el conjunto de datos quedó repartido de la siguiente forma:

* 1600 imágenes para entrenamiento
* 457 imágenes para test
* 229 imágenes para validación

Con esta nueva partición, se procedió a entrenar nuevamente la red neuronal ResNet18, utilizando la misma arquitectura adaptada para trabajar con los 10 slices centrales de cada volumen. En este segundo entrenamiento, se utilizó la siguiente configuración de hiperparámetros:

* Batch size: 16
* Número de épocas: 100
* Tasa de aprendizaje (Learning rate): 0.001
* Red empleada: ResNet18
* Slices centrales utilizados: 10

Este segundo entrenamiento, al contar con un conjunto de datos considerablemente mayor, permitió evaluar la capacidad del modelo para generalizar mejor y detectar patrones más robustos que diferencien las imágenes ponderadas en T1 de las ponderadas en T2.

* + - 1. **Entrenamiento y Métricas de validación**

Tras completar el segundo entrenamiento, se evaluó el rendimiento del modelo sobre el conjunto de validación, obteniéndose los siguientes resultados:

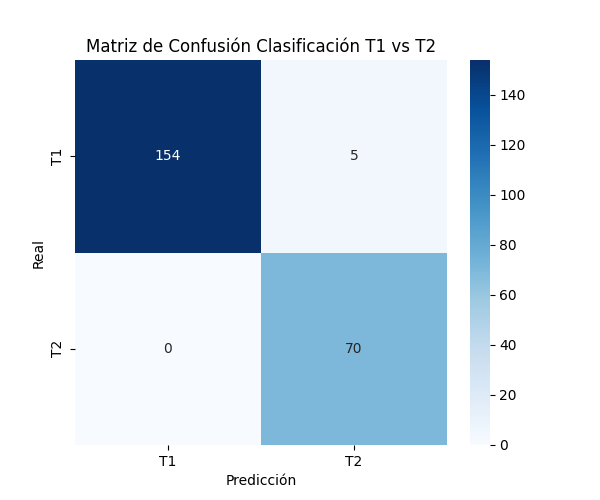


FIgura 11 Matriz de confusión

Reporte de Clasificación:

| **Clase** | **Precisión** | **Recall** | **F1-Score** | **Soporte** |
| --- | --- | --- | --- | --- |
| T1 | 0.98 | 0.96 | 0.97 | 159 |
| T2 | 0.92 | 0.96 | 0.94 | 70 |

La precisión global (accuracy) alcanzada fue de 0.96 sobre un total de 229 imágenes del conjunto de validación.

Tras finalizar el proceso de entrenamiento, se registró el progreso de la pérdida (loss) y la precisión (accuracy) tanto en el conjunto de entrenamiento como en el de validación a lo largo de las 100 épocas. En la Figura X se muestra la evolución de estas métricas durante el entrenamiento.

En la gráfica superior se representa la evolución de la función de pérdida (loss) para los conjuntos de entrenamiento y validación. Se observa una clara tendencia decreciente en ambas curvas, lo que indica que el modelo es capaz de ajustar progresivamente sus parámetros para minimizar el error. Sin embargo, se aprecia una mayor inestabilidad en la curva de validación, lo que sugiere cierta sensibilidad a las características específicas de los datos de validación, algo habitual en conjuntos de datos médicos debido a la alta variabilidad anatómica y de adquisición entre pacientes.

En la gráfica inferior se muestra la evolución de la precisión (accuracy) sobre el conjunto de entrenamiento. Se observa que la precisión aumenta de forma rápida en las primeras épocas y se estabiliza posteriormente, alcanzando valores cercanos al 100% en algunas fases. Estas gráficas permiten observar la progresión general del aprendizaje del modelo y detectar posibles síntomas de sobreajuste, que en este caso parecen controlados gracias a la consistencia entre ambas curvas.

Gráfico, Histograma

Descripción generada automáticamente

FIgura 12 Gráfica de Progreso del entrenamiento

Una vez completado el entrenamiento, se procedió a evaluar el modelo sobre el conjunto de test, obteniéndose los siguientes resultados:

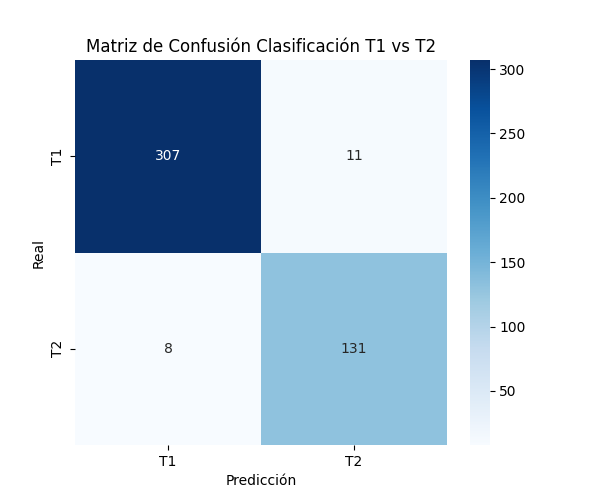


Figura 13 Matriz de confusión

Reporte de Clasificación:

| **Clase** | **Precisión** | **Recall** | **F1-Score** | **Soporte** |
| --- | --- | --- | --- | --- |
| T1 | 0.97 | 0.97 | 0.97 | 318 |
| T2 | 0.92 | 0.94 | 0.93 | 139 |

La precisión global (accuracy) alcanzada sobre el conjunto de test fue de 95.39%, con un macro promedio de precisión, recall y f1-score de 0.95, lo que refleja un rendimiento elevado y equilibrado entre ambas clases, manteniéndose en valores similares a los obtenidos durante la validación.

