



UNIVERSITAT DE
BARCELONA

Trabajo de final de grado

GRADO DE INGENIERÍA
INFORMÁTICA

Facultat de Matemàtiques i Informàtica
Universitat de Barcelona

CycleGAN para la predicción de
la respuesta patológica completa
en cáncer de mama con imágenes
DCE-MRI

Autor: Eloi Busquets García-González

Director: Dr. Oliver Díaz Montesdeoca
Realizado a: Departamento
de Matemáticas e Informàtica
Barcelona, 10 de junio de 2025

Resumen

El cáncer de mama es una de las principales causas de mortalidad a nivel mundial. La predicción anticipada de la respuesta patológica completa (PCR) tras la quimioterapia neoadyuvante es clave para personalizar tratamientos y mejorar el pronóstico de los pacientes. Este trabajo investiga el uso de arquitecturas *Cycle-Generative Adversarial Networks* (CycleGAN) para la predicción de la PCR en el cáncer de mama a partir de imágenes de Resonancia Magnética Dinámica con Contraste (DCE-MRI), sin requerir imágenes emparejadas.

Los recursos utilizados están formados por un conjunto de datos de más de 1500 casos de DCE-MRI junto con sus variables clínicas. Se diseñó un modelo AC-CycleGAN, que extiende la arquitectura tradicional de una Cycle-GAN añadiendo una salida de clasificación auxiliar. Además, se incorporó una función de pérdida basada en máscaras para focalizar el aprendizaje en regiones de interés, como las tumorales, utilizando para ello diferentes segmentaciones. Los resultados muestran que la arquitectura propuesta no logra aprender representaciones discriminativas entre los dominios PCR0 y PCR1 en un conjunto de datos heterogéneo, especialmente en la clase minoritaria PCR1. En el caso del modelo base sin adversarialidad, los resultados son comparables, mostrando igualmente dificultades para mejorar significativamente la clasificación de la clase PCR1. Además, se integraron mapas de atención para facilitar la interpretación clínica de las predicciones.

Este trabajo concluye que, aunque el uso de CycleGAN para tareas de clasificación ha resultado ineficiente, su rendimiento ha sido comparable al del modelo base. Como líneas futuras, se plantea la exploración de enfoques multimodales, el uso de arquitecturas híbridas con mecanismos de atención y la evaluación de una arquitectura combinada con un modelo externo para mejorar su capacidad de generalización.

Abstract

Breast cancer is one of the leading causes of mortality worldwide. Early prediction of the pathological complete response (PCR) after neoadjuvant chemotherapy is crucial for personalizing treatments and improving patient prognosis. This work investigates the use of Cycle-Generative Adversarial Networks (CycleGAN) architectures for predicting PCR in breast cancer based on Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI), without requiring paired images.

The resources used include a dataset of over 1,500 DCE-MRI cases along with their associated clinical variables. An AC-CycleGAN model was designed, extending the traditional CycleGAN architecture by adding an auxiliary classification output. Additionally, a mask-based loss function was incorporated to focus the learning process on regions of interest, such as tumors, using different segmentation strategies. The results show that the proposed architecture fails to learn discriminative representations between the PCR0 and PCR1 domains in a heterogeneous dataset, particularly for the minority class PCR1. In the baseline model without adversarial training, the results are comparable, also showing difficulties in improving the classification of the PCR1 class. Moreover, attention maps were integrated to facilitate the clinical interpretation of the predictions.

This work concludes that although the use of CycleGAN for classification tasks has proven inefficient, its performance has been comparable to that of the baseline model. Future work includes the exploration of multimodal approaches, the use of hybrid architectures with attention mechanisms, and the evaluation of a combined architecture with an external model to enhance generalization capability.

Resum

El càncer de mama és una de les principals causes de mortalitat a nivell mundial. La predicció anticipada de la resposta patològica completa (PCR) després de la quimioteràpia neoadjuvant és clau per personalitzar els tractaments i millorar el pronòstic de les pacients. Aquest treball investiga l'ús d'arquitectures *Cycle-Generative Adversarial Networks* (CycleGAN) per a la predicció de la PCR en el càncer de mama a partir d'imatges de Resonància Magnètica Dinàmica amb Contrast (DCE-MRI), sense necessitat d'imatges aparellades.

Els recursos utilitzats estan formats per un conjunt de dades de més de 1.500 casos de DCE-MRI juntament amb les seves variables clíniques. Es va dissenyar un model AC-CycleGAN, que amplia l'arquitectura tradicional de CycleGAN afegint una sortida de classificació auxiliar. A més, es va incorporar una funció de pèrdua basada en màscares per focalitzar l'aprenentatge en regions d'interès, com ara les tumorsals, utilitzant diferents segmentacions. Els resultats mostren que l'arquitectura proposada no aconsegueix aprendre representacions discriminatives entre els dominis PCR0 i PCR1 en un conjunt de dades heterogeni, especialment per a la classe minoritària PCR1. En el cas del model base sense adversarietat, els resultats són comparables, mostrant també dificultats per millorar significativament la classificació de la classe PCR1. A més, es van integrar mapes d'atenció per facilitar la interpretació clínica de les prediccions.

Aquest treball conclou que, tot i que l'ús de CycleGAN per a tasques de classificació ha resultat ineficient, el seu rendiment ha estat comparable al del model base. Com a línies futures, es proposa explorar enfocaments multimodals, l'ús d'arquitectures híbrides amb mecanismes d'atenció i l'avaluació d'una arquitectura combinada amb un model extern per millorar la seva capacitat de generalització.

Agradecimientos

Quisiera expresar mi más sincero agradecimiento a todas las personas que han contribuido de alguna forma a la realización de este Trabajo de Fin de Grado.

En primer lugar, me gustaría agradecer a mi tutor, Oliver Díaz, por su orientación, apoyo y valiosas sugerencias a lo largo de todo el proyecto, que han sido cruciales no solo para el desarrollo de este proyecto, sino también para mi crecimiento personal y académico. Pero, sobre todo, le agradezco su cercanía, paciencia y calidad humana.

También quiero agradecer a mi familia, y en especial a mi madre y mi hermana, por su apoyo, comprensión y ánimo. Su presencia, siempre constante, incluso en los momentos más difíciles, ha sido fundamental para llegar hasta aquí. Del mismo modo, agradezco a mis amigos, en particular a Mar, Fanny y Julia, por estar a mi lado, por su cariño y amistad durante estos años universitarios.

A todas las personas que, de una u otra manera, han formado parte de este camino, gracias.

Índice

1. Introducción	1
1.1. Contexto	1
1.2. Motivación del problema	4
1.3. Planificación	5
1.3.1. Definición de objetivos	5
1.3.2. Planificación de tareas	6
2. Estado del arte	9
2.1. Limitaciones en el análisis de imágenes médicas	9
2.2. Generalización entre Dominios en Imágenes Médicas	9
2.2.1. Transferencia del Aprendizaje	9
2.2.2. Métodos de creación de imágenes sintéticas	11
2.3. Predicción de respuesta patológica completa (PCR) en cáncer de mama	16
3. Materiales y metodología	18
3.1. Materiales	18
3.1.1. Datos clínicos y de imagen	18
3.1.2. División del conjunto de datos	29
3.1.3. Lenguaje de programación y librerías	30
3.2. Metodología	31
3.2.1. Segmentación	31
3.2.2. Modelos propuestos	34
4. Implementación	39
4.1. Estructura del proyecto	39
4.2. <i>Pipeline</i>	39
4.2.1. Preprocesamiento	39
4.2.2. Imágenes sintéticas y Clasificación	45
5. Análisis y discusión de resultados	49
5.1. Análisis de resultados del modelos basado en una CyclesGAN	49
5.2. Análisis de resultados del modelo base	55
5.3. Comparativa entre modelos	55

6. Conclusiones y trabajo futuro	58
6.1. Trabajo futuro	58

Capítulo 1

1. Introducción

1.1. Contexto

El cáncer de mama es una de las principales causas de mortalidad a nivel mundial (véase [Figura 1.1](#)). Según estimaciones de la Organización Mundial de la Salud (WHO, por sus siglas en inglés), en 2022 se diagnosticaron 2.3 millones de nuevos casos y se registraron 670.000 muertes por esta enfermedad [1]. Además, fue la principal causa de muerte por cáncer entre mujeres ese mismo año [2].

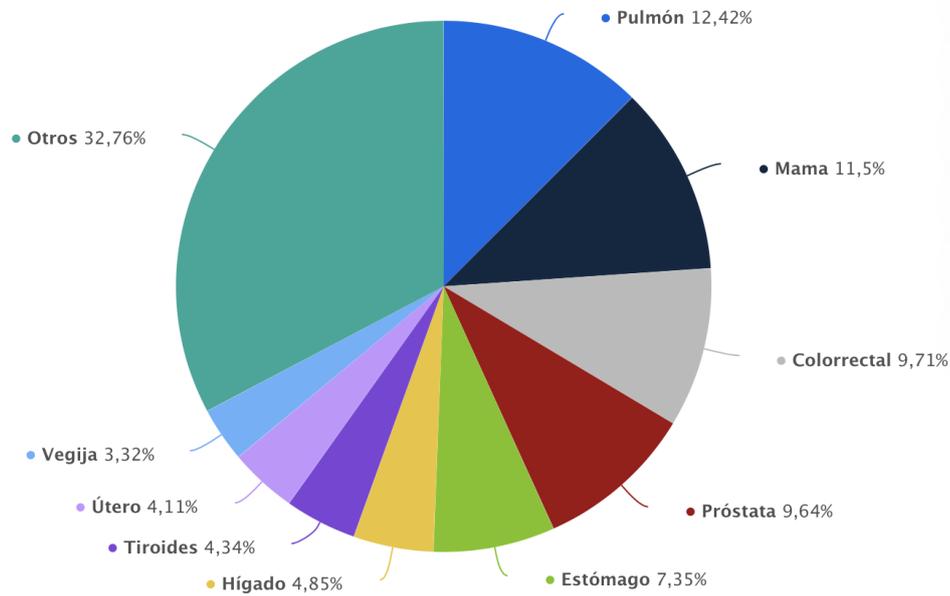


Figura 1.1: Distribución porcentual de las muertes por cáncer más frecuentes entre la población mundial en 2022 [3].

Según un estudio publicado por la *International Agency for Research on Cancer* (IARC) 1 de cada 20 mujeres en el mundo será diagnosticada con cáncer de mama a lo largo de su vida. Además, estiman que si las tasas actuales se mantienen, para 2050 podrían alcanzarse 3.2 millones de casos nuevos y 1.1 millones de muertes relacionadas anualmente (véase [Figura 1.2](#)) [4].

El diagnóstico temprano es un factor crucial para mejorar el pronóstico, aumentar las tasas de supervivencia y reducir la morbilidad y la intensidad de los tratamientos. Las opciones terapéuticas incluyen cirugía, quimioterapia, hormonoterapia, inmunoterapia y radioterapia. En este contexto, la terapia neoadyuvante (NACT, por sus siglas en inglés), introducida en la década de 1970, se ha convertido en un estándar en el tratamiento del cáncer de mama localmente avanzado. Su propósito principal es reducir el tamaño tumoral antes de la cirugía, permitiendo procedimientos menos invasivos y proporcionando una eficacia del tratamiento [5].

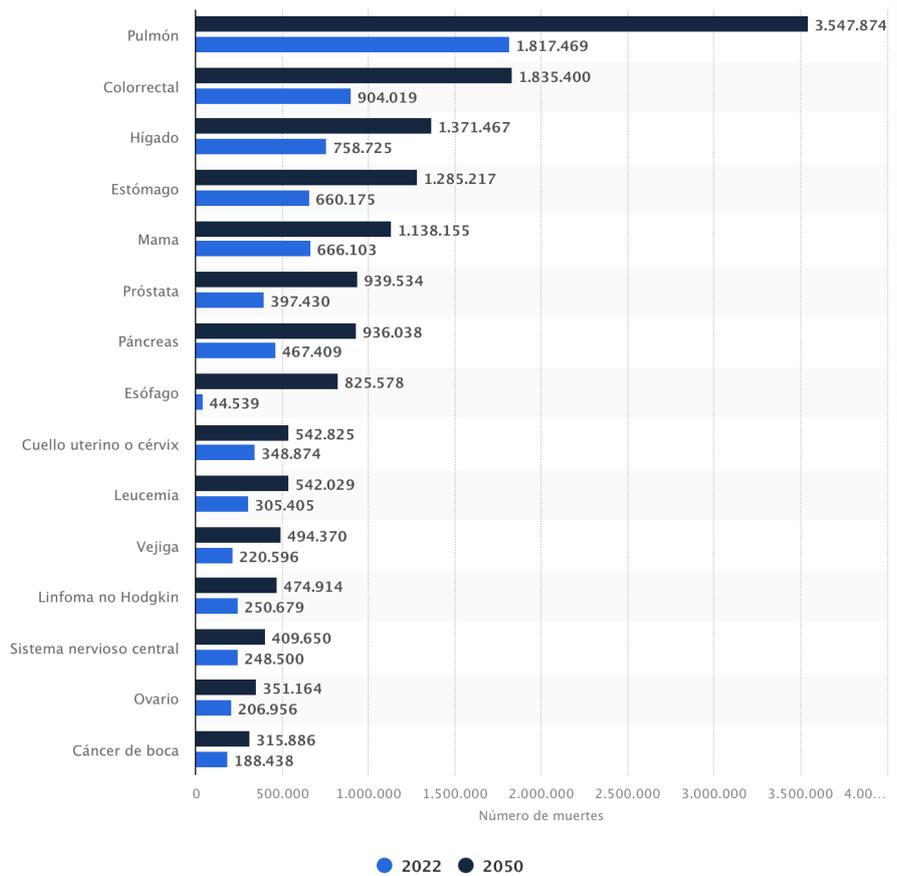


Figura 1.2: Tipos de cáncer con mayor número de nuevos casos detectados a nivel mundial en 2022 y 2050 [6].

La evaluación de la eficacia del tratamiento puede realizarse a través de diferentes indicadores clínicos, radiológicos o patológicos, además de métricas como el periodo de Supervivencia Libre de Enfermedad (*Disease-Free Survival*, DFS) o la Supervivencia Global (*Overall Survival*, OS). Sin embargo, el indicador más fiable de la supervivencia es la Respuesta Patológica Completa (*Pathological Complete Response*, PCR), definida como la ausencia completa de células tumorales invasivas en la mama y en los ganglios linfáticos tras el tratamiento [7]. Estos indicadores, junto con sus criterios clínicos, se resumen en la [Tabla 1](#).

A pesar de sus beneficios, la terapia NACT puede ocasionar efectos secundarios, lo que remarca de nuevo la importancia de una planificación terapéutica. En este contexto, resulta especialmente relevante poder predecir de forma anticipada si un paciente responderá favorablemente al tratamiento, siendo la PCR un indicador clave en esta evaluación.

La Resonancia Magnética (*Magnetic Resonance Imaging*, MRI) se considera una de las modalidades de imagen más sensibles para el diagnóstico del cáncer de mama, particularmente útil en la planificación preoperatoria y en la evaluación de la respuesta al tratamiento. Además de permitir la medición de la extensión del tejido fibroglandular —que incluye tanto el tejido fibroso como el glandular, componentes

Tabla 1: Indicadores de evaluación terapéutica y sus criterios clínicos

Métrica	Criterio Clínico
Respuesta Patológica Completa (PCR)	Desaparición total de todas las lesiones objetivo. Cualquier ganglio linfático patológico debe reducir su eje corto a menos de 10 mm.
Supervivencia Libre de Enfermedad (DFS)	Tiempo transcurrido desde el final del tratamiento hasta la detección de una recurrencia local, regional o distante del cáncer, o la aparición de un segundo tumor primario. Se considera un indicador de control del cáncer a medio plazo.
Supervivencia Global (OS)	Tiempo total desde el diagnóstico o inicio del tratamiento hasta el fallecimiento. Se considera el criterio definitivo de efectividad terapéutica a largo plazo.

esenciales de la mama—, la MRI mamaria ofrece información funcional de gran valor mediante el uso de agentes de contraste intravenosos, como los quelatos de gadolinio. Estos agentes acortan el tiempo de relajación longitudinal o T1 -tiempo que tarda la magnetización de los protones en recuperarse tras haber sido excitados por un pulso de un campo magnético- en los tejidos donde se acumulan, lo que genera una señal más intensa en las imágenes ponderadas en T1. Esta propiedad permite visualizar con alta sensibilidad cambios fisiológicos y biológicos in vivo relacionados con el riesgo y la presencia de cáncer de mama [8]. Esta herramienta se conoce como la Resonancia Magnética Dinámica con Contraste (DCE-MRI, por sus siglas en inglés).