Con el objetivo de comprender mejor las limitaciones del modelo y analizar los casos más problemáticos, se realizó una revisión visual de las imágenes que fueron clasificadas erróneamente. Esta inspección manual permitió observar que muchas de estas imágenes presentaban características atípicas o condiciones particulares, como artefactos de imagen, zonas de bajo contraste o ruidos que dificultaban la diferenciación entre secuencias T1 y T2. Este análisis post-entrenamiento permitió identificar posibles áreas de mejora en el preprocesamiento de las imágenes o en la incorporación de técnicas adicionales para aumentar la robustez del modelo ante este tipo de casos complejos.

¿Debería de poner fotos de ejemplo de las imágenes ruidosas?

* + 1. **Tercer entrenamiento**
       1. **División de conjunto de datos y configuración del entrenamiento**

Tras analizar la evolución de las curvas de pérdida y precisión durante el segundo entrenamiento, se observó que la red aún no había alcanzado un estado completamente estable y que existía margen de mejora si se continuaba el proceso de entrenamiento durante más épocas. Por este motivo, se decidió realizar un reentrenamiento de la red de clasificación de imágenes T1 y T2, manteniendo el mismo conjunto de datos, la misma arquitectura y configuración general, con la única modificación de aumentar el número de épocas de entrenamiento.

Para este reentrenamiento, se reutilizaron los mismos archivos Excel y CSV previamente generados, manteniendo los siguientes tamaños de los subconjuntos:

* 1600 imágenes para entrenamiento
* 457 imágenes para test
* 229 imágenes para validación

El reentrenamiento se llevó a cabo utilizando la red ResNet18 adaptada, con la misma configuración de hiperparámetros que en el entrenamiento anterior, salvo por el aumento en el número de épocas:

* Batch size: 16
* Número de épocas: 200
* Tasa de aprendizaje (Learning rate): 0.001
* Red empleada: ResNet18
* Slices centrales utilizados: 10
  + - 1. **Entrenamiento y Métricas de validación**

Una vez finalizado el reentrenamiento, se evaluó el modelo sobre el conjunto de validación, obteniéndose las siguientes métricas:

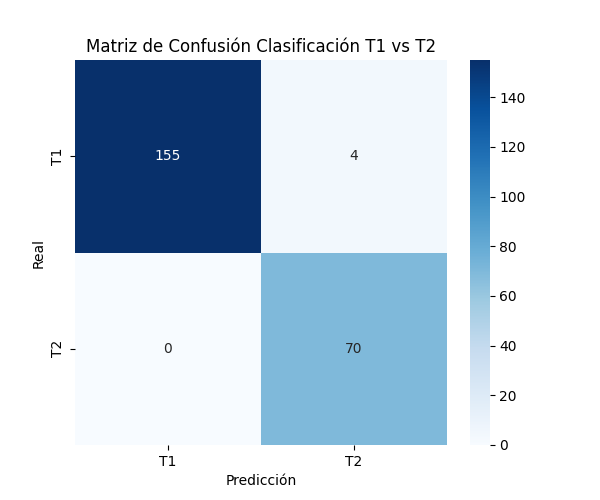


FIgura 14 Matriz de confusión

Reporte de Clasificación:

| **Clase** | **Precisión** | **Recall** | **F1-Score** | **Soporte** |
| --- | --- | --- | --- | --- |
| T1 | 0.98 | 0.97 | 0.98 | 159 |
| T2 | 0.94 | 0.96 | 0.95 | 70 |

La precisión global (accuracy) alcanzada sobre el conjunto de validación fue de 0.97, con un macro promedio de precisión, recall y f1-score de 0.96, lo que supone una ligera mejora respecto al entrenamiento anterior. Estos resultados muestran un buen equilibrio entre ambas clases, manteniendo un alto grado de acierto en la identificación tanto de imágenes T1 como de imágenes T2.

El incremento en el número de épocas permitió al modelo aprovechar al máximo la información disponible, consolidando un mejor ajuste a las características del conjunto de datos y una mayor estabilidad en las métricas obtenidas. Este último modelo se considera el **modelo definitivo** para la tarea de clasificación automática de imágenes de resonancia magnética T1 y T2, sobre el conjunto de datos utilizado.

Para registrar la evolución del proceso de entrenamiento, se generó una gráfica donde se representa la evolución de la función de pérdida (loss) y la precisión (accuracy) a lo largo de las 200 épocas, tanto para el conjunto de entrenamiento como para el de validación. Esta representación permite observar visualmente el comportamiento del modelo y detectar posibles signos de sobreajuste o problemas de convergencia.

A continuación, se muestra la gráfica correspondiente al progreso del entrenamiento tras el reentrenamiento con 200 épocas.

Imagen que contiene Gráfico

Descripción generada automáticamente

FIgura 15 Gráfica del progreso del entrenamiento

Una vez finalizado el entrenamiento, se evaluó el rendimiento del modelo sobre el conjunto de test, obteniéndose las siguientes métricas:

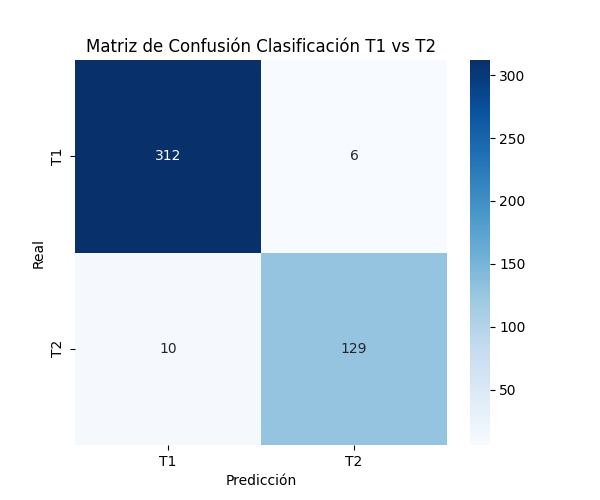


FIgura 16 Matriz de confusión

Reporte de Clasificación:

| **Clase** | **Precisión** | **Recall** | **F1-Score** | **Soporte** |
| --- | --- | --- | --- | --- |
| T1 | 0.97 | 0.98 | 0.97 | 318 |
| T2 | 0.96 | 0.93 | 0.94 | 139 |

La precisión global (accuracy) alcanzada fue de 0.96, lo que indica que el modelo es capaz de clasificar correctamente el 96% de las imágenes de test. Además, se obtuvo una Macro-accuracy de 95.39%, confirmando un buen equilibrio entre ambas clases.

Estos resultados evidencian una mejora significativa respecto al primer modelo entrenado, alcanzando un alto grado de acierto y estabilidad. Tanto la clase T1 como la clase T2 presentan valores de recall y precisión elevados, lo que confirma que el modelo es capaz de distinguir correctamente entre ambas secuencias de resonancia magnética, incluso ante imágenes no vistas previamente.

* + 1. **Separación del conjunto de datos en subtipos de T1 y T2**

Tras el segundo entrenamiento y la evaluación del modelo de clasificación binaria entre imágenes ponderadas en T1 y T2, surgió la posibilidad de abordar un nivel de clasificación más detallado. En el marco de una formación complementaria sobre técnicas de imagen en resonancia magnética, se profundizó en las distintas fases de adquisición dentro de las secuencias ponderadas en T1. A raíz de esta formación, se identificó que, además de diferenciar entre T1 y T2, las imágenes de resonancia magnética ponderadas en T1 pueden subdividirse en tres categorías adicionales, correspondientes a las diferentes fases de adquisición tras la administración de contraste:

* T1 basal (previa a la administración de contraste)
* T1 arterial (fase temprana tras la administración de contraste)
* T1 venosa (fase tardía tras la administración de contraste)

Esto llevó a plantear un nuevo objetivo dentro del proyecto, consistente en desarrollar un modelo capaz de clasificar las imágenes en estas cuatro categorías: T1 basal, T1 arterial, T1 venoso y T2. Para ello, fue necesario someter nuevamente el conjunto de datos disponibles (las 2286 imágenes seleccionadas tras el filtrado) a un proceso de revisión y reanotación por parte del equipo de técnicos de imagen médica y radiólogos. Este proceso implicó revisar cada imagen para determinar no solo si pertenecía a una secuencia T1 o T2, sino también, en el caso de T1, identificar la fase específica dentro de la adquisición con contraste.

El resultado de este proceso fue un nuevo conjunto de datos, ya no dividido únicamente en dos clases (T1 y T2), sino reorganizado en cuatro clases: T1 basal, T1 arterial, T1 venoso y T2. Esta nueva clasificación permitió plantear el desarrollo y entrenamiento de un nuevo modelo de clasificación, capaz de distinguir entre estas cuatro categorías, lo que supone un incremento en la complejidad de la tarea, pero también una mejora en la utilidad clínica del sistema.

Para ilustrar las diferencias visuales entre los distintos subtipos de imágenes ponderadas en T1, a continuación, se presentan ejemplos representativos de cada una de las categorías utilizadas en el nuevo proceso de clasificación. Estas imágenes permiten observar cómo varían el contraste y la visualización de las estructuras anatómicas en función de la fase de adquisición, lo que justifica la necesidad de diferenciarlas correctamente en un modelo de clasificación automático.

Tras la revisión exhaustiva de las 2286 imágenes disponibles por parte del equipo de técnicos de imagen médica y radiólogos, se estableció la siguiente subdivisión final del conjunto de datos:

* T2: 698 imágenes
* T1 venoso: 770 imágenes
* T1 basal: 478 imágenes
* T1 arterial: 340 imágenes

Esta reorganización del conjunto de datos permite disponer de una base de datos actualizada y adaptada a la realidad clínica, reflejando las distintas fases de adquisición de las imágenes ponderadas en T1. Esta subdivisión servirá como base para el entrenamiento de un nuevo modelo de clasificación multiclase, capaz de distinguir entre estas cuatro categorías.