Estas imágenes se obtienen con escáneres con una intensidad de campo magnético de al menos 1.5 teslas. La DCE-MRI ponderadas en T1 pueden realizarse con o sin supresión de grasa, y habitualmente se adquieren en el plano axial, ya que este es más rápido que la adquisición sagital y proporciona una mejor visión general de ambas mamas (bilateral) [9] (véase la [Figura 1.3](#)).

El protocolo debe obtener una adquisición basal o fase pre-contraste -antes de la administración del contraste- ponderada en T1. Tras la administración del contraste, se repite la adquisición ponderada en T1 para visualizar las anomalías que realzan (fases tempranas y tardías). Es crucial capturar una imagen aproximadamente 60–90 segundos después de la administración del contraste, ya que la mayoría de los cánceres de mama muestran su pico de realce en ese momento. La detección de lesiones se basa principalmente en estas imágenes posteriores al contraste (véase la [Figura 1.4](#)). Cuando no se emplea supresión de grasa, es necesario generar imágenes por sustracción entre las adquisiciones pre y post-contraste. Incluso con supresión de grasa, las imágenes por sustracción resultan útiles para diferenciar verdaderas lesiones con realce de aquellas con señal T1 alta en el estado basal. Sin embargo,

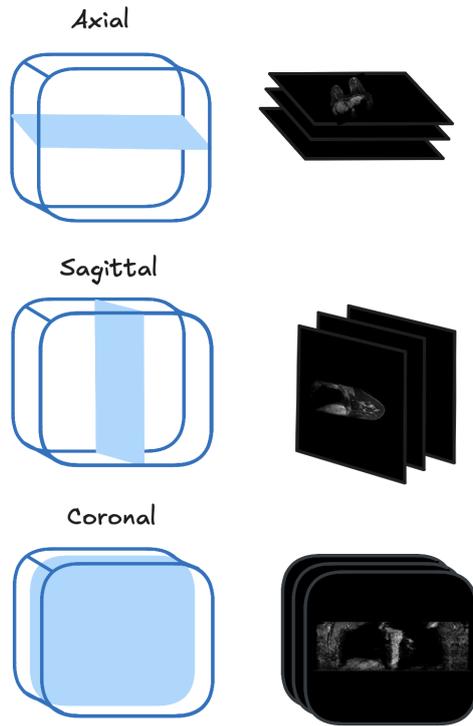


Figura 1.3: Representación de los tres planos anatómicos principales en imagen médica: corte axial (horizontal), sagittal (lateral) y coronal (frontal), junto con sus respectivas visualizaciones.

los artefactos por movimiento, los artefactos por desplazamiento químico y una supresión de grasa deficiente pueden ocultar lesiones pequeñas en las imágenes. Por convención, la MRI mamaria debe ser capaz de mostrar todos los cánceres que realzan y que tengan un tamaño de al menos 5 mm. Por ello, el grosor de corte en las imágenes ponderadas en T1 no debe superar los 2.5 mm. Para una evaluación morfológica precisa, el tamaño del píxel en el plano debe ser de 1×1 mm o menor. La adquisición de tan solo dos series ponderadas en T1 —una basal y otra aproximadamente 90 segundos después del contraste— ha demostrado ser suficiente para la detección eficaz de lesiones, como evidencian los protocolos abreviados de MRI mamaria [10].

1.2. Motivación del problema

Como ya se ha mencionado, el cáncer de mama es uno de los mayores desafíos en salud a nivel global. A pesar de los avances en las técnicas de imágenes médicas y en el análisis computacional, la detección temprana y la planificación de un tratamiento adecuado siguen siendo aspectos fundamentales para mejorar el pronóstico de los pacientes.

No obstante, todavía existen obstáculos que dificultan el desarrollo de herramientas clínicas más eficaces. Entre ellos se encuentran la falta de datos médicos

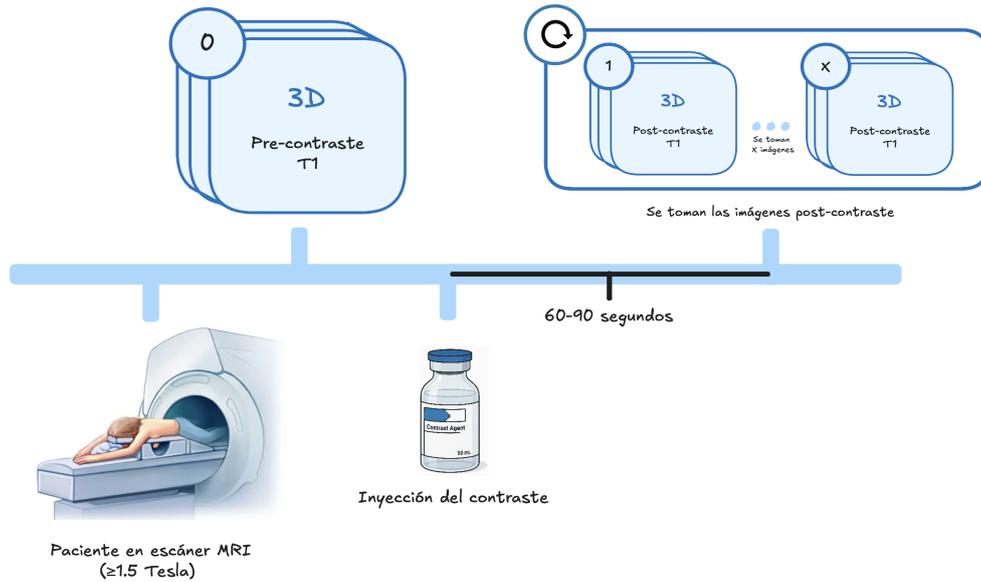


Figura 1.4: Protocolo típico de adquisición de imágenes de mama mediante MRI en secuencia T1, incluyendo la fase pre-contraste (0) y las fases post-contraste dinámicas (1 a X) tras la administración del agente de contraste.

etiquetados —ya que su interpretación requiere personal especializado—, el coste y el tiempo asociados a la adquisición de imágenes, así como la variabilidad en la procedencia y calidad de los datos [11].

En este sentido, los métodos de traducción de dominio no supervisado, como las CycleGAN (*Cycle-Generative Adversarial Networks*) ofrecen una vía para abordar algunos de estos retos, y serán explorados en este trabajo para transformar imágenes médicas entre dominios.

1.3. Planificación

La planificación es fundamental para organizar las tareas, establecer objetivos y asegurar un uso eficiente del tiempo y de los recursos disponibles. En esta sección se definen todos los elementos anteriores y se presenta la distribución de las tareas a lo largo del tiempo mediante un diagrama de Gantt.

1.3.1. Definición de objetivos

El objetivo principal de este trabajo es explorar la viabilidad del uso de una arquitectura basada en CycleGAN para la clasificación de imágenes DCE-MRI en función de la PCR.

La hipótesis principal que se plantea es que, al entrenar la CycleGAN para generar imágenes sintéticas como mecanismo de *data augmentation*, el modelo aprende implícitamente las distribuciones estadísticas de ambos dominios (origen: PCR0,

destino: PCR1). Como resultado, se espera que este conocimiento aprendido pueda aprovecharse para discriminar si una imagen pertenece a uno u otro dominio.

Para alcanzar el objetivo, se plantean los siguientes objetivos específicos:

- Investigar el estado del arte en técnicas de *image-to-image translation*, con énfasis en aplicaciones médicas y arquitecturas tipo CycleGAN.
- Revisar literatura relacionada con el uso de imágenes DCE-MRI en la predicción de la PCR en cáncer de mama.
- Recopilar, procesar y preparar un conjunto de datos adecuado de imágenes DCE-MRI.
- Diseñar e implementar una arquitectura basada en CycleGAN, adaptada al problema de clasificación según la PCR.
- Entrenar y validar el modelo.
- Analizar los resultados obtenidos para extraer conclusiones y proponer posibles mejoras o líneas de investigación futura.

1.3.2. Planificación de tareas

En esta sección se presentan tanto las etapas y tareas previstas inicialmente como una comparación con la planificación real ejecutada al final del trabajo.

Etapas y tareas A continuación se enumeran las principales etapas del proyecto junto con las tareas correspondientes:

- **Estado del arte**
 - Contextualizar la investigación mediante revisión bibliográfica.
- **Implementación**
 - Recopilación y preprocesamiento de datos.
 - Desarrollo del modelo base.
 - Implementación de técnicas de clasificación por traducción de imágenes (*CycleGAN*).
 - Evaluación y comparación de los resultados.
- **Elaboración de la memoria**
 - Redacción del documento.
 - Incorporación de correcciones.

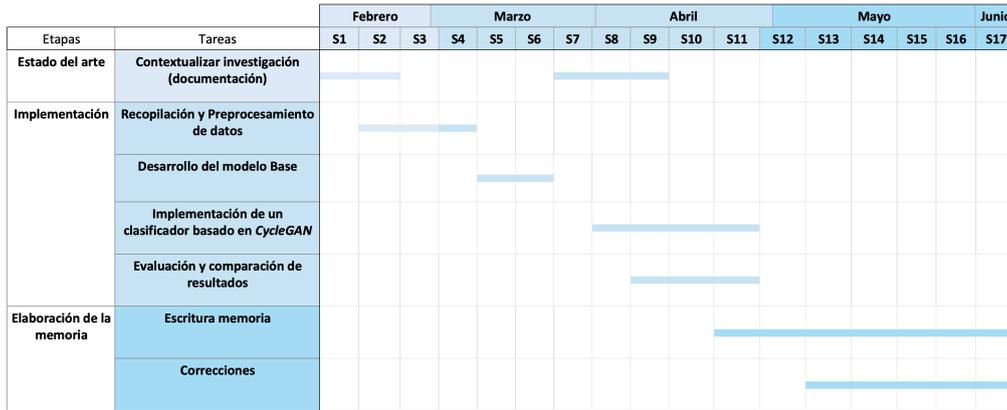


Figura 1.5: Planificación inicial del proyecto.

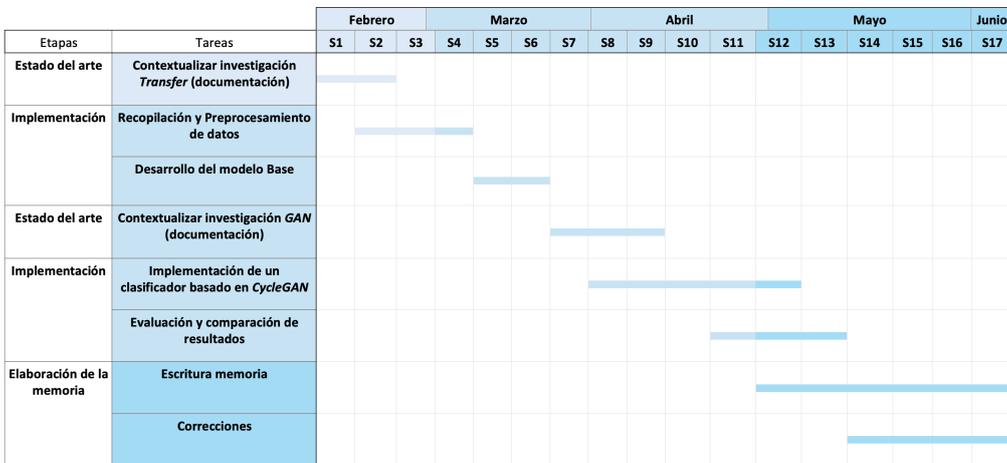


Figura 1.6: Planificación real ejecutada al finalizar el proyecto.

Diagramas de Gantt Se presentan varios diagramas de Gantt. El primero (Figura 1.5) muestra la planificación inicial definida al inicio del proyecto, mientras que el segundo (Figura 1.6) refleja la planificación real llevada a cabo.

Inicialmente, este trabajo tenía como objetivo analizar los efectos del *Transfer Learning* en la predicción de la PCR en el cáncer de mama, centrándose en conjuntos de datos provenientes de diferentes poblaciones. Sin embargo, este enfoque se vio afectado por la imposibilidad de acceder a los datos previstos. Durante la exploración de técnicas de *data augmentation*, surgió una nueva hipótesis, descrita anteriormente: si un modelo generativo como CycleGAN es capaz de sintetizar imágenes entre dominios, ¿podría también aprender implícitamente las distribuciones estadísticas subyacentes y, en consecuencia, discriminar entre los dominios de origen? Esta nueva línea de trabajo implicó adaptar la planificación. En el primer diagrama (Figura 1.5) se observa la incorporación de una nueva etapa de contextualización teórica centrada en modelos de generación de imágenes sintéticas, que reemplaza al bloque original de *Transfer Learning* presente en la planificación original (Figura 1.7). Dado que el proyecto ya había comenzado cuando se realizó el cambio de enfoque, ambas planificaciones —la inicial (Figura 1.5) y la real (Fi-

gura 1.6)— mantienen la misma estructura temporal con diferencias mínimas. No obstante, la fase de implementación, entrenamiento y validación del nuevo enfoque requirió algo más de tiempo del previsto originalmente debido a limitaciones computacionales.



Figura 1.7: Planificación inicial del proyecto para el *Transfer Learning*.

Capítulo 2

2. Estado del arte

2.1. Limitaciones en el análisis de imágenes médicas

En el análisis de imágenes médicas mediante técnicas de aprendizaje automático, es común asumir que los datos del dominio origen y los datos del dominio objetivo comparten la misma distribución. Sin embargo, esta suposición rara vez se cumple, debido a la heterogeneidad en las imágenes médicas, causadas, por ejemplo, por el uso de distintos escáneres o protocolos de adquisición [12, 13]. Esto se conoce como el problema del *domain shift*. Diferentes estudios han demostrado que el rendimiento de los modelos en la fase de test se ve significativamente afectado por este fenómeno, evidenciando que los errores se incrementan proporcionalmente a las diferencias de distribución de entrenamiento y evaluación [14].

Este problema también está estrechamente relacionado con el sesgo en la selección de muestras (*sample selection bias*), que ocurre cuando los datos utilizados para entrenar un modelo no son representativos de la tarea objetivo. Este sesgo limita la capacidad del modelo para generalizar a otros entornos clínicos [15].

Además, a diferencia del análisis de imágenes naturales —donde existen grandes conjuntos de datos etiquetados, como ImageNet[16]—, en el ámbito médico se enfrenta a una escasez de datos anotados. La anotación de imágenes médicas es un proceso costoso, lento y tedioso que, además, requiere la participación de expertos clínicos especializados [17]. Esta limitación restringe el desarrollo de modelos robustos y generalizables en entornos clínicos reales.

2.2. Generalización entre Dominios en Imágenes Médicas

Como respuesta al problema del *domain shift*, los métodos basados en el dominio han ganado considerable atención en los últimos años. Su objetivo principal es minimizar las diferencias de distribución entre el dominio fuente y el dominio objetivo, de forma que los modelos entrenados en un entorno puedan generalizar correctamente a datos provenientes de otros contextos clínicos. A continuación, se muestran diferentes métodos de generalización entre dominios.

2.2.1. Transferencia del Aprendizaje

El *transfer learning* (TL) tiene como objetivo preservar y reutilizar el conocimiento previamente adquirido en un dominio fuente para abordar nuevos problemas en un dominio objetivo, proporcionando soluciones más rápidas y eficaces [11, 14, 15, 17, 18].

Dos conceptos claves en el aprendizaje por transferencia son el dominio y la tarea. Un dominio se refiere al espacio de características de una colección de datos

específica y la distribución estadística de estas características. Por otro lado, una tarea está relacionada con el espacio de etiquetas y la función de predicción que debe aprenderse a partir de los datos del dominio [14]. Por tanto, podemos decir que el objetivo del TL es transferir el conocimiento aprendido de la tarea T_a en el dominio A a la tarea T_b en el dominio B.

Uno de los principales desafíos del TL en el contexto de las imágenes médicas es la diferencia entre conjuntos de datos. Esta diferencia puede manifestarse de diversas maneras. Por ejemplo, muchos modelos se entrenan utilizando conjuntos como ImageNet[16] de imágenes compuestas de tres canales, RGB (sigla del inglés *Red, Green, Blue*); estas redes pueden no ser adecuadas para imágenes médicas que suelen estar en escala de grises, como las radiografías (Rayos X) o la MRI, lo que limita su capacidad de generalización. Otro inconveniente habitual sería que los modelos entrenados en imágenes naturales —como personas, animales u objetos— presentan dificultades cuando se aplican a imágenes médicas. Estas últimas poseen características particulares: menor variabilidad en la textura, formas más homogéneas, diferentes resoluciones y dimensiones espaciales [11, 18]. Diversos estudios han demostrado que los modelos preentrenados con datos de dominios no médicos pueden dar una transferencia negativa, es decir, cuando la relación entre el dominio de origen y el dominio objetivo es incompatible, lo que conlleva una reducción del rendimiento del modelo [18, 19, 20, 21]. Una estrategia para abordar el problema es la adoptada por Heker y Greenspan (2020) [22], quienes entrenan un modelo sobre imágenes hepáticas para transferir el aprendizaje al análisis y segmentación de lesiones hepáticas.