Imagen que contiene tabla, foto, comida, alimentos

Descripción generada automáticamente

FIgura 17 T1 Basal MRI Image of Abdomen

Imagen que contiene interior, foto, pequeño, tabla

Descripción generada automáticamente

FIgura 18 T1 Arterial MRI Image of Abdomen

Imagen que contiene tabla, comida, pequeño, sándwich

Descripción generada automáticamente

FIgura 19 T1 Venous MRI Image of Abdomen

* + 1. **Primer entrenamiento subclases T1 y T2**
       1. **División de conjunto de datos y configuración del entrenamiento**

Tras completar el proceso de revisión y reanotación de las imágenes por parte del equipo de técnicos de imagen médica y radiólogos, se decidió realizar un nuevo entrenamiento enfocado en la **clasificación multiclase**, que permitiera diferenciar entre las distintas fases de las secuencias T1 (T1 basal, T1 arterial y T1 venoso), además de las imágenes T2. Este nuevo enfoque amplía significativamente el objetivo inicial, pasando de un modelo binario (T1 vs T2) a un modelo de clasificación en **cuatro clases**.

Para este nuevo entrenamiento, se reutilizó la estructura de trabajo previamente desarrollada, adaptando los archivos Excel y CSV para reflejar las nuevas etiquetas de cada imagen. Se mantuvieron los mismos tamaños de los subconjuntos de datos utilizados en los entrenamientos anteriores, quedando divididos de la siguiente forma:

* 1600 imágenes para entrenamiento
* 457 imágenes para test
* 229 imágenes para validación

El entrenamiento se llevó a cabo utilizando nuevamente la red **ResNet18**, adaptada para trabajar con los **10 slices centrales** de cada volumen. En este caso, la configuración de hiperparámetros utilizada fue:

* Batch size: 16
* Número de épocas: 250
* Tasa de aprendizaje (Learning rate): 0.001
* Red empleada: ResNet18
* Slices centrales utilizados: 10
  + - 1. **Entrenamiento y Métricas de validación**

Una vez finalizado el entrenamiento, se evaluó el rendimiento del modelo sobre el conjunto de validación, obteniéndose los siguientes resultados:

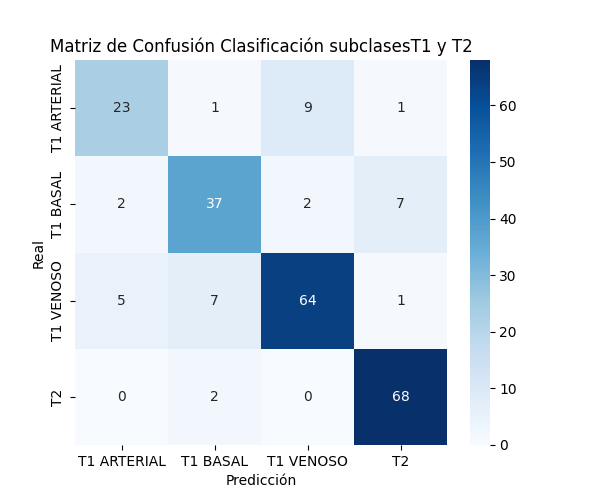


FIgura 20 matriz de confusión

Reporte de Clasificación:

| **Clase** | **Precisión** | **Recall** | **F1-Score** | **Soporte** |
| --- | --- | --- | --- | --- |
| T1 Arterial | 0.50 | 0.74 | 0.60 | 34 |
| T1 Basal | 0.79 | 0.65 | 0.71 | 48 |
| T1 Venoso | 0.78 | 0.64 | 0.70 | 77 |
| T2 | 0.87 | 0.96 | 0.91 | 70 |

La precisión global (accuracy) alcanzada sobre el conjunto de validación fue de 0.75, con un macro promedio de precisión, recall y f1-score de 0.74, 0.74 y 0.73 respectivamente, lo que refleja la mayor complejidad de la tarea al tratarse de una clasificación multiclase con categorías que presentan características visuales similares entre sí. El promedio ponderado de precisión, recall y f1-score fue de 0.77, 0.75 y 0.75 respectivamente, mostrando un desempeño razonable, aunque mejorable, especialmente en la diferenciación entre las distintas fases de adquisición de las imágenes ponderadas en T1. El modelo presenta un comportamiento más robusto al identificar las imágenes T2, mientras que en las subclases de T1 el rendimiento es algo inferior, lo que sugiere posibles líneas de mejora a través de técnicas adicionales de preprocesamiento o de aumento de datos.

Para registrar la evolución del proceso de entrenamiento, se generó una gráfica donde se representa la evolución de la función de pérdida (loss) y la precisión (accuracy) a lo largo de las 250 épocas, tanto para el conjunto de entrenamiento como para el de validación. Esta representación permite observar visualmente el comportamiento del modelo y detectar posibles signos de sobreajuste o dificultades de convergencia, especialmente al tratarse de una tarea de clasificación multiclase con categorías que presentan características visuales similares entre sí.

A continuación, se muestra la gráfica correspondiente al progreso del entrenamiento tras el entrenamiento con 250 épocas para la clasificación en cuatro clases.