Tabla 2: Estrategia de TL según la nueva tarea y el tamaño del *dataset*.

Nuevo <i>dataset</i>	Tarea parecida	Tarea diferente
<i>Dataset</i> pequeño	Solo se ajusta (reentrena) la capa de salida.	Congelar capas iniciales y entrenar las finales.
<i>Dataset</i> grande	Solo se ajusta (reentrena) la capa de salida.	Entrenamiento o reentrenamiento de todas las capas.

Una vez se ha identificado qué información del modelo fuente se puede reutilizar para la nueva tarea, se diseña la estrategia para la transferencia del aprendizaje. Esta puede incluir la extracción de características (*feature extraction*), el ajuste fino (*fine-tuning*) o la adaptación de dominio (*domain adaptation*), cada una depende del tamaño del nuevo conjunto de datos y la similitud con la tarea original (véase [Tabla 2](#)). En la extracción de características, se utilizan las capas iniciales de un modelo preentrenado como detectores de patrones genéricos [15]. En el ajuste fino, se reutilizan los pesos preentrenados como punto de partida para reentrenar parte o la totalidad del modelo (según los datos y la tarea) [11, 15, 18]. Por su parte, la adaptación de dominio busca ajustar un modelo entrenado en un dominio fuente para que funcione adecuadamente en un dominio destino con distribuciones marginales diferentes [14, 15]. En esta última, se asume que el espacio de características

y de tareas permanece igual entre dominios. En todas estas técnicas se toman los pesos del modelo preentrenado, dominio origen, y se utilizan nuevos datos.

2.2.2. Métodos de creación de imágenes sintéticas

El uso de la síntesis de imágenes en el contexto médico no solo es relevante por su capacidad actual para mejorar la precisión del diagnóstico, sino también por el potencial para mitigar diversas limitaciones asociadas al uso de datos clínicos, las limitaciones —como la heterogeneidad, la escasez, el coste y el tiempo requerido para su obtención y anotación— expuestas en la [Subsección 2.1](#).

Los modelos de creación de imágenes sintéticas en la Inteligencia Artificial generativa se pueden agrupar en redes generativas adversarias (GAN, por sus siglas en inglés), *diffusion models* y *autoencoders*. Las GANs aprenden una pérdida que trata de clasificar si la imagen creada es real o falsa, mientras simultáneamente entrena un modelo generador para minimizar esta pérdida [23]. Los *diffusion models* se basan en la reconstrucción progresiva de una imagen a partir de ruido. Los *autoencoders* están compuestos por un codificador, que reduce la dimensión de la imagen a un espacio latente, y un decodificador, que reconstruye la imagen original a partir de esa representación latente [24].

La GAN es una Red Neuronal Convolutiva (CNN, por sus siglas en inglés) que está compuesta por un generador y un discriminador (véase [Figura 2.1](#)). El generador, una CNN, intenta aprender la distribución latente de los datos reales para generar información sintética a partir de una muestra aleatoria de ruido. Por otro lado, el discriminador, una CNN, tiene como objetivo distinguir entre las imágenes reales y las imágenes sintéticas. De este modo, el entrenamiento del generador busca mejorar el proceso de generación para engañar al discriminador, mientras que este último intenta mejorar su capacidad para no dejarse engañar, lo que constituye un entrenamiento adversarial [25].

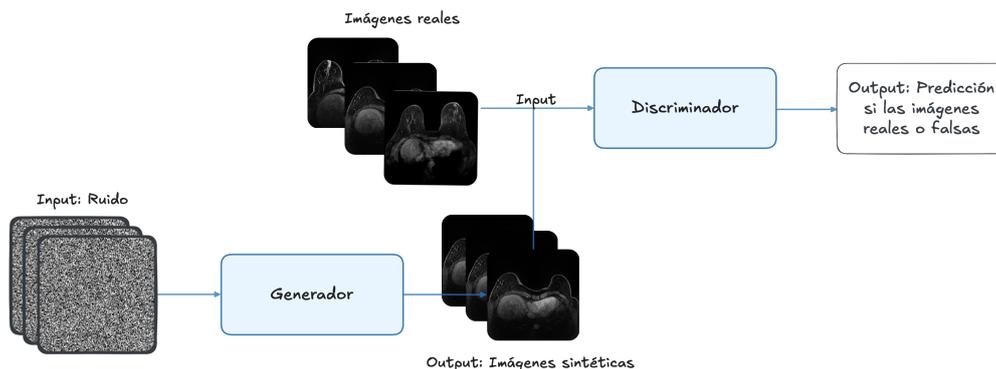


Figura 2.1: Esquema de la arquitectura GAN. Está compuesta por un generador, que recibe un vector de ruido como entrada y produce imágenes sintéticas, y un discriminador, que intenta distinguir entre imágenes reales y generadas.

En este trabajo nos hemos centrado en las GANs por la limitación en los datos médicos, la posibilidad de incorporar funciones de pérdida y, sobre todo, porque en

los modelos *diffusion* y *autoencoders* el entrenamiento suele ser computacionalmente más costoso.

A continuación se describen algunas variantes de las GANs aplicadas al procesamiento de imágenes médicas.

Pix2Pix El objetivo en esta tarea de *image to image traslation* consiste en aprender el mapeo de una imagen de entrada x y su correspondiente imagen de salida y , utilizando un conjunto de imágenes alineadas (véase [Ecuación 2.1](#)).

$$G : x \rightarrow y \quad (2.1)$$

Para ello, en la arquitectura de un Pix2Pix (véase [Figura 2.2](#)) se combina una pérdida adversarial, \mathcal{L}_{GAN} , ([Ecuación 2.2](#)), que fuerza al generador G a producir imágenes realistas que el discriminador D no pueda distinguir, y una pérdida de reconstrucción, \mathcal{L}_{L1} , ([Ecuación 2.3](#)), que penaliza las diferencias entre la imagen generada y la imagen real. La combinación de ambas da lugar a la pérdida total del modelo, $\mathcal{L}_{\text{Pix2Pix}}$, ([Ecuación 2.4](#)).

$$\mathcal{L}_{\text{GAN}}(G, D) = \mathbb{E}_y[\log D(y)] + \mathbb{E}_x[\log(1 - D(G(x)))] \quad (2.2)$$

$$\mathcal{L}_{\text{L1}}(G) = \mathbb{E}_{x,y}[\|y - G(x)\|_1] \quad (2.3)$$

$$\mathcal{L}_{\text{Pix2Pix}}(G, D) = \mathcal{L}_{\text{GAN}}(G, D) + \lambda \mathcal{L}_{\text{L1}}(G), \quad (2.4)$$

donde:

- x : imagen de entrada del dominio origen.
- y : imagen real correspondiente del dominio destino.
- G : generador que traduce imágenes x a $G(x)$.
- D : discriminador que distingue si una imagen es real, y , o generada, $G(x)$.
- $D(y)$: probabilidad de que y sea una imagen real.
- $\|y - G(x)\|_1$: error absoluto medio entre la imagen real y y la imagen generada $G(x)$.
- λ : peso de la pérdida \mathcal{L}_{L1} .

En general, la arquitectura Pix2Pix está compuesta por un generador tipo U-Net y un discriminador tipo PatchGAN [[23](#)].

Sin embargo, esto requiere que las imágenes de entrada y salida estén emparejadas, es decir, que exista una correspondencia directa entre los píxeles de ambas

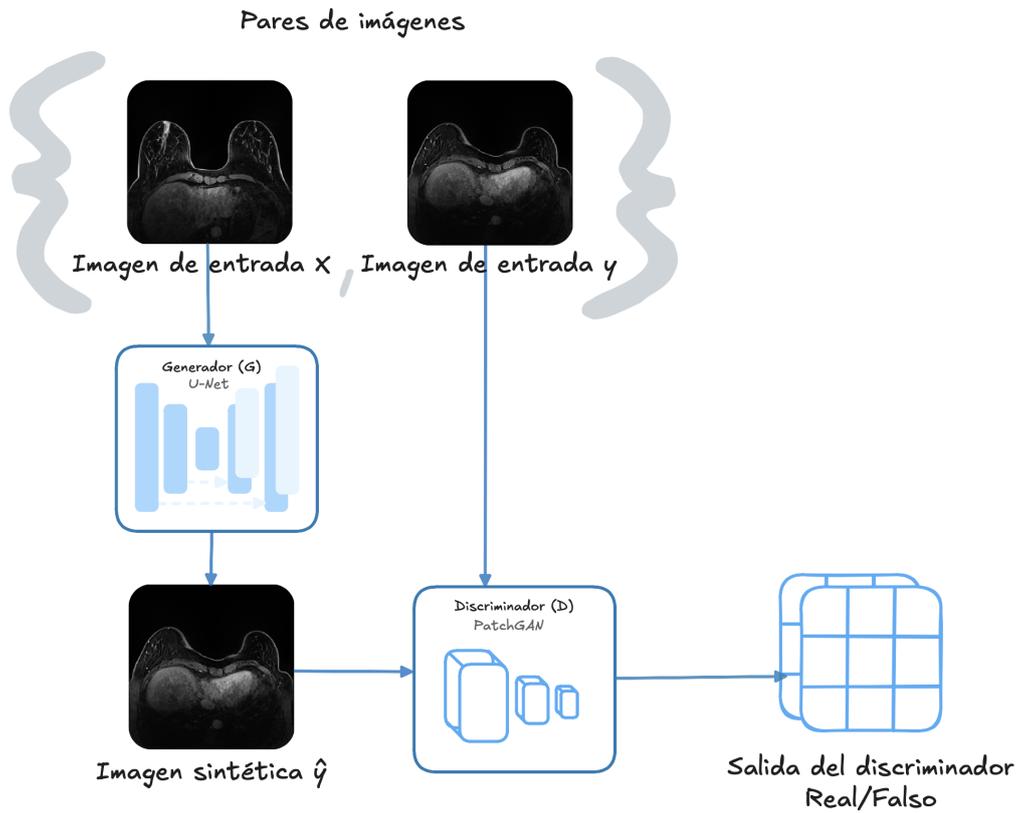


Figura 2.2: Esquema de la arquitectura Pix2Pix. A partir de una imagen de entrada x y su correspondiente objetivo y

imágenes. En el ámbito médico, esta condición rara vez se cumple, debido a la variabilidad anatómica entre pacientes y a la dificultad de obtener pares de imágenes clínicamente equivalentes.

A pesar de las limitaciones, algunos estudios recientes han demostrado el potencial de este en tareas de síntesis de imágenes médicas emparejadas. En el trabajo de Garrucho et al. (2024) [26], se entrenó un modelo Pix2Pix para traducir imágenes de MRI con y sin supresión de grasa, con el fin de abordar la heterogeneidad de los datos médicos (mencionada en la [Subsección 2.1](#)). Para alinear los datos, se aplica un registro rígido-afín entre ambos tipos de imágenes del mismo paciente. Posteriormente, le ha permitido realizar una segmentación automática de tumores en mama. Este enfoque fue efectivo para imágenes DCE-MRI emparejadas, alcanzando resultados prometedores en métricas de evaluación de la calidad de las imágenes generadas como PSNR (*Peak Signal-to-Noise Ratio*) con valor de 25.58 y SSIM (*Structural Similarity Index Measure*) de 0.75.

Redes Generativas Antagónicas Cíclicas En las imágenes médicas, es común que los conjuntos de datos disponibles sean no emparejados, es decir, que no exista una correspondencia directa entre las imágenes del dominio origen y las del dominio objetivo. En particular, en las imágenes DCE-MRI de la región mamaria, esta situación es común debido a la variabilidad en los protocolos de adquisición entre

distintos centros y fabricantes de escáneres [12, 13].

Para abordar este problema, las CycleGAN permiten realizar traducciones entre dominios sin requerir pares de imágenes alineadas, introduciendo una forma de supervisión débil o *weakly supervised* [27].

En CycleGAN, se asume que existe una relación subyacente entre los dos dominios, origen y objetivo, y se aprenden dos funciones de traducción:

$$G : X \rightarrow Y \quad \text{y} \quad F : Y \rightarrow X, \quad (2.5)$$

donde G traduce de un dominio X a un dominio Y , y F realiza la traducción inversa. El objetivo es que ambas funciones sean aplicaciones biyectivas, es decir, inyectivas (sin colisiones en la salida) y exhaustivas (cubren completamente el co-dominio).

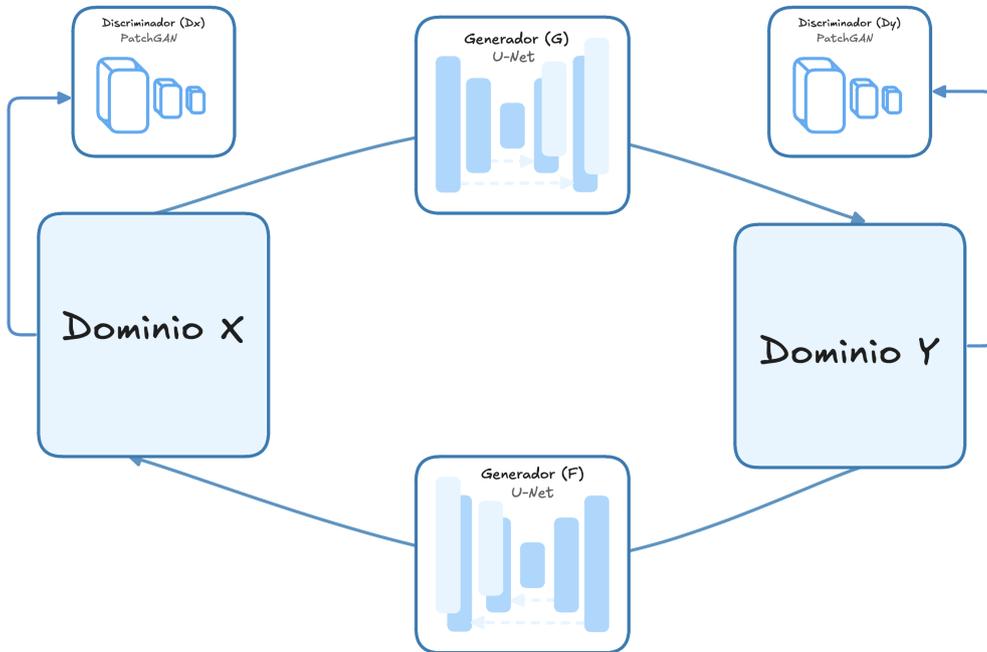


Figura 2.3: Esquema de la arquitectura CycleGAN. A partir dos dominios X e Y .

La arquitectura (véase [Figura 2.3](#)) está formada por dos generadores, G y F , y dos discriminadores, (D_Y y D_X). Los discriminadores determinan si una imagen pertenece al dominio real o ha sido generada. De forma similar a Pix2Pix, en CycleGAN la función de pérdida total se compone de tres términos principales: la pérdida adversarial \mathcal{L}_{GAN} ([Ecuación 2.6](#)), la pérdida de consistencia cíclica \mathcal{L}_{cyc} ([Ecuación 2.7](#)), que garantiza que una traducción de ida y vuelta recupere la imagen original, y la pérdida de identidad $\mathcal{L}_{identity}$ ([Ecuación 2.8](#)). Esta última es especialmente importante en contextos no emparejados, donde no se dispone de una correspondencia exacta entre dominios, ya que desalienta que los generadores modifiquen características irrelevantes de la imagen, como el color o la iluminación. La pérdida total del modelo, $\mathcal{L}(G, F, D_X, D_Y)$, queda definida en la [Ecuación 2.9](#).

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x)))] \quad (2.6)$$

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1] \quad (2.7)$$

$$\mathcal{L}_{\text{identity}}(G, F) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(y) - y\|_1] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(x) - x\|_1] \quad (2.8)$$

$$\begin{aligned} \mathcal{L}(G, F, D_X, D_Y) &= \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ &\quad + \lambda_{\text{cyc}} \mathcal{L}_{\text{cyc}}(G, F) + \lambda_{\text{id}} \mathcal{L}_{\text{identity}}(G, F), \end{aligned} \quad (2.9)$$

donde:

- $x \sim p_{\text{data}}(x)$: muestras reales del dominio X .
- $y \sim p_{\text{data}}(y)$: muestras reales del dominio Y .
- G : generador que traduce de X a Y .
- F : generador que traduce de Y a X .
- D_Y : discriminador del dominio Y , distingue y de $G(x)$.
- D_X : discriminador del dominio X , distingue x de $F(y)$.
- λ_{cyc} : peso que regula la importancia de la pérdida de ciclo.
- λ_{id} : peso que regula la pérdida de identidad.

Por ejemplo, en el estudio de Modanwal et al. (2021) [28], aborda las variaciones en la intensidad y distribución del ruido que presentan las imágenes DCE-MRI obtenidas de diferentes escáneres con una CycleGAN. La CycleGAN aprende una correspondencia bidireccional entre imágenes de diferentes escáneres sin necesidad de datos emparejados, permitiendo así la normalización de imágenes en entornos clínicos heterogéneos. También experimentan con la adición de nuevas pérdidas a la función de coste total para mantener la estructura de la región mamaria. Se mejoró la consistencia morfológica, alcanzando un coeficiente de Dice -métrica que mide la superposición entre dos regiones- de hasta 0.98 para la máscara del seno tras la normalización.

Alternativamente, Muramatsu et al. (2020) [29] utilizaron una CycleGAN como *data augmentation*, transformando nódulos pulmonares en imágenes sintéticas de masas mamarias con el fin de mejorar la clasificación entre lesiones benignas y malignas en mamografías. Entrenaron redes convolucionales (ResNet-50) con combinaciones de datos reales, aumentados y sintéticos, alcanzando una precisión de hasta 81.4% y un AUC (*total area under the receiver operating characteristic (ROC) curve*) de 0.884.

Redes Generativas Antagónicas con Clasificación Auxiliar Las Redes Generativas Antagónicas con Clasificación Auxiliar (*Auxiliary Classifier Generative Adversarial Network*, AC-GAN) añaden una salida adicional al discriminador que permite predecir la clase de la imagen, además de distinguir entre imágenes reales y generadas. De este modo, el discriminador aprende dos funciones simultáneamente: la fuente de la imagen y su clase [30]. Esta arquitectura es adecuada para tareas de clasificación con datos escasos, como ocurre en las imágenes médicas, permitiendo incorporar información diagnóstica directamente en el proceso generativo.

2.3. Predicción de respuesta patológica completa (PCR) en cáncer de mama

Tras revisar la literatura sobre la predicción de la PCR en cáncer de mama, se han identificado dos desafíos principales (expuestos en la [Subsección 2.1](#)): la disponibilidad y la heterogeneidad de los datos etiquetados.

Por un lado, muchas investigaciones se enfrentan a una limitada disponibilidad de datos, restringiendo su capacidad de generalizar. En algunos casos, los conjuntos están formados por solo 95 pacientes [31]. Otros solucionan estas restricciones mediante el uso de datos longitudinales, es decir, imágenes del mismo paciente obtenidas, por ejemplo, en diferentes fases del tratamiento. Investigaciones como Syed et al. [32] y Huang et al. [33] desarrollaron modelos que combinan imágenes pre-NACT (antes de la NACT), post-NACT (después de la NACT) y la diferencia entre ambas, mejorando los resultados en comparación con los modelos entrenados en una única fase. Huang et al. reportaron un AUC promedio de 0.90 para su modelo combinado, frente a 0.78–0.84 en los modelos unifase.

Otro aspecto relacionado con la disponibilidad de datos es el desequilibrio de clases, habitual en tareas de predicción de la PCR, donde la clase positiva suele estar infrarrepresentada. Por ejemplo, Yang et al. [31] en los datos usados solo el 25 % alcanzó PCR, lo que puede introducir sesgo y afectar el rendimiento del modelo en pacientes minoritarios. Para abordar este problema, diversas investigaciones han implementado técnicas de aumento de datos y estrategias específicas para equilibrar las clases. Por ejemplo, Sutton et al. (2020) [34] para generar muestras sintéticas de la clase minoritaria, PCR1, en conjuntos de datos de RMI utilizó la técnica *Synthetic Minority Over-sampling Technique* o Escobar et al. (2024) [35] utilizó una GAN para reducir el desequilibrio de clases en imágenes histopatológicas.

Finalmente, la interpretabilidad de los modelos desde un punto de vista clínico resulta clave para su integración en la práctica médica. En este sentido, Mao et al. (2025) [36] propusieron el uso de mapas de atención en redes convolucionales para identificar visualmente las regiones más relevantes de la imagen en la predicción de PCR, lo que facilita la asociación con cambios biológicos.

A continuación se citan algunas investigaciones realizadas sobre imágenes DCE-MRI para predecir la PCR en pacientes con cáncer de mama:

- Mao et al. (2025) [36]: Desarrollaron un sistema multimodal completamen-

te automático (*Multimodal Integrated Fully Automated Pipeline System*, MI-FAPS) que integra DCE-MRI pre-NACT, imágenes histopatológicas y datos clínicos para predecir la PCR en un conjunto de datos recopilado en múltiples centros con 1004 pacientes. Alcanzaron una AUC de 0.882 en test externo y 0.909 en una cohorte prospectiva, superando modelos unimodales.

- Chang et al. (2024) [37]: Entrenaron una CNN 3D con DCE-MRI post-NACT en 852 pacientes HER2+ o triple negativo. Al usar regiones tumorales recortadas, lograron AUC=0.74, y observaron un descenso sustancial del rendimiento al usar imágenes completas (AUC 0.45–0.54), subrayando la importancia de centrarse en la lesión.
- Syed et al. (2023) [32]: Aplicaron XGBoost (*eXtreme Gradient Boosting*) sobre características de textura extraídas de DCE y imágenes por difusión (DWI, por sus siglas en inglés) en 117 pacientes (antes, durante y después de NACT), junto con datos clínicos. Obtuvieron AUC hasta 0.951 combinando todas las modalidades, destacando el valor de datos longitudinales y multimodales.
- Xu et al. (2025) [38]: Estudiaron el cáncer de mama triple negativo usando un modelo 3D ResNeXt50 con DCE y DWI basales más datos clínicos en 282 pacientes. Lograron AUC de 0.76 en test interno, y AUC 0.72 en test externo, evaluando distintas variantes de preprocesamiento.
- D’Anna et al. (2025) [39]: Usaron radiómica clásica en DCE-MRI con 55 pacientes. Extrajeron 1446 características y aplicaron ComBat para corregir el sesgo por escáneres. Random Forest logró AUC de 0.98 ± 0.06 . Aunque es una muestra pequeña, muestra el potencial de técnicas tradicionales bien preprocesadas.
- Lv et al. (2025) [40]: Compararon arquitecturas profundas aplicando transferencia de aprendizaje sobre DCE-MRI basal en 313 pacientes. Tras extraer 256 características de la capa *fully-connected* (FC), usaron *support vector machine* SVM y XGBoost, siendo ResNet18 la más eficaz (AUC 0.87).

Capítulo 3

3. Materiales y metodología

3.1. Materiales

En esta sección describiremos los recursos utilizados, que incluyen los datos clínicos disponibles, el proceso de segmentación y las herramientas computacionales.

3.1.1. Datos clínicos y de imagen

Los datos utilizados proceden del conjunto de datos de MAMA-MIA [41], que integra 1506 casos de DCE-MRI en pacientes con cáncer de mama tratados con quimioterapia neoadyuvante. El conjunto unió varios estudios públicos (ISPY1 [42], ISPY2 [43], Duke [44] y NACT [45]), e incluye segmentaciones expertas del tumor primario.

Las imágenes DCE-MRI fueron adquiridas antes del tratamiento e incluyen volúmenes tridimensionales obtenidos en distintas fases, antes y después de la inyección del agente de contraste. Los criterios de selección de MAMA-MIA [41] incluyeron casos previos al tratamiento, y para los cuales se disponía de información clínica, como la PCR o información de supervivencia a cinco años [46].

Además, el conjunto proporciona un archivo XLSX, *Excel Open XML Spreadsheet*, (véase Figura 3.1) que contiene 50 variables clínicas y demográficas armonizadas, entre ellas la PCR, información sobre el subtipo tumoral y estado de supervivencia.

	patient_id	dataset	bilateral_breast_cancer	multifocal_cancer	...	tcia_series_uid
0	DUKE_001	DUKE	0	0.0	...	1.3.6.1.4.1.14519.5.2.1.1857778498036652445367...
1	DUKE_002	DUKE	0	0.0	...	1.3.6.1.4.1.14519.5.2.1.2920879504444133146439...
2	DUKE_005	DUKE	0	1.0	...	1.3.6.1.4.1.14519.5.2.1.3082672827908226352755...
3	DUKE_009	DUKE	0	1.0	...	1.3.6.1.4.1.14519.5.2.1.1955995935901349215652...
4	DUKE_010	DUKE	0	1.0	...	1.3.6.1.4.1.14519.5.2.1.1390661710325525999628...

Figura 3.1: Archivo que contiene los datos clínico de cada paciente.

Análisis exploratorio de las variables clínicas y radiológicas Con el objetivo de comprender mejor la composición del conjunto de datos, se realizó un análisis exploratorio de las variables clínicas disponibles. Se distinguieron las variables categóricas, 33, de las numéricas, 17, y se representaron gráficamente.

En primer lugar, para las variables categóricas se observa que aproximadamente el 29.2% de las pacientes alcanzaron una PCR (PCR = 1.0), mientras que el resto presentó enfermedad residual (PCR = 0) (véase Figura 3.2).

Para variables relacionadas con las características técnicas de los escáneres y la procedencia de los datos, se observan tres fabricantes, siendo Genreal Electric (GE)

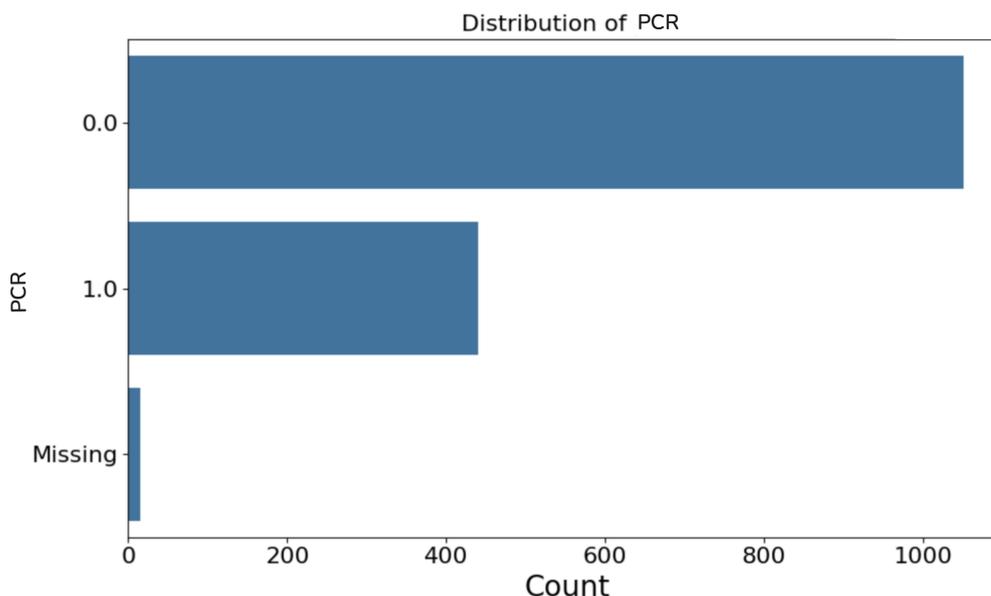


Figura 3.2: Distribución de casos etiquetados sin PCR (0) y con la PCR (1) dentro del MAMA-MIA.

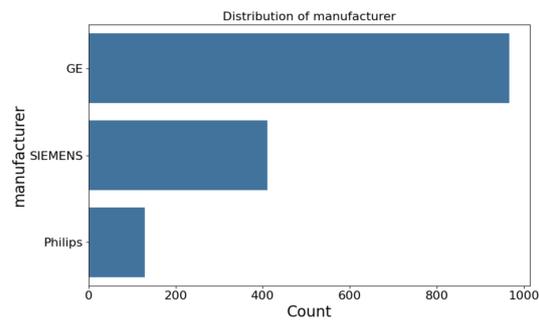
el más frecuente, presente en más del 60 % de los casos, seguidos por Siemens y Philips. En total hay 22 modelos distintos, los principales: Signa HDxt (38.2 %), Avanto (12.4 %) y GENESIS-SIGNA (10.7 %). En cuanto al campo magnético, predomina de 1.5T (72 %), mientras que el resto corresponde a equipos de 3T. Respecto al origen de los datos, la mayor parte proviene del estudio ISPY2 [43], con 980 casos, seguido de DUKE [44] con 291, ISPY1 [42] con 171 y NACT [45] con 64.

Para las características demográficas, la mayoría de los pacientes son caucásicos (74.9 %), seguidos por afroamericanos (16 %) y asiáticos (5.7 %). El resto de las categorías étnicas tienen una representación mínima. Respecto al índice de masa corporal (BMI, por sus siglas en inglés), la mayoría de los casos se agrupan dentro de las categorías normal y sobrepeso, seguidos por pacientes con obesidad de tipo 1. Además, se observan valores faltantes con cierta frecuencia en esta variable (véase Figura 3.4).

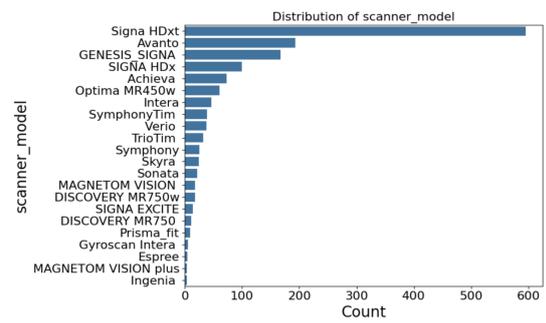
Para las características anatómicas e imagenológicas, se observa que la gran mayoría de las imágenes fueron adquiridas con supresión de grasa. Solo en el 31 % de los pacientes se incluyen imágenes de ambas mamas, lo que indica que la mayoría de las adquisiciones son unilaterales. Tanto la presencia de implantes mamarios (30 pacientes) como el cáncer bilateral (también presente en 30 casos) tienen una baja frecuencia dentro del conjunto de datos (véase Figura 3.5).

En cuanto a las características tumorales, el subtipo más frecuente es el *triple negative*, considerado el más agresivo, con 499 casos. Le siguen el *luminal a* con 381, *luminal* con 211 y *her2 enriched* con 169, *luminal b* con 155 y *her2 pure* con 65. Respecto a la presencia de cáncer multifocal ¹, solo el 35.8 % lo presenta, mientras

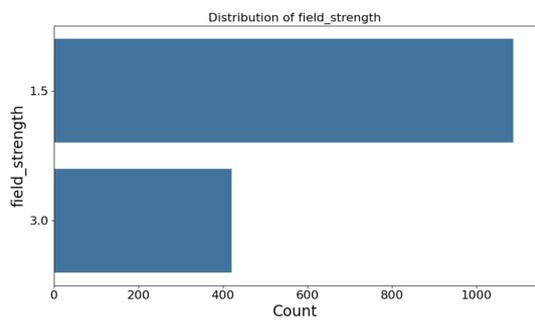
¹El cáncer multifocal se refiere a la presencia de 2 o más focos tumorales en un mismo cuadrante y a menos de 5 centímetros del foco primario.