Gráfico

Descripción generada automáticamente con confianza media

FIgura 21 grafica de evolución de entrenamiento

Una vez completado el entrenamiento de la red para la clasificación multiclase, se procedió a evaluar el rendimiento final sobre el conjunto de test, compuesto por 457 imágenes que no habían sido utilizadas durante el entrenamiento ni en la fase de validación. El objetivo de esta evaluación es comprobar la capacidad de generalización del modelo frente a datos completamente nuevos, asegurando que el comportamiento observado no sea producto de un sobreajuste a las imágenes utilizadas previamente.

Los resultados obtenidos en el test fueron los siguientes:

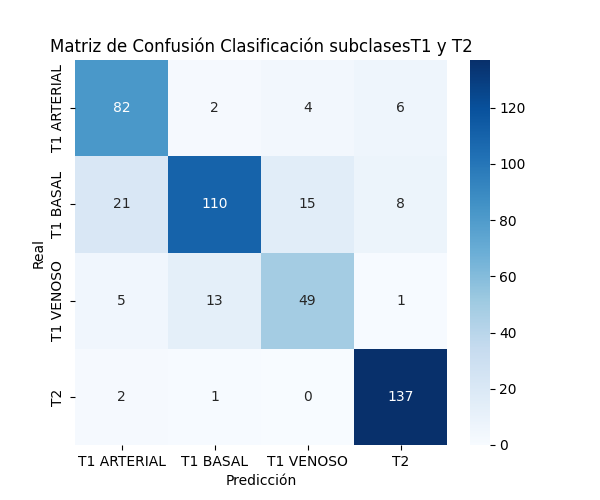


FIgura 22 matriz de confusión

Reporte de Clasificación (test):

| **Clase** | **Precisión** | **Recall** | **F1-Score** | **Soporte** |
| --- | --- | --- | --- | --- |
| T1 Arterial | 0.72 | 0.72 | 0.72 | 68 |
| T1 Basal | 0.75 | 0.87 | 0.81 | 95 |
| T1 Venoso | 0.87 | 0.71 | 0.79 | 154 |
| T2 | 0.90 | 0.98 | 0.94 | 140 |

La precisión global (accuracy) alcanzada en el test fue de **0.83,** mientras que la **Macro-accuracy** fue de **82.18%.** Estas métricas reflejan un desempeño general sólido, con un balance razonable entre las distintas clases, a pesar de la mayor complejidad que supone distinguir entre subclases de la secuencia T1 y diferenciar entre secuencias con y sin contraste.

En cuanto al desglose por clase, el modelo muestra un comportamiento especialmente robusto en la clasificación de imágenes T2, alcanzando un recall del 98%, lo que indica que prácticamente todas las imágenes T2 son identificadas correctamente. Las imágenes correspondientes a las fases T1 arterial, basal y venosa presentan un rendimiento algo más variable, siendo la fase venosa la que obtiene mejor precisión (0.87), mientras que la fase arterial muestra un balance entre precisión y recall algo más bajo, lo que refleja que es la fase más compleja de diferenciar para el modelo.

Estos resultados, si bien son satisfactorios considerando la dificultad de la tarea, podrían **mejorarse significativamente** mediante el **aumento del conjunto de datos disponible**, especialmente en las fases arterial y basal de la secuencia T1. Contar con un mayor número de ejemplos por cada subclase permitiría al modelo capturar mejor las **características distintivas** de cada fase, reduciendo la confusión entre ellas y mejorando la robustez de la clasificación.

* + 1. **Prueba de otras redes para la clasificación de T1, T2 y subclases de T1**

Poner mucha información, es decir tanto las métricas de entrenamiento como test, ¿o solo las métricas de test?

Con el objetivo de evaluar el impacto de la arquitectura de la red neuronal en el rendimiento del modelo de clasificación, se decidió realizar un proceso de comparación utilizando diferentes redes ya existentes. Este análisis permite valorar si arquitecturas con distintas profundidades, complejidades y enfoques de diseño pueden ofrecer un mejor desempeño en la tarea de clasificación de imágenes de resonancia magnética, tanto para la distinción binaria entre T1 y T2, como para la clasificación multiclase en las subclases de T1.

Para ello, se seleccionaron cuatro arquitecturas ampliamente utilizadas en tareas de visión por computador, adaptadas en este caso al formato de entrada de las imágenes médicas utilizadas:

* AlexNet
* VGG
* SqueezeNet
* MobileNet

El análisis comparativo de estas redes permitirá observar las diferencias en términos de precisión, recall y estabilidad del entrenamiento, así como la capacidad de cada arquitectura para adaptarse a un conjunto de datos médico relativamente reducido y con cierta complejidad visual, derivada de la similitud entre las distintas subclases de T1.

A continuación, se describen los procesos de adaptación y entrenamiento realizados con cada una de estas redes, así como los resultados obtenidos y su comparación con la red inicial utilizada en este trabajo (ResNet18).

SEGUIR CON ESTE PUNTO YA QUE NO LO TENGO DEL TODO HECHO

* 1. **Fase 2: Redes de Segmentación para T1, T2 y RM**

Tras la fase de clasificación, el siguiente paso consiste en desarrollar un sistema de segmentación automática capaz de identificar y delimitar diferentes estructuras anatómicas dentro de las imágenes de resonancia magnética. Esta fase tiene como objetivo aplicar redes de segmentación adaptadas, permitiendo obtener máscaras anatómicas que faciliten el análisis clínico y el procesamiento posterior de las imágenes.