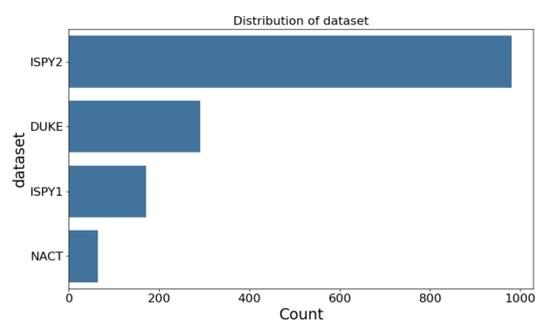
(a)



(b)

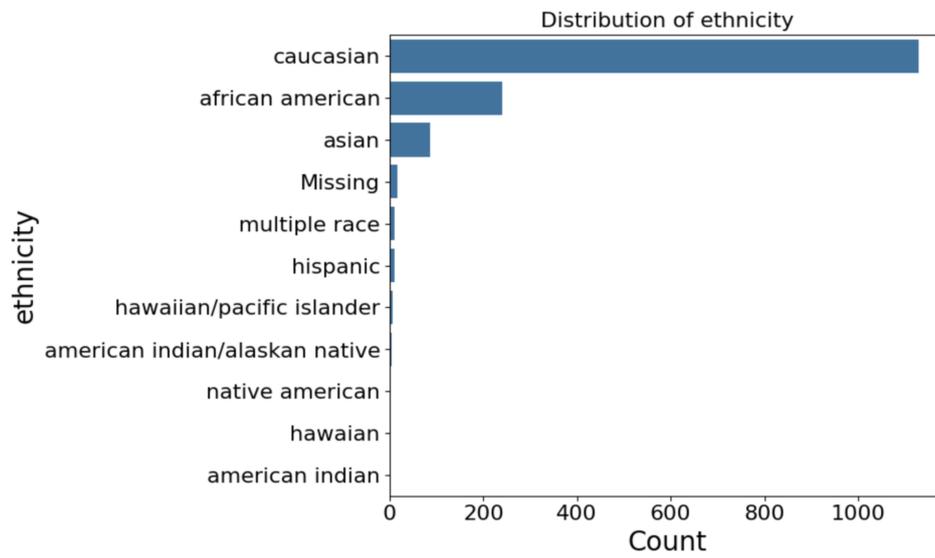


(c)

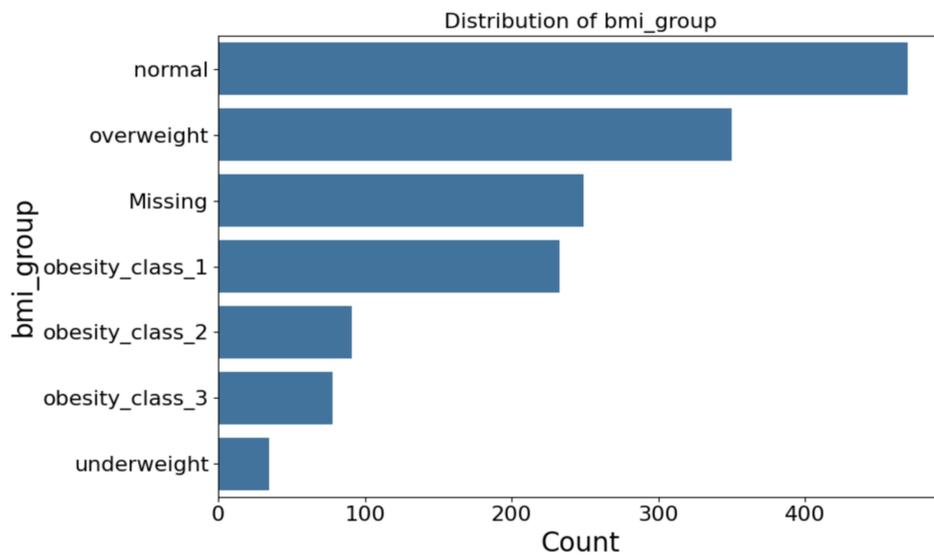


(d)

Figura 3.3: Distribución de las características técnicas de los escáneres y la procedencia de los datos: (a) fabricante del escáner, (b) modelo, (c) campo magnético y (d) origen del conjunto de datos.

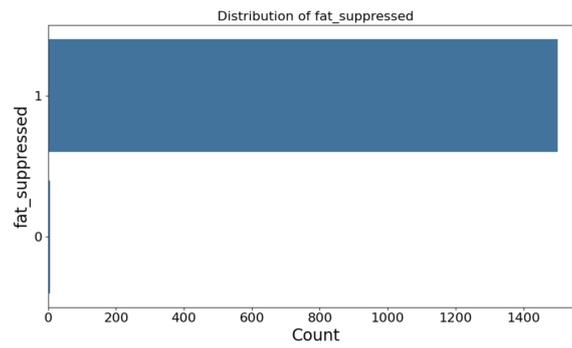


(a)

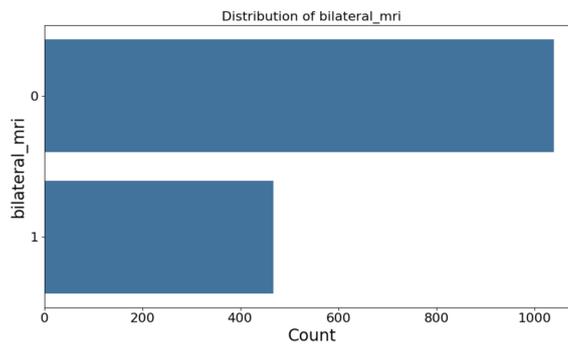


(b)

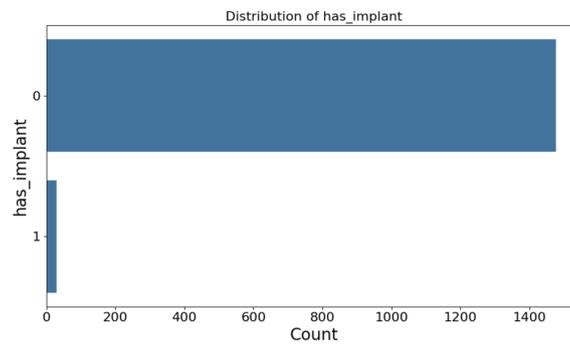
Figura 3.4: Distribución de las características (a) perfil étnico y (b) BMI.



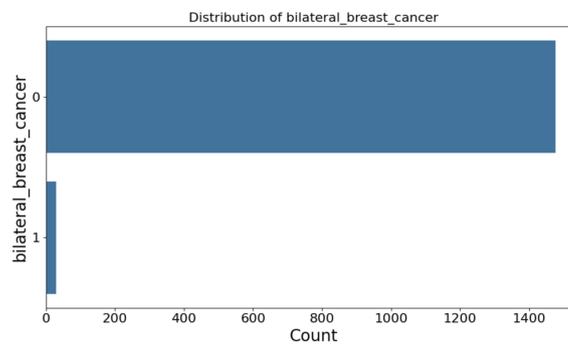
(a)



(b)



(c)



(d)

Figura 3.5: Distribución de las características anatómicas e imagenológicas: (a) supresión de grasa, (b) imagen bilateral, (c) presencia de implantes mamarios y (d) cáncer bilateral.

que en más del 50 % no se detecta (véase [Figura 3.6](#)).

En segundo lugar, al analizar las variables numéricas relacionadas con eventos a lo largo del tiempo, se observa una disponibilidad limitada. Específicamente, los días de seguimiento están registrados para 517 pacientes (34.3 %), la recurrencia tumoral para 523 (34.7 %), la metástasis para solo 64 pacientes (4.3 %) y el fallecimiento para 459 pacientes (30.5 %). Los datos de metástasis no se analizan porque no son extrapolables al resto de pacientes al ser una muestra demasiado pequeña (véase la [Figura 3.7](#)).

Para la variable correspondiente a los días de seguimiento, se observa una distribución asimétrica con un sesgo a la derecha: la mayoría de los pacientes tienen seguimientos inferiores a 2000 días (aproximadamente 5.5 años). Aunque existen casos con valores cercanos a 10 años. En contraste, las variables de días hasta la recurrencia y días hasta el fallecimiento presentan distribuciones fuertemente sesgadas hacia valores bajos, con una gran concentración de casos en cero.

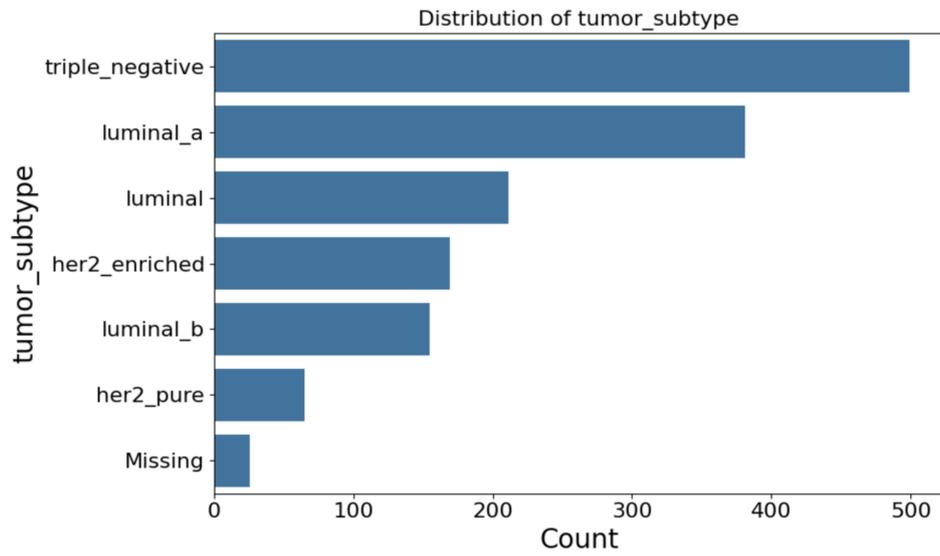
En las variables demográficas -edad, peso y estatura- (véase [Figura 3.8](#)), se observa que la mayoría de los pacientes tienen una edad alrededor de los 50 años. El peso se concentra principalmente entre 55 kg y 90 kg, mientras que la estatura muestra una distribución más discreta, con picos evidentes en torno a 1.60 m, 1.65 m y 1.70 m. Estas distribuciones individuales son coherentes con lo observado anteriormente en el BMI, donde predominan las categorías normal y sobrepeso (véase [Figura 3.4](#)).

En tercer lugar, los parámetros técnicos de las imágenes DCE-MRI muestran que el máximo número de fases varía entre un mínimo de 3 y un máximo de 6. La mayoría de las imágenes (916 casos) presentan dimensiones de 512×512 píxeles (filas y columnas), aunque también se encuentran resoluciones inferiores como 256×256 (251 casos). Para la distribución del número de cortes por volumen es más dispersa, con modos en torno a 150 y 160 cortes (véase la [Figura 3.9](#)).

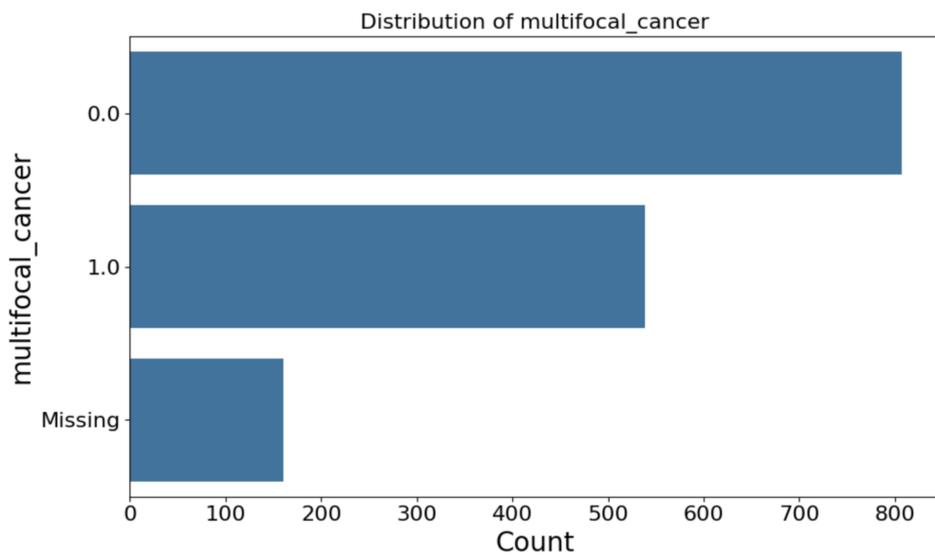
Finalmente, se analizaron también variables relacionadas con la adquisición de las imágenes médicas. En particular, se evaluó la distribución de la vista de adquisición y la orientación anatómica del escaneo, ambas categóricas (véase la [Figura 3.10](#)).

Se observa que la gran mayoría de los estudios fueron adquiridos en vista axial, mientras que una menor proporción utiliza la vista sagital. En cuanto a la orientación anatómica, el 85 % de las imágenes están orientadas en formato LAI (*Left-Anterior-Inferior*), mientras que un pequeño porcentaje (alrededor del 15 %) corresponde a PSL (*Posterior-Superior-Left*).

En conclusión, el conjunto de datos presenta una heterogeneidad, reflejada en aspectos como los centros de origen, modelos de escáner empleados, campo magnético y características técnicas de las imágenes. Esta variabilidad puede influir tanto en la capacidad de generalización como en la robustez del modelo. También, la representación de las distintas etnias desequilibrada puede introducir sesgos étnicos en los modelos.

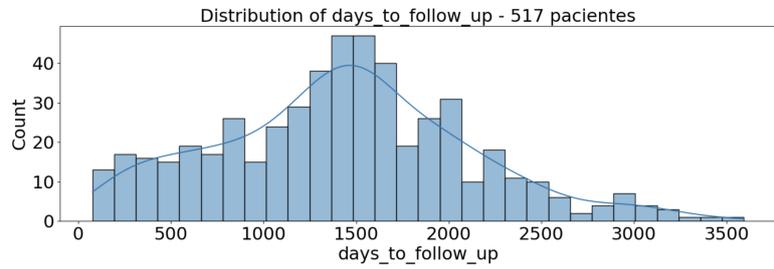


(a)

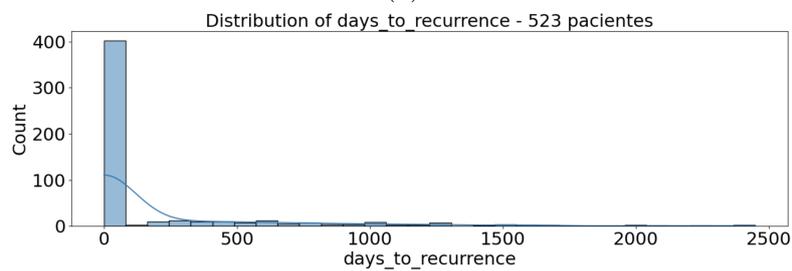


(b)

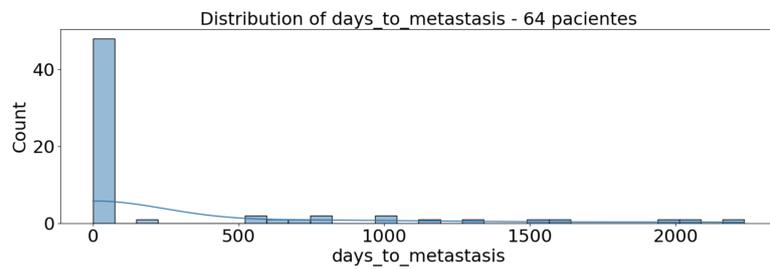
Figura 3.6: Distribución de las características tumorales: (a) subtipo de tumor y (b) cáncer multifocal.



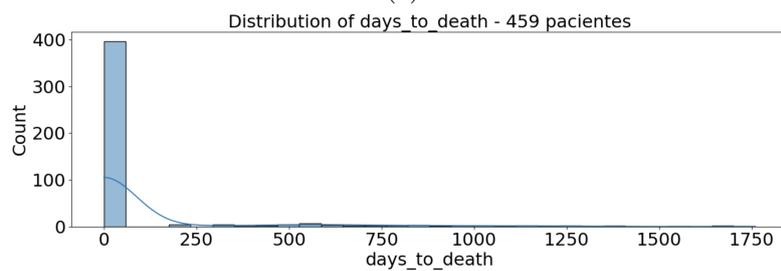
(a)



(b)

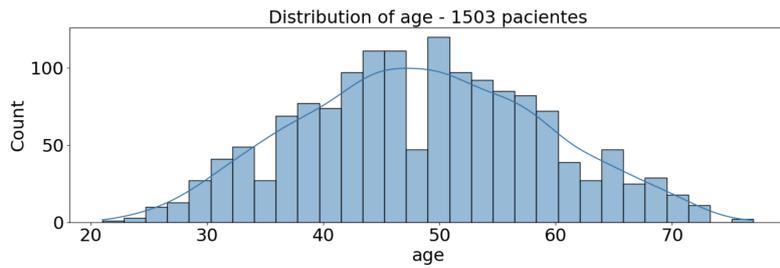


(c)

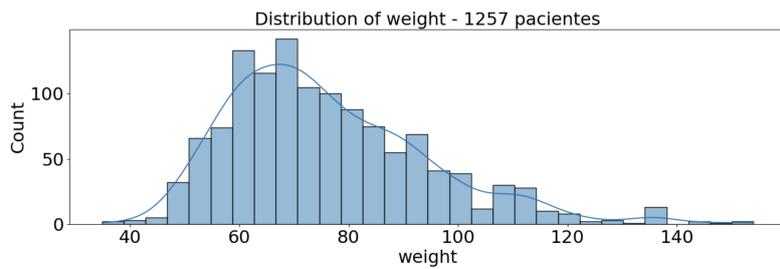


(d)

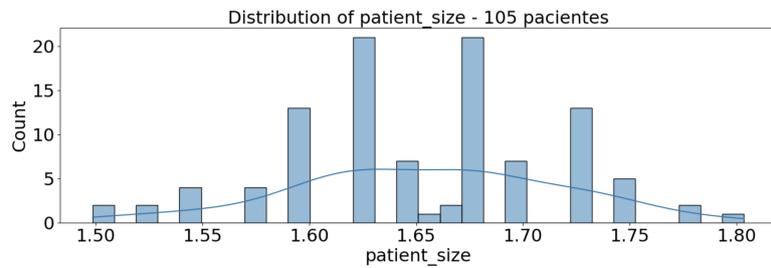
Figura 3.7: Distribución de las características relacionadas con eventos a lo largo del tiempo: (a) días del seguimiento, (b) recurrencia, (c) metástasis y (d) fallecimiento.