Debido al tamaño disponible del conjunto de datos, se ha optado por dividir el proceso de segmentación en tres categorías generales:

* T1, correspondiente a imágenes ponderadas en secuencia T1.
* T2, correspondiente a imágenes ponderadas en secuencia T2.
* MR, que agrupa conjuntamente las imágenes T1 y T2, permitiendo aprovechar el máximo número de casos disponibles.

La cantidad de datos disponible para cada una de estas categorías es la siguiente:

* T1: 174 casos.
* T2: 386 casos.
* MR: 560 casos (resultado de combinar los casos T1 y T2).

Dado que el proceso de segmentación puede beneficiarse de un enfoque más específico por tipo de estructura, el planteamiento de esta fase contempla el desarrollo de múltiples redes de segmentación especializadas, cada una enfocada en un conjunto concreto de estructuras anatómicas. En lugar de entrenar un único modelo encargado de segmentar todas las estructuras de forma simultánea, se opta por dividir el proceso en **etapas consecutivas**, donde cada red se especializa en una categoría concreta de tejidos u órganos.

Este enfoque modular tiene varias ventajas. Por un lado, permite optimizar el uso de memoria y reducir el riesgo de fugas de memoria que podrían surgir al procesar demasiadas clases o estructuras en un único modelo. Al dividir la tarea, cada red trabaja con un **número reducido de clases**, lo que facilita la convergencia y reduce la complejidad interna del modelo.

Además, al encadenar redes en un flujo secuencial, es posible utilizar la segmentación obtenida por una red como **punto de partida o restricción** para las siguientes, de modo que las predicciones de cada etapa se ajustan mejor a las regiones colindantes ya detectadas. Esta estrategia permite que las siguientes redes se enfoquen únicamente en áreas donde hay mayor probabilidad de encontrar estructuras específicas, mejorando así la precisión general. Dependiendo de la **confianza con la que la red anterior haya realizado sus predicciones**, las siguientes redes podrán ajustar su comportamiento, incrementando o reduciendo el nivel de detalle en función de la fiabilidad de los datos previos.

En conjunto, este flujo modular y progresivo contribuye a optimizar los recursos computacionales, facilitar el entrenamiento y mejorar la precisión final de las segmentaciones, adaptándose mejor a la complejidad de las imágenes médicas de resonancia magnética.

A lo largo de esta fase se describirá el proceso de preparación de los datos, el diseño del flujo de segmentación y el desarrollo de las distintas redes, empleando como base la estructura de trabajo proporcionada por nnU-Net, una de las arquitecturas más utilizadas y validadas para segmentación de imágenes médicas en la actualidad.

* + 1. **Preparación del conjunto de datos para la segmentación**

Para llevar a cabo la fase de segmentación, fue necesario preparar un **conjunto de datos específico**, compuesto por imágenes de resonancia magnética y sus correspondientes segmentaciones anatómicas, que constituyen el **ground truth** de cada caso. Estas segmentaciones fueron proporcionadas por el equipo de técnicos de imagen médica y radiólogos, quienes revisaron y anotaron manualmente las estructuras anatómicas de interés en cada volumen.

Con el objetivo de organizar adecuadamente el conjunto de datos, las imágenes y sus segmentaciones asociadas fueron clasificadas en **tres carpetas principales**, en función de la secuencia de adquisición:

* Carpeta T1, que contiene las imágenes y segmentaciones obtenidas en secuencias ponderadas en T1.
* Carpeta T2, que contiene las imágenes y segmentaciones obtenidas en secuencias ponderadas en T2.
* Carpeta RM, que agrupa conjuntamente las imágenes y segmentaciones de T1 y T2, creando un conjunto combinado más amplio.

Esta organización facilita tanto la preparación de datos como la futura comparación de resultados entre las redes especializadas en cada secuencia y las redes entrenadas sobre el conjunto combinado.

* + 1. **Selección de las estructuras anatómicas a segmentar**

Dado que el objetivo es desarrollar un flujo modular de segmentación progresiva, se decidió comenzar por bloques de **segmentación,** dividiendo el proceso en diferentes categorías de estructuras. Este enfoque permite entrenar redes especializadas en subconjuntos concretos de estructuras anatómicas, facilitando la creación de un flujo de segmentación progresivo y optimizado.

Las estructuras serán seleccionadas por su importancia clínica y por la necesidad de diferenciarlas correctamente en procedimientos como la resección tumoral o la planificación de trasplantes, donde la correcta identificación de algunas de estas estructuras anatómicas es clave.

Antes de iniciar el desarrollo de las redes de segmentación para cada una de estas estructuras, se llevó a cabo un análisis exploratorio del contenido real de los datos disponibles. Para ello, se desarrolló un **código de análisis de segmentos anatómicos (¿explicar el código o poner fotos?)**, que recorre todo el conjunto de datos (clasificado previamente en las carpetas T1, T2 y RM) e identifica qué estructuras anatómicas están presentes en cada caso.

Este análisis genera tres gráficas, una por cada categoría (T1, T2 y RM), donde se representa el **recuento de cada estructura** dentro de cada modalidad de imagen. De esta forma, es posible visualizar la **disponibilidad real de cada segmento anatómico** en el dataset, permitiendo tomar decisiones informadas sobre qué estructuras tienen **suficientes ejemplos** para entrenar redes de segmentación y cuáles presentan una presencia insuficiente, lo que podría dificultar o impedir su segmentación automática.