(a)

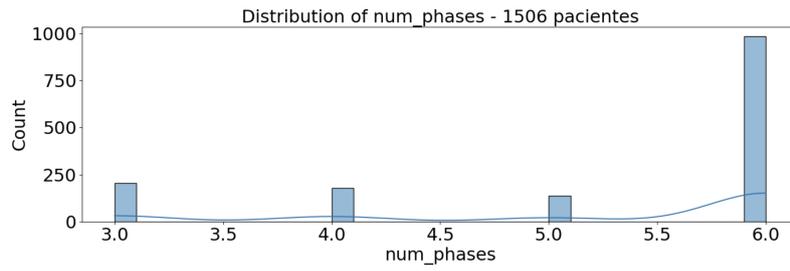


(b)

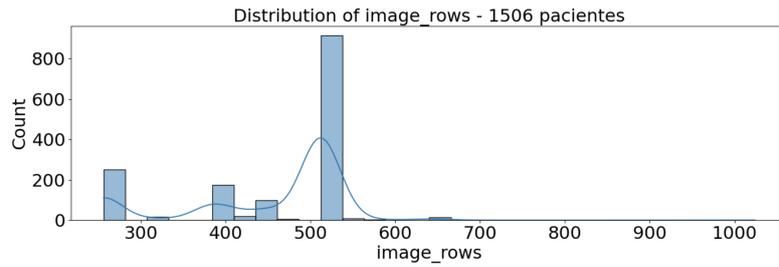


(c)

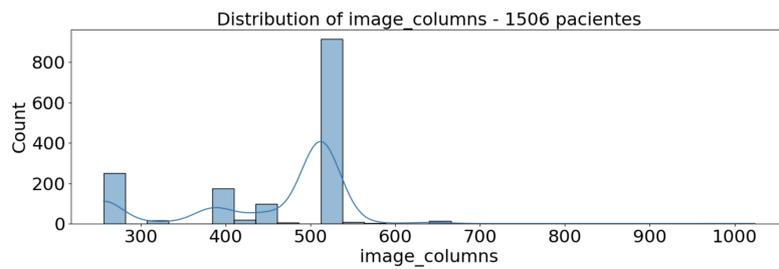
Figura 3.8: Distribución de las características demográficas: (a) edad, (b) peso y (c) altura.



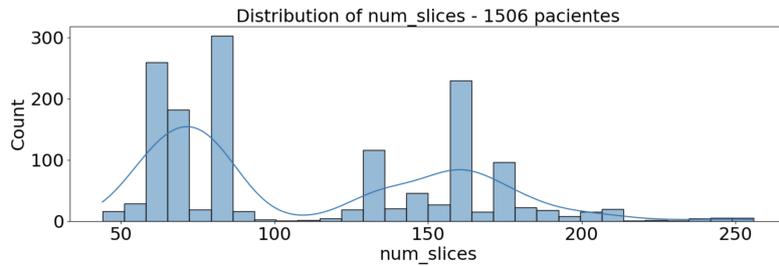
(a)



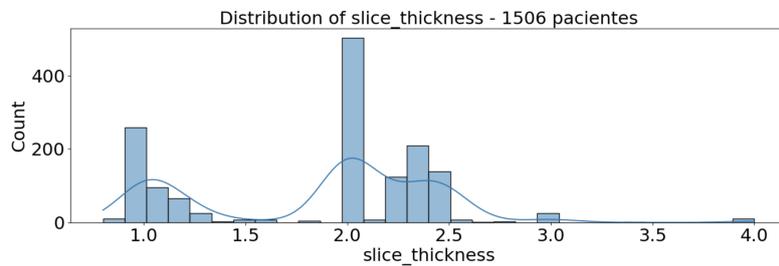
(b)



(c)

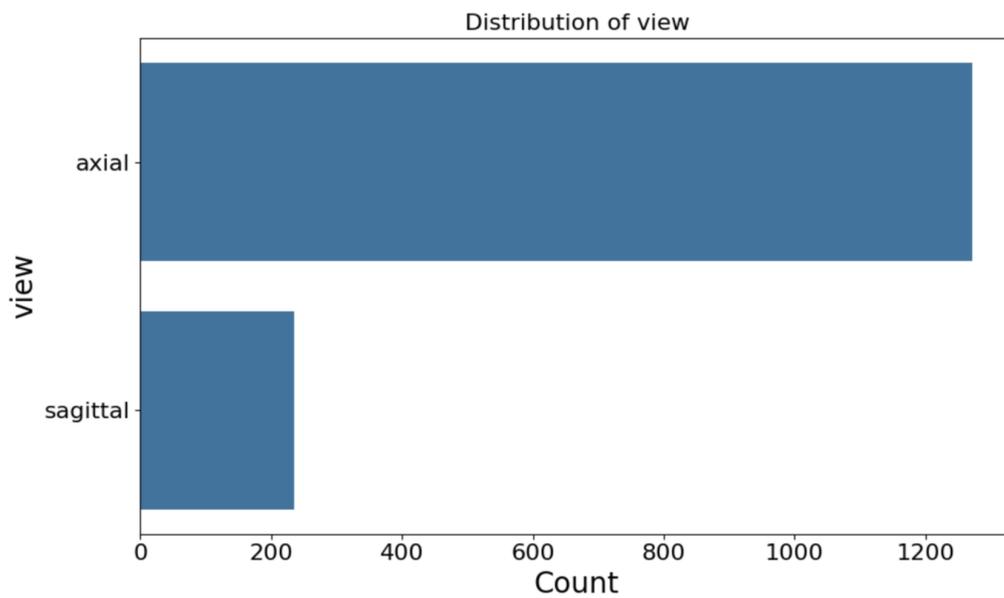


(d)

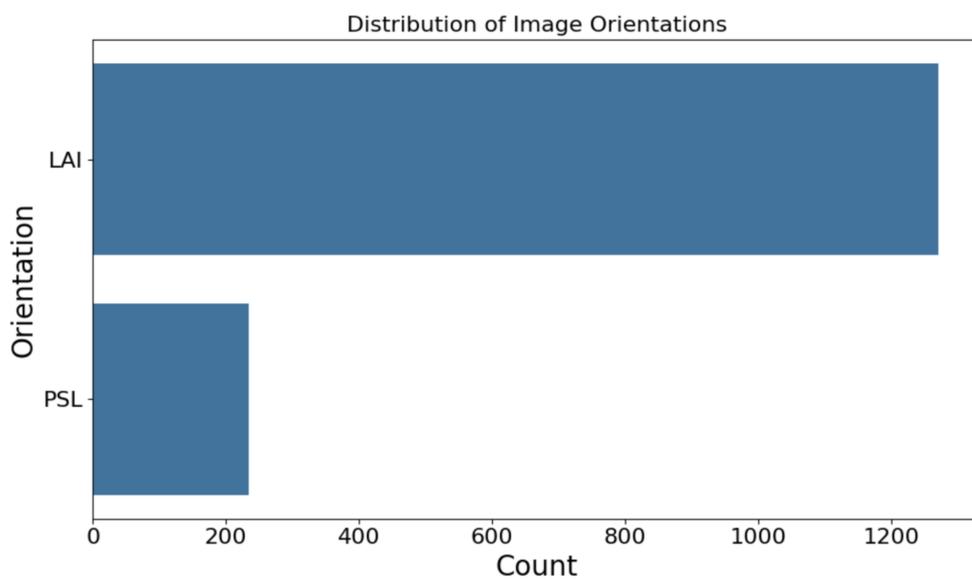


(e)

Figura 3.9: Distribución de las características de las DCE-MRI: (a) número de fases, (b) filas, (c) columnas, (d) cortes y (e) grosor del corte.



(a)



(b)

Figura 3.10: Distribución de la (a) vista de adquisición y la (b) orientación anatómica en las imágenes DCE-MRI del conjunto MAMA-MIA.

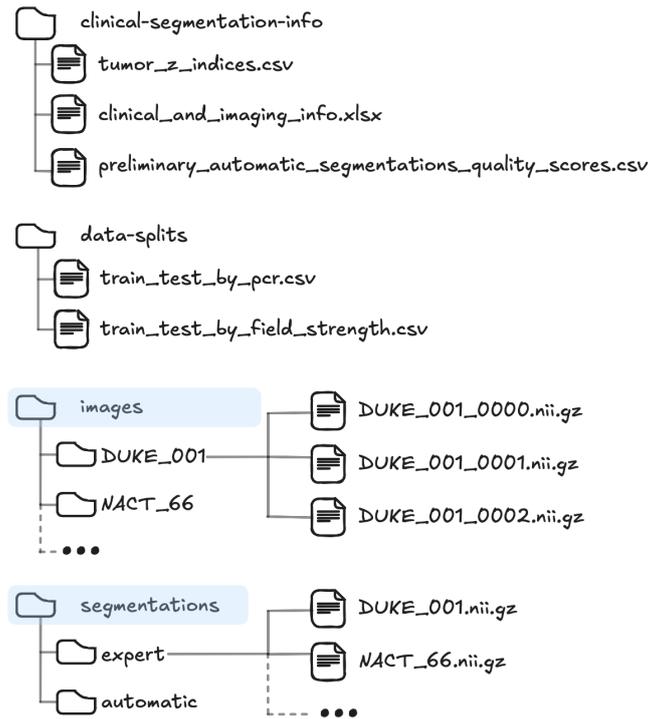


Figura 3.11: Estructura de directorios de los datos utilizados.

3.1.2. División del conjunto de datos

Este trabajo parte de un conjunto de imágenes DCE-MRI en formato `.nii.gz`, que constituyen los datos base del proyecto. Con el objetivo de garantizar la escalabilidad y facilitar la automatización del *pipeline*, se han creado distintos archivos en formato *comma-separated values* (CSV, por sus siglas en inglés) que definen la división del conjunto de datos en entrenamiento y test.

Cada archivo CSV contiene ocho columnas: cuatro correspondientes al conjunto de entrenamiento y cuatro al de test (véase [Figura 3.11](#)). En estos se incluyen los identificadores de los pacientes -que coinciden con el nombre de las carpetas- y los índices del eje Z utilizados para extraer los cortes axiales (véase [Figura 3.12](#)). En los conjuntos de entrenamiento y test se mantiene la misma distribución de las características clínicas.

Además, para la clasificación explícita de la PCR, se incorporan columnas adicionales con las etiquetas correspondientes a cada paciente. Estas etiquetas permiten distinguir entre pacientes con PCR1 y PCR0, tanto para el entrenamiento como para test (véase [Figura 3.13](#)).

Como se ha mencionado, los archivos CSV anteriores incluyen una columna con los índices del eje Z por paciente, que indica la coordenada axial donde se encuentra el tumor. Esta información se extrae de un archivo adicional denominado `tumor_z_indices.csv` (véase [Figura 3.14](#)). Este archivo contiene los valores del eje Z en los que el tumor está presente, garantizando así que los cortes seleccionados

	train_A_id	train_A_z_index	train_B_id	train_B_z_index	test_A_id	test_A_z_index	test_B_id	test_B_z_index
0	DUKE_001	39	DUKE_005	35.0	DUKE_019	35.0	DUKE_044	43.0
1	DUKE_002	30	DUKE_005	40.0	DUKE_021	43.0	DUKE_044	40.0
2	DUKE_010	50	DUKE_009	32.0	DUKE_028	26.0	DUKE_055	26.0
3	DUKE_012	32	DUKE_009	37.0	DUKE_046	37.0	DUKE_055	27.0
4	DUKE_022	22	DUKE_045	27.0	DUKE_057	30.0	DUKE_099	34.0

Figura 3.12: Archivo `train_test_by_pcr.csv` con ocho columnas: cuatro para el entrenamiento y cuatro para test. En ambos casos, los pacientes están divididos en dos dominios: A (PCR0) y B (PCR1).

	train_A_id	train_A_z_index	train_A_label	train_B_id	train_B_z_index	train_B_label	test_A_id	test_A_z_index	test_A_label	test_B_id	test_B_z_index	test_B_label
0	DUKE_001	39	0	DUKE_002	30	0	DUKE_019	35.0	0.0	DUKE_028	26.0	0.0
1	DUKE_005	35	1	DUKE_009	32	1	DUKE_021	43.0	0.0	DUKE_028	34.0	0.0
2	DUKE_012	32	0	DUKE_010	50	0	DUKE_044	43.0	1.0	DUKE_046	37.0	0.0
3	DUKE_022	22	0	DUKE_040	40	0	DUKE_057	30.0	0.0	DUKE_046	29.0	0.0
4	DUKE_032	43	0	DUKE_041	37	0	DUKE_069	41.0	0.0	DUKE_055	26.0	1.0

Figura 3.13: Archivo `train_test_by_field_strength.csv` con doce columnas: seis para entrenamiento y seis para test. En ambos casos, los pacientes están divididos en dos dominios según la intensidad del campo magnético del escáner: A (1.5T) y B (3T).

contengan tejido tumoral visible. Para la construcción del archivo se utilizan las máscaras tumorales, de forma que se calcula la suma de los valores de la máscara en cada plano axial; los cortes cuya suma sea distinta de cero -contienen tumor- se consideran relevantes.

	patient_id	z_index	tumor_content
0	DUKE_019	35	382.0
1	DUKE_019	27	128.0
2	DUKE_019	39	55.0
3	DUKE_019	26	25.0
4	DUKE_019	33	421.0

Figura 3.14: Archivo CSV que contiene tres columnas: el identificador del paciente, el índice en el eje Z y la cantidad de tumor presente en el corte, según la máscara de segmentación.

3.1.3. Lenguaje de programación y librerías

El código del proyecto se ha desarrollado en el lenguaje de programación Python (versión 3.9.21), utilizando el entorno de desarrollo Visual Studio Code. Las librerías empleadas, junto con sus versiones y funciones, se detallan en la [Tabla 3](#). Para la ejecución del modelo se ha utilizado una GPU T4 x 2 proporcionada por Kaggle [51].

Tabla 3: Librerías python utilizadas en el proyecto, junto con su versión y funcionalidad principal.

Librería	Versión	Funcionalidad
<code>tqdm</code>	4.67.1	Mostrar barras de progreso en bucles.
<code>numpy</code>	1.26.4	Manipulación de arrays.
<code>pandas</code>	2.2.3	Análisis y manipulación de datos tabulares (CSV, DataFrames).
<code>matplotlib</code>	3.9.4	Visualización de datos.
<code>SimpleITK</code>	2.5.0	Lectura, escritura y procesamiento de imágenes médicas.
<code>torch</code>	2.2.2	Framework para crear redes neuronales profundas (PyTorch).
<code>torchio</code>	0.20.8	Preprocesamiento y augmentación de imágenes médicas 3D.
<code>scikit-learn</code>	1.6.1	Algoritmos y herramientas para aprendizaje automático, como TSNE para reducción de dimensionalidad.
<code>plotly</code>	6.0.1	Visualización interactiva de gráficos.
<code>ipywidgets</code>	8.1.6	Widgets interactivos para notebooks Jupyter.
<code>IPython.display</code>	8.18.1	Funcionalidades para mostrar widgets.
<code>skimage</code>	0.24.0	Procesamiento de imágenes, detección de contornos, regiones, etc.
<code>seaborn</code>	0.13.2	Visualización estadística basada en matplotlib, especialmente para correlaciones y distribuciones.

3.2. Metodología

En esta sección se describen los diferentes procesos de segmentación. Posteriormente, se presentan los modelos desarrollados para la tarea de clasificación de imágenes.

3.2.1. Segmentación

La segmentación consiste en separar regiones de interés dentro de una imagen, permitiendo aislar estructuras específicas del fondo o de otros tejidos. En imágenes médicas, se utiliza para identificar y delimitar órganos, tejidos o lesiones, como tumores. Los mecanismos de segmentación pueden clasificarse en:

- **Manual:** Realizada por un experto, suele ser precisa pero requiere mucho tiempo. Esta sujeta a la variabilidad interobservador.
- **Automática:** Realizada por algoritmos sin intervención humana directa. Es

rápida y escalable, pero su precisión depende del método y la calidad de los datos.

- **Semiautomática:** Combina los algoritmos con la intervención del experto, buscando un equilibrio entre precisión y eficacia.

Como se mencionó en la [Subsubsección 3.1.1](#), se dispone de segmentaciones tumorales para cada paciente, obtenidas a través de un enfoque semiautomático que combina la intervención de un experto con la segmentación generada por un modelo U-Net.

En este trabajo, debido a ciertas limitaciones observadas en el desempeño del modelo en fases iniciales, se han explorado diversas técnicas de segmentación autónoma para diferenciar la región de la mama del fondo. A continuación, se describen estas técnicas:

Algoritmo de Otsu Es un método de segmentación automática basado en el umbral. El objetivo del algoritmo es determinar un umbral óptimo para separar la imagen en dos clases distintas. Para ello, analiza la distribución de los píxeles en el histograma de intensidades y selecciona el valor de umbral que minimiza la varianza dentro de cada clase o, equivalentemente, que maximiza la varianza entre clases (véase [Figura 3.15](#)).

La fórmula que define la varianza entre clases es:

$$\sigma_b^2(t) = P_0(t)P_1(t) [\mu_0(t) - \mu_1(t)]^2, \quad (3.1)$$

donde:

- $P_0(t)$ y $P_1(t)$: probabilidades de los píxeles que pertenecen a las clases 0 y 1, es decir, la proporción de píxeles con valores menores o mayores que el umbral t respectivamente.
- $\mu_0(t)$ y $\mu_1(t)$: medias de intensidades de cada clase, calculadas para los píxeles por debajo y por encima del umbral t .

El umbral óptimo t^* es aquel que maximiza $\sigma_b^2(t)$, la varianza entre clases [\[47\]](#). En el contexto de imágenes médicas, es útil para segmentar estructuras relevantes, como para el posterior análisis con recorte, que permita centrar el procesamiento en la región de interés.

Operaciones Morfológicas Son técnicas basadas en la teoría de conjuntos y la topología, aplicadas a imágenes binarias para refinar máscaras segmentadas mediante la modificación de su estructura. Las operaciones fundamentales son la dilatación y la erosión, que sirven como base para operaciones más complejas como la apertura y el cierre [\[48\]](#). En la [Figura 3.16](#) se muestra el efecto de las operaciones morfológicas para una imagen.

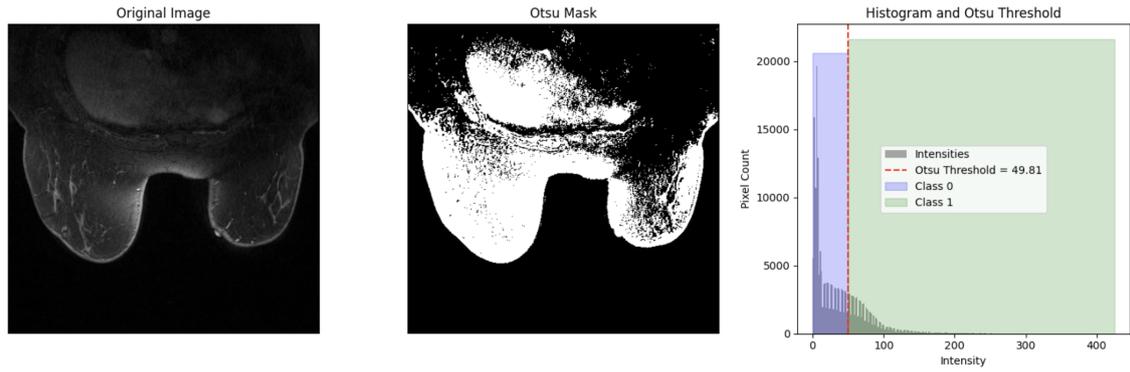


Figura 3.15: De izquierda a derecha: (1) imagen original en un corte axial del paciente DUKE_001, (2) máscara binaria resultante tras aplicar el umbral de Otsu, y (3) histograma de intensidades de la imagen con la línea de umbral óptimo marcada en rojo. Se observa cómo Otsu separa automáticamente las intensidades en dos clases basadas en su distribución.

Dilatación La dilatación expande las regiones blancas (valor 1) de una imagen binaria. Es útil para rellenar pequeños huecos en las máscaras o conectar regiones cercanas [49]. Se formaliza como:

$$A \oplus B = \{z \mid (B)_z \cap A \neq \emptyset\}, \quad (3.2)$$

donde:

- A : imagen binaria original.
- B : elemento estructurante.
- $(B)_z$: traslación de B centrada en el punto z .
- $A \oplus B$: resultado de la dilatación, incluye todos los puntos z donde B se superpone parcialmente con A .

Erosión La erosión reduce las regiones blancas, eliminando píxeles en los bordes de los objetos. Se utiliza para eliminar ruido pequeño o separar objetos conectados débilmente [49].

$$A \ominus B = \{z \mid (B)_z \cap A^c \neq \emptyset\}, \quad (3.3)$$

donde:

- A : imagen binaria original.
- A^c : es el complemento de A .
- B : elemento estructurante.

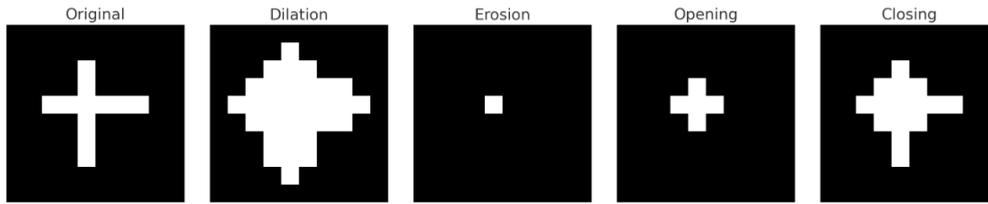


Figura 3.16: De izquierda a derecha: imagen binaria original con forma de cruz; dilatación que expande las regiones blancas; erosión que reduce las regiones blancas; apertura (erosión seguida de dilatación); y cierre (dilatación seguida de erosión).

- $(B)_z$: traslación de B al punto z .
- $A \ominus B$: resultado de la erosión, incluye todos los puntos z donde B está completamente contenido dentro de A .

Operaciones combinadas Combinando dilatación y erosión se obtienen operaciones más complejas:

- **Apertura:** Es la erosión seguida de una dilatación. Se utiliza para eliminar pequeños objetos o imperfecciones en los bordes, suavizando contornos.

$$A \circ B = (A \ominus B) \oplus B, \quad (3.4)$$

donde A es la imagen binaria sobre la que se aplica la operación, y B es el elemento estructurante. Primero se realiza una erosión (\ominus), seguida de una dilatación (\oplus).

- **Cierre:** Es la dilatación seguida de la erosión. Se utiliza para rellenar pequeños huecos.

$$A \bullet B = (A \oplus B) \ominus B, \quad (3.5)$$

donde A es la imagen binaria y B es el elemento estructurante. Primero se realiza una dilatación (\oplus), seguido de una erosión (\ominus).

3.2.2. Modelos propuestos

En esta sección se describen los modelos de clasificación utilizados para la predicción de la PCR.

AC-CycleGAN En la [Figura 3.17](#) se muestra la arquitectura del modelo CycleGAN con clasificador auxiliar (AC-CycleGAN, por sus siglas en inglés), propuesto para este proyecto. Este modelo, al igual que la arquitectura clásica de CycleGAN [27] expuesta en la [Sección 2.2.2](#), está compuesto por dos generadores (G y F) y dos discriminadores (D_a y D_b). Los generadores aprenden a traducir imágenes entre

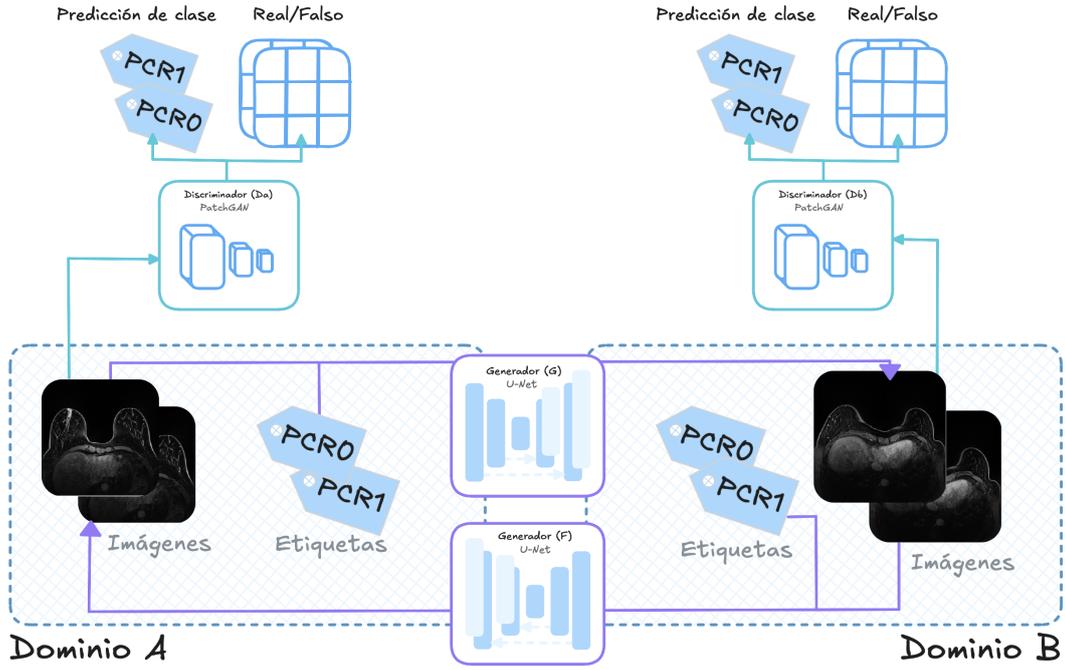


Figura 3.17: Esquema de la arquitectura AC-CycleGAN propuesta, con generadores condicionados, (G y F), y discriminadores (D_a y D_b) que realizan una predicción de clase.

dos dominios: G transforma imágenes del dominio A al dominio B , mientras que F realiza la operación inversa. Por su parte, los discriminadores D_a y D_b intentan distinguir entre imágenes reales y sintéticas de sus respectivos dominios.

El cambio respecto al CycleGAN es que el generador (véase Tabla 4) recibe como *input* imágenes del dominio A con su respectiva etiqueta. Y el discriminador (véase Tabla 5) da como salida no solo si la imagen es falsa o real, sino también la predicción de clase.

A las funciones de pérdida, descritas en la Sección 2.2.2, se le incorpora una nueva pérdida de clasificación auxiliar (Ecuación 3.6) para el discriminador. Esta mide la capacidad del discriminador para predecir correctamente la clase real, mediante una entropía cruzada entre las etiquetas verdaderas y las predichas.

$$\mathcal{L}_{\text{aux}}(D) = \mathbb{E}_{a \sim p_{\text{data}}(a)}[\text{CE}(C(a), c_a)] + \mathbb{E}_{b \sim p_{\text{data}}(b)}[\text{CE}(C(b), c_b)], \quad (3.6)$$

donde:

- $a \sim p_{\text{data}}(a)$: muestras reales del dominio A .
- $b \sim p_{\text{data}}(b)$: muestras reales del dominio B .
- CE es la pérdida de entropía cruzada.
- $C(\cdot)$ es la cabeza de clasificación del discriminador D

- c_a y c_b son las etiquetas verdaderas de las imágenes reales a y b .

Además, la forma de calcular la norma $L1$ en las pérdidas de identidad, $\mathcal{L}_{\text{identity}}$, y cíclica, \mathcal{L}_{cyc} , se modifica para introducir una ponderación basada en una máscara. Esta consiste en una pérdida $L1$ ponderada por una ROI, L_{mask} , donde los píxeles dentro de dicha región reciben un mayor peso. La fórmula es la siguiente:

$$\mathcal{L}_{\text{mask}} = \frac{\sum_i w_i \cdot |\hat{y}_i - y_i|}{\sum_i w_i},$$

donde:

- \hat{y}_i es el valor predicho.
- y_i es el valor real.
- $w_i = 1 + (\lambda_{\text{mask}} - 1) \cdot m_i$ es el peso aplicado a cada píxel.
- λ_{mask} es el peso adicional asignado a los píxeles dentro de la ROI.
- $m_i \in \{0, 1\}$ es la máscara binaria (1 si el píxel pertenece a la ROI, 0 en caso contrario).

De este modo, los píxeles fuera de la ROI (donde $w_i = 1$) tienen un peso menor en comparación con los que están dentro (donde $w_i = \lambda$), lo que permite focalizar el aprendizaje en la región de interés.

Modelo base Como modelo base para la predicción de la PCR se emplea la misma arquitectura del discriminador con la cabeza auxiliar en el modelo AC-CycleGAN. Se reutiliza la misma estructura pero sin la tarea adversarial, como punto de comparación con el modelo AC-CycleGAN. Esto nos permitirá evaluar si la incorporación de la traducción entre dominios en un entrenamiento adversarial aporta beneficios respecto a un enfoque de clasificación clásico.

Tabla 4: Arquitectura del generador en el modelo AC-CycleGAN, para una imagen de un canal de entrada $(B, 1, 256, 256)$, donde B indica el *batch size*.

Módulo	Operación / Capa	Detalles
Label Input	Embedding + Fully Connected	Codificación de etiquetas con Embedding (número de clases a predecir, 32), seguido de una capa totalmente conectada que expande a 256×256 , luego se reestructura como una imagen $(B, 1, 256, 256)$
Image Input	Entrada de imagen	$(B, 1, 256, 256)$
Concatenación	Concatenate (canales)	Imagen + etiqueta $\rightarrow (B, 2, 256, 256)$
Initial Conv	Conv2D + Norm + ReLU	Filtros: 64, kernel: 7×7 , stride: 1, padding: reflect + pad3
Downsampling 1	Conv2D + Norm + ReLU	Filtros: 128, kernel: 3×3 , stride: 2, padding: 1
Downsampling 2	Conv2D + Norm + ReLU	Filtros: 256, kernel: 3×3 , stride: 2, padding: 1
Residual Blocks	$6 \times$ ResNet Blocks	Cada bloque opera con 256 filtros, incluye padding, norm, ReLU
Upsampling 1	ConvTranspose2D + Norm + ReLU	Filtros: 128, kernel: 3×3 , stride: 2, padding: 1, output padding: 1
Upsampling 2	ConvTranspose2D + Norm + ReLU	Filtros: 64, kernel: 3×3 , stride: 2, padding: 1, output padding: 1
Output Layer	Conv2D + Tanh	Filtros: 1, kernel: 7×7 , padding: reflect + pad3
Output Imagen	Resultado final	Dimensión de salida: $(B, 1, 256, 256)$

Tabla 5: Arquitectura del discriminador en AC-CycleGAN para una entrada de $(B, 1, 256, 256)$, donde B indica el *batch size*.

Módulo	Operación / Capa	Detalles
Input Imagen	Entrada de imagen	$(B, 1, 256, 256)$
Conv1	Conv2D + LeakyReLU	Filtros: 64, kernel: 4×4 , stride: 2, padding: 1 $\rightarrow (B, 64, 128, 128)$
Conv2	Conv2D + Norm + LeakyReLU	Filtros: 128, kernel: 4×4 , stride: 2, padding: 1 $\rightarrow (B, 128, 64, 64)$
Conv3	Conv2D + Norm + LeakyReLU	Filtros: 256, kernel: 4×4 , stride: 2, padding: 1 $\rightarrow (B, 256, 32, 32)$
Conv4	Conv2D + Norm + LeakyReLU	Filtros: 512, kernel: 4×4 , stride: 1, padding: 1 $\rightarrow (B, 512, 31, 31)$
Validez Patch	Conv2D	Filtros: 1, kernel: 4×4 , stride: 1, padding: 1 $\rightarrow (B, 1, 30, 30)$
Cabeza auxiliar	Bloques secuenciales	
AdaptiveAvgPool	Reducción espacial	$(B, 512, 1, 1)$
Flatten	Aplanamiento	$(B, 512)$
Linear	Linear	$(B, 512) \rightarrow (B, \text{num_classes})$
Salida Final	Tupla (validación, clase)	$(B, 1, 32, 32), (B, \text{num_classes})$

Capítulo 4

4. Implementación

Este proyecto tiene como objetivo comprobar si es posible aprovechar el conocimiento aprendido para generar imágenes sintéticas para discriminar a qué clase pertenece una imagen (véase el objetivo definido en la [Subsubsección 1.3.1](#)).

Para ello, se ha desarrollado una arquitectura modular que permite reutilizar y adaptar el flujo de trabajo (*pipeline*), facilitando su escalabilidad.

4.1. Estructura del proyecto

Con la intención de desarrollar un proyecto con un diseño modular, mutable y flexible se siguen los principios SOLID (*Single responsibility, Open-closed, Liskov substitution, Interface segregation* y *Dependency inversion*) y se respetan, también, los principios de asignación de responsabilidades propuestos por GRASP (*General Responsibility Assignment Software Patterns*). Para lograrlo, se aplican diferentes patrones de diseño.

Un ejemplo es la combinación de los patrones *Strategy* y *Factory* para la creación de los *datasets* (véase [Figura 4.1](#)), según las necesidades del modelo: datos emparejados para Pix2Pix, datos no emparejados para CycleGAN, o datos con etiquetas para AC-CycleGAN.

Estos patrones no solo se aplican en la gestión de los conjuntos de datos, sino también en la definición de las funciones de pérdida del modelo -donde se incluye el patrón *Singleton*-, o en la creación de los modelos entrenados.

4.2. Pipeline

En esta sección describiremos los pasos metodológicos seguidos para el desarrollo de este trabajo. El *pipeline global* (véase [Figura 4.2](#)) está formado por dos fases, cada una recibe un *input* específico y envía un *output* para la siguiente fase.