Estas gráficas sirven como herramienta de apoyo para definir la estrategia de segmentación modular, ya que permiten priorizar el desarrollo de redes especializadas en aquellas estructuras con mayor presencia en el dataset y, al mismo tiempo, identificar qué estructuras requieren **alguna estrategia como el aumento de datos.**

El análisis de estos gráficos será el punto de partida para decidir el orden y el alcance de los bloques de segmentación que conformarán el flujo modular de segmentación anatómica.

Gráfico, Histograma

Descripción generada automáticamente

FIgura 23 Estructuras para T1

Gráfico, Histograma

Descripción generada automáticamente

FIgura 24 Estructuras para T2

Gráfico, Histograma

Descripción generada automáticamente

FIgura 25 Estructuras para RM

Tras analizar las gráficas, se decidió segmentar las siguientes estructuras para las 3 carpetas del conjunto de datos:

* **Vasculatura portal**, que incluye la vena porta y sus ramas principales.
* **Vasculatura arterial**, que comprende la arteria hepática y sus principales bifurcaciones, no estoy seguro.
* **Vasculatura venosa**, que abarca las venas hepáticas y la vena cava inferior, tampoco estoy seguro.
* Páncreas
* Hígado
* … (seguir)
  + 1. **Bloques de segmentación:** 
       1. **Vasculaturas**

El primer bloque de segmentación desarrollado corresponde a la **segmentación de vasculaturas**, por tratarse de un conjunto de estructuras anatómicas esenciales en la planificación quirúrgica, especialmente en cirugías abdominales y hepáticas. La correcta identificación y delimitación de las principales estructuras vasculares permite mejorar la seguridad de los procedimientos y proporciona un contexto anatómico indispensable para el análisis y segmentación de otros tejidos y órganos.

Para preparar el conjunto de datos específico para este bloque, se desarrolló un **código de extracción** que permite generar un nuevo subconjunto de datos a partir de las segmentaciones completas proporcionadas por el equipo médico. Este código identifica y separa exclusivamente las regiones etiquetadas como vasculatura portal, vasculatura venosa y vasculatura arterial.

Este proceso de extracción genera un conjunto de datos específico, en el que cada imagen está asociada únicamente a la **máscara de su componente vascular correspondiente**, eliminando el resto de las estructuras segmentadas. De este modo, se facilita el entrenamiento de redes especializadas en la detección y segmentación de cada tipo de vaso sanguíneo de forma independiente.

El objetivo de este primer bloque es que el modelo aprenda a identificar y segmentar con precisión las principales estructuras vasculares, creando una base anatómica clara que podrá ser utilizada como **referencia o restricción** para los siguientes bloques de segmentación. De esta forma, las futuras redes enfocadas en tejidos blandos, órganos sólidos u otras estructuras anatómicas podrán beneficiarse de la información vascular ya segmentada, mejorando la coherencia y precisión global del flujo de segmentación modular

**Entrenamiento inicial de la red de segmentación de vasculaturas en imágenes T1**

El primer entrenamiento específico dentro del bloque de segmentación de vasculaturas se llevó a cabo utilizando exclusivamente los casos correspondientes a **secuencias T1**, con el objetivo de evaluar el rendimiento de la red en este tipo concreto de adquisición. Tras la ejecución del proceso de extracción de vasculaturas a partir de las segmentaciones completas, se obtuvo un total de **126 casos válidos con presencia de estructuras vasculares**.

Este conjunto de datos fue dividido en tres subconjuntos:

* 100 casos para entrenamiento.
* 26 casos para test.

Para este primer entrenamiento, se utilizó el framework **nnU-Net**, adaptando los siguientes hiperparámetros para ajustarse mejor a las características específicas de las vasculaturas y al tamaño limitado del conjunto de datos:

* Tasa de aprendizaje inicial (initial\_lr): 1e-2
* Parámetro de regularización (weight\_decay): 3e-5
* Porcentaje de oversampling de foreground (oversample\_foreground\_percent): 0.33
* Número de iteraciones por época (num\_iterations\_per\_epoch): 250
* Número de iteraciones de validación por época (num\_val\_iterations\_per\_epoch): 50
* Número total de épocas (num\_epochs): 1000
* Activación de deep supervision (enable\_deep\_supervision): True

Además, se dejó abierta la posibilidad de ajustar la tasa de aprendizaje inicial y otros hiperparámetros, en función de la evolución de las métricas de entrenamiento y validación. Este tipo de ajuste dinámico es especialmente relevante en tareas de segmentación de estructuras vasculares, dado que su morfología alargada y variable, junto con el alto contraste en ciertas zonas, puede generar dificultades adicionales en el proceso de segmentación automática.

Este primer entrenamiento sobre **T1** servirá como base de referencia para comparar posteriormente con los entrenamientos realizados sobre **T2** y el conjunto combinado **RM**, evaluando si existen diferencias relevantes en el desempeño de la red en función de la secuencia de adquisición.

Tras completar el entrenamiento de la red de segmentación para vasculaturas en el conjunto de datos de T1, se obtuvieron las siguientes métricas:

* Mean Validation Dice: 0.9019

Este valor representa el promedio de la métrica Dice obtenido durante el proceso de validación, indicando un alto grado de solapamiento entre las segmentaciones predichas por la red y las segmentaciones reales (ground truth) proporcionadas por el equipo médico. Este resultado es indicativo de una buena capacidad de la red para ajustar sus predicciones a las regiones anatómicas de interés durante el entrenamiento y validación.

Para visualizar la evolución del proceso de entrenamiento, a continuación, se muestra la gráfica que recoge la evolución de la función de pérdida(loss) y la métrica Dice, tanto en el conjunto de entrenamiento como en el de validación, a lo largo de las 1000 épocas. Esta representación gráfica permite observar la convergencia de la red y detectar posibles signos de sobreajuste o problemas de estabilidad.

Interfaz de usuario gráfica, Aplicación, Word

Descripción generada automáticamente

FIgura 26 Curvas de entrenamiento T1

Una vez finalizado el entrenamiento, se evaluó el modelo sobre el conjunto de test, compuesto por **26 casos**, obteniéndose las siguientes métricas globales:

* **Foreground Mean Dice**: 0.7438
* **Foreground IoU**: 0.6174

Además de las métricas globales, se analizaron de forma independiente las métricas para cada una de las tres estructuras vasculares segmentadas:

* **Vasculatura portal** (Clase 1):
  + Dice: 0.6455
  + IoU: 0.4885
  + Falsos Negativos (FN): 10073.8
  + Falsos Positivos (FP): 4865.3
  + Verdaderos Positivos (TP): 11821.8
  + Verdaderos Negativos (TN): 11946112.9
* **Vasculatura arterial** (Clase 2):
  + Dice: 0.7419
  + IoU: 0.6103
  + Falsos Negativos (FN): 10329.6
  + Falsos Positivos (FP): 6658.4
  + Verdaderos Positivos (TP): 22422.6
  + Verdaderos Negativos (TN): 11933463.3
* **Vasculatura venosa** (Clase 3):
  + Dice: 0.8440
  + IoU: 0.7534
  + Falsos Negativos (FN): 5520.7
  + Falsos Positivos (FP): 4180.1
  + Verdaderos Positivos (TP): 21716.7
  + Verdaderos Negativos (TN): 11941456.4

**Explicación de Métricas en segmentación de Imagen Médica**

**Dice Coefficient (Dice)**: Mide el solapamiento entre la segmentación predicha y la segmentación de referencia (ground truth). Su valor varía entre 0 (sin coincidencia) y 1 (coincidencia perfecta). Es la métrica más utilizada en segmentación médica, ya que evalúa cuánto de la predicción coincide con la segmentación manual.

**Intersection over Union (IoU)**: Similar al Dice, pero algo más estricto, ya que mide el área de intersección dividida entre el área total unida de predicción y ground truth. Penaliza más las discrepancias y suele ser inferior al Dice.

**Verdaderos Positivos (TP)**: Píxeles correctamente segmentados que coinciden con el ground truth.

**Falsos Positivos (FP)**: Píxeles segmentados que **no pertenecen** al ground truth. Esta métrica indica **píxeles segmentados de más** y es clave para detectar sobresegmentación.

**Falsos Negativos (FN)**: Píxeles presentes en el ground truth que el modelo no segmentó. Representa los **píxeles omitidos** por el modelo, indicando infrasegmentación.

**Verdaderos Negativos (TN)**: Píxeles correctamente identificados como fondo. En segmentación médica es menos relevante, ya que el interés se centra en las estructuras de interés y no en el fondo.

Es importante señalar que, en segmentación médica, las segmentaciones manuales proporcionadas por los técnicos de imagen médica o radiólogos (ground truth) **no son siempre perfectas**. Existen áreas grises donde la delimitación exacta de una estructura puede ser subjetiva o depender del criterio personal de cada experto. En algunos casos, el modelo puede detectar y segmentar estructuras o ramificaciones vasculares que el experto no incluyó inicialmente en la segmentación manual.

Cuando esto ocurre, el modelo **pinta de más** en comparación con el ground truth, generando un aumento en el número de **falsos positivos (FP)**, lo que impacta negativamente en métricas como el Dice y el IoU. Sin embargo, este comportamiento **no siempre implica un error real**. En ocasiones, el modelo puede estar **segmentando correctamente estructuras anatómicas que simplemente no estaban reflejadas en la segmentación manual original**, lo que indica que el modelo podría ser incluso **más sensible o preciso que el propio ground truth**.

Este fenómeno es relativamente frecuente en estructuras vasculares, donde pequeñas ramas o conexiones finas pueden ser omitidas en la segmentación manual, pero sí detectadas por el modelo.

En términos generales, el modelo muestra un rendimiento aceptable para segmentar estructuras vasculares en imágenes T1. Sin embargo, la presencia de falsos positivos y falsos negativos, así como la posible **subanotación** presente en el ground truth, deben ser consideradas al interpretar las métricas obtenidas. El objetivo no es únicamente maximizar los valores de Dice e IoU, sino también garantizar que la segmentación generada por el modelo sea **anatómicamente coherente y clínicamente útil**, incluso si esto implica desviarse ligeramente de la segmentación manual original.

Lo hago como memoria la siguiente parte, o lo hago como trabajo final, ¿sin importar mucho lo que he ido haciendo?

Las métricas de la matriz de confusión no tienen mucho valor