4.2.1. Preprocesamiento

Como ya se ha expuesto previamente en la [Subsección 2.1](#), la heterogeneidad y la variabilidad propias de las imágenes médicas hacen imprescindible un preprocesamiento para afrontar problemas como el *domain shift* y el sesgo en los datos. Además, la exploración realizada en la [Subsubsección 3.1.1](#) resalta la importancia de definir un preprocesamiento.

El *pipeline* de preprocesamiento (véase [Figura 4.3](#)) está formado por cinco fases, con sus específicos *inputs* y *outputs*. En este trabajo, partimos de imágenes DCE-MRI en formato `.nii.gz`. El objetivo principal del preprocesamiento es ex-

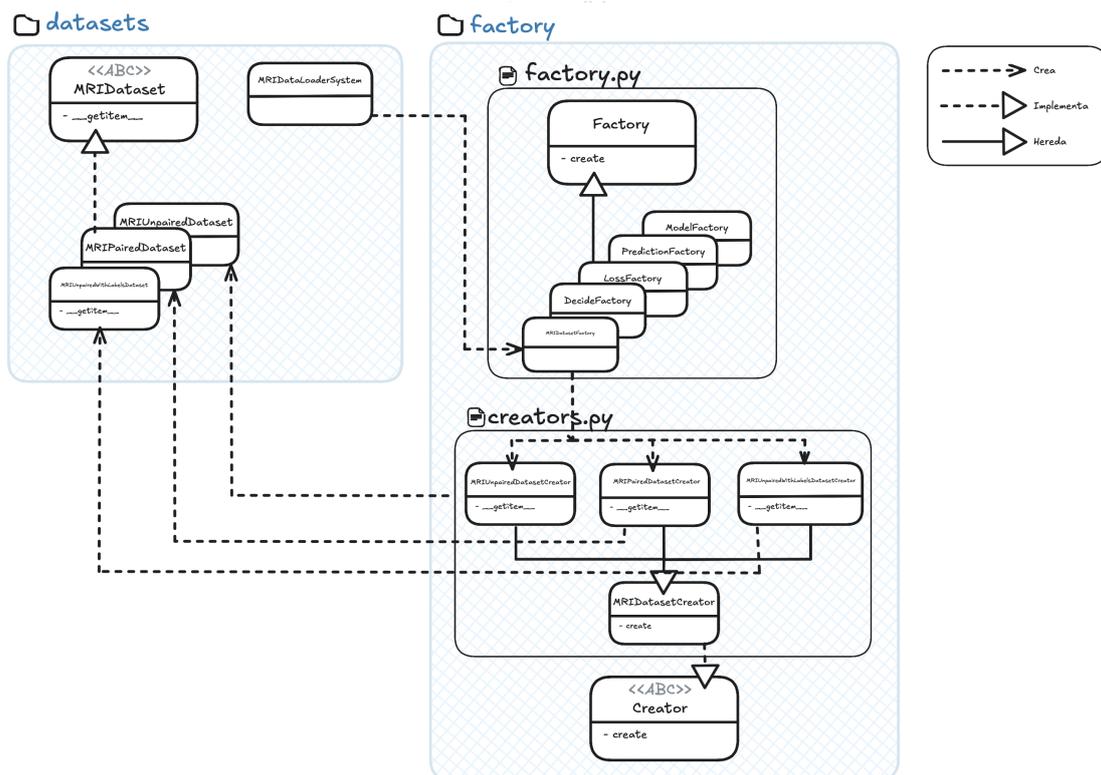


Figura 4.1: Ejemplo en el proyecto de la implementación de los patrones de diseño Factory y Strategy para la creación de los diferentes tipos de *datasets* utilizados.

traer cortes axiales normalizados y alineados espacialmente, los cuales son los más adecuados para el análisis debido a las ventajas mencionadas en la [Subsección 1.1](#), tales como una mejor visualización global de ambas mamas. Además, la mayoría de las imágenes del conjunto de datos se encuentran originalmente en el plano de adquisición axial, lo que refuerza la idoneidad de trabajar en dicho plano.

En la primera fase del *pipeline* se ha experimentado con diferentes tipos de segmentaciones. La intención es comparar su utilidad en la ejecución y determinar cuál representa mejor las características clínicas para la predicción. Cada corte se acompaña de una máscara, esta puede ser de tres tipos, en función de la región anatómica de interés (véase [Figura 4.4](#)):

- **Máscara tumoral:** Es la máscara disponible en el material, que corresponde a la segmentación semiautomática del tumor. Esta máscara se carga empleando la función `ReadImage` de la librería SimpleITK.
- **Máscara del primer plano:** Es la máscara que se obtiene al aplicar el algoritmo de Otsu ([Sección 3.2.1](#)) mediante el filtro `OtsuThresholdImageFilter` de la librería SimpleITK, junto con operaciones morfológicas ([Sección 3.2.1](#)). Esta máscara delimita el tejido a partir de la imagen DCE-MRI. El algoritmo de Otsu calcula un umbral óptimo entre dos clases: fondo (regiones de menor intensidad) y tejido (regiones de mayor intensidad). A continuación, se aplica una operación de dilatación (`BinaryDilate` de la librería Sim-

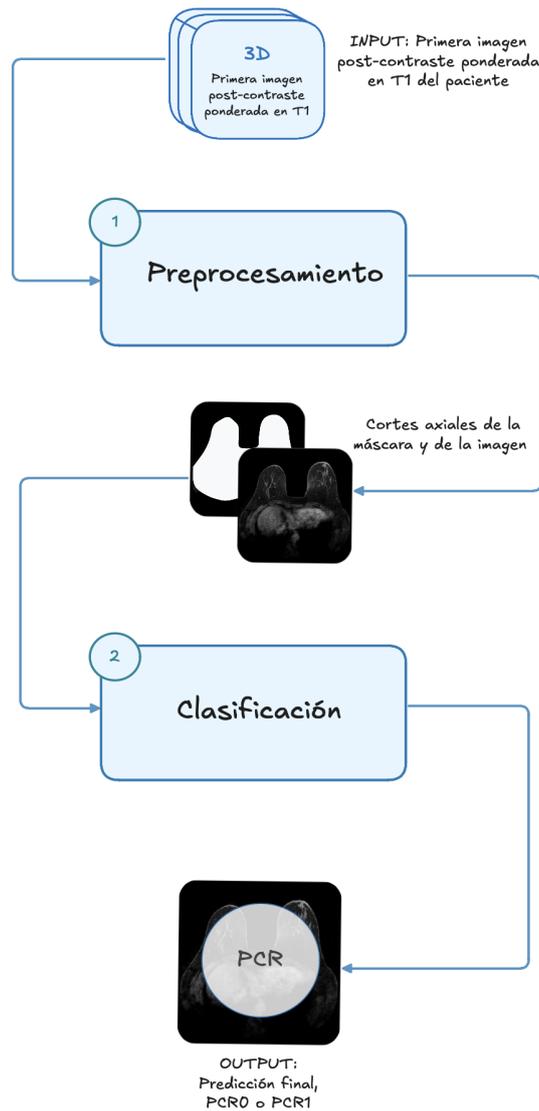


Figura 4.2: *Pipeline* global para una imagen 3D de entrada. El *pipeline* se compone de dos etapas principales: (1) preprocesamiento y (2) clasificación.

pleITK) para ampliar la región segmentada y mejorar la continuidad en los bordes, seguida de un cierre morfológico (`BinaryMorphologicalClosing` librería SimpleITK), para rellenar posibles agujeros. Finalmente, se eliminan pequeñas estructuras no relevantes mediante análisis de componentes conectados (`ConnectedComponent` librería SimpleITK), conservando únicamente la región de mayor tamaño (`LabelShapeStatisticsImageFilter` librería SimpleITK).

- **Máscara peritumoral:** Es una máscara binaria que se obtiene a partir de la máscara tumoral después de aplicar una dilatación (`BinaryDilate` de la librería SimpleITK). El objetivo es definir una región anular alrededor del tumor, para capturar señales clínicas relevantes para el diagnóstico, como cambios en la vascularización, edema o infiltración [50].

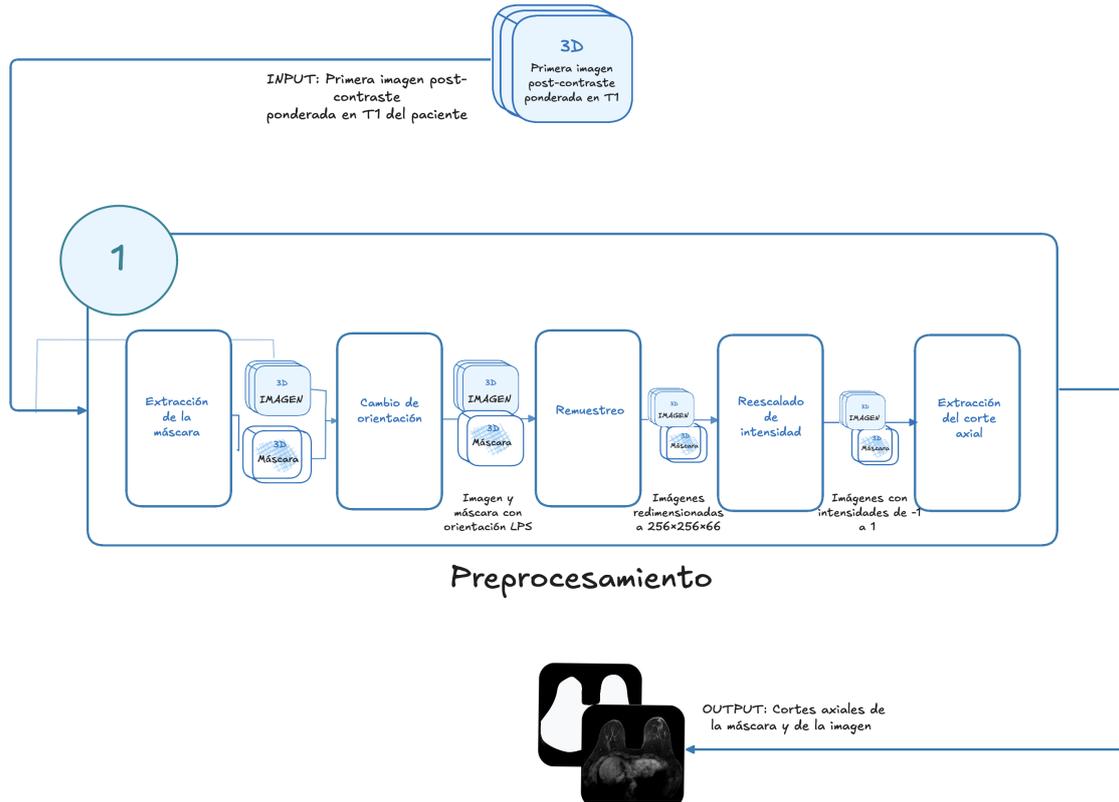


Figura 4.3: *Pipeline* de preprocesamiento que recibe una imagen DCE-MRI y devuelve un corte axial de la imagen, junto con su máscara. El *pipeline* está formado por cinco fases: extracción de las máscara, cambio de orientación, remuestreo, rescalado de intensidades y extracción del corte axial.

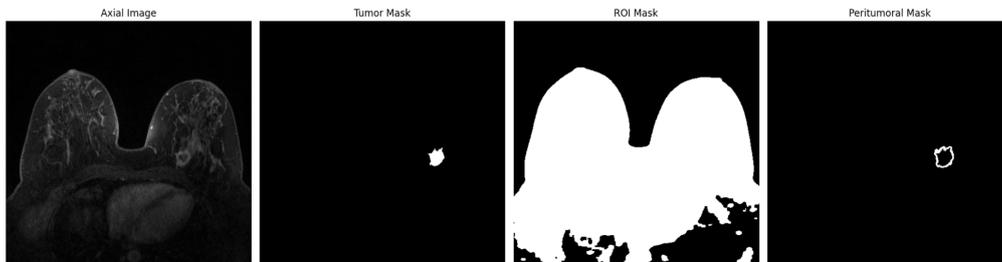


Figura 4.4: Ejemplo de diferentes máscaras utilizadas en el estudio, extraídas de un corte axial de una imagen DCE-MRI del paciente DUKE.001. De izquierda a derecha: imagen axial original, máscara tumoral segmentada, máscara del primer plano, y máscara peritumoral obtenida mediante una dilatación sobre la máscara tumoral.

Una vez extraída la segmentación, el siguiente paso consiste en dotar a todas las imágenes de una orientación común. Durante la exploración de los datos, se observó que las imágenes presentan dos orientaciones principales (Figura 3.10) según el estándar DICOM [52] (*Digital Imaging and Communication in Medicine*): LAI (*Left-Anterior-Inferior*) y PSL (*Posterior-Superior-Left*). Para cumplir el objetivo, coherencia espacial, se ha normalizado la orientación de todas las imágenes a LPS

(*Left-Posterior-Superior*) utilizando el `DICOMorientImageFilter` de la librería `SimpleITK` para establecerlo. Este sistema se alinea con las coordenadas anatómicas del paciente: el eje X apunta hacia la izquierda del paciente, el eje Y hacia su parte posterior y el eje Z hacia su parte superior. En cortes axiales, las imágenes suelen interpretarse como si el paciente estuviera tumbado boca arriba, con los pies hacia el observador y la cabeza al fondo de la imagen (véase [Figura 4.5](#)). Esto no solo facilita la interpretación de las imágenes, sino que también mejora el procesamiento por el modelo, que depende de la consistencia espacial para analizar correctamente las características.

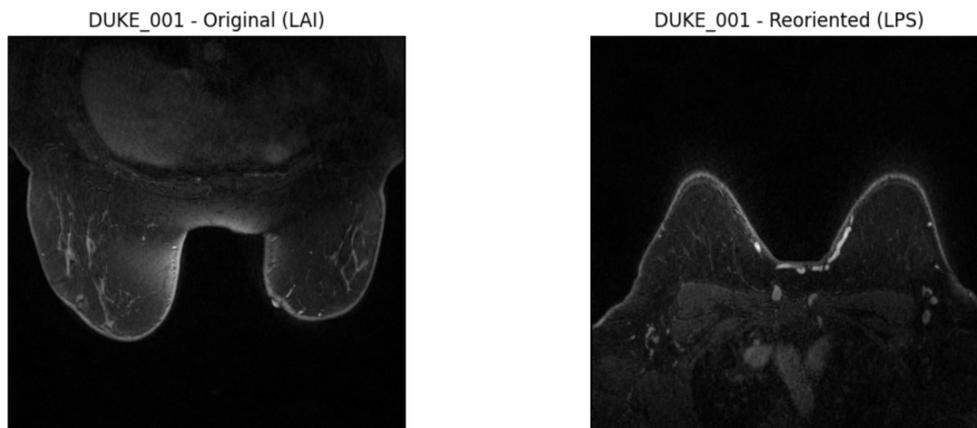


Figura 4.5: Comparación visual de la reorientación espacial de imágenes DCE-MRI al sistema de referencia anatómico LPS: imagen del paciente DUKE_001 originalmente en orientación LAI (izquierda) y tras la conversión a LPS (derecha).

Como se ha observado en la [Sección 3.1.1](#), existen diferentes tamaños que varían desde $256 \times 256 \times 44$ hasta $512 \times 512 \times 256$. Para unificar el tamaño de entrada para el modelo, todas las imágenes se redimensionan a unas dimensiones comunes de $256 \times 256 \times 66$. Para ello se utiliza una función basada en `SimpleITK`. Esta configura el (`ResampleImageFilter`) para conservar la orientación, el origen y el tipo de píxel de la imagen original. Luego, se aplica el tipo de interpolación definido, ya sea `sitkBSpline` o `sitkNearestNeighbor`.

Durante el proceso de reorientación, se ha detectado que algunas imágenes originalmente en orientación PSL, al aplicar la nueva orientación, alteran el orden de los ejes de x, y, z a z, x, y . Eso puede implicar que en algunos casos la dimensión correspondiente al eje x sea inferior a 256. Si se aplicara una interpolación directamente, se generarían deformaciones no deseadas en las estructuras anatómicas, produciendo resultados poco realistas o inconsistentes (véase [Figura 4.6](#)). Para evitar este problema, si alguna dimensión es menor que el tamaño esperado, se realiza un relleno (*padding*) con el valor mínimo de la imagen, de forma que se preserve la morfología original sin introducir distorsiones. Para el *padding*, se calcula cuánta altura y anchura adicional se necesita, aplicándolo de forma asimétrica —solo en la parte superior (eje vertical) y en los laterales (eje horizontal)— con el fin de preservar la estructura anatómica de la imagen. El relleno se realiza utilizando la función `numpy.pad`. En caso de que las dimensiones excedan los valores objetivo, se aplica

un rescalado mediante interpolación B-Spline para las imágenes y interpolación por vecino más cercano en las máscaras.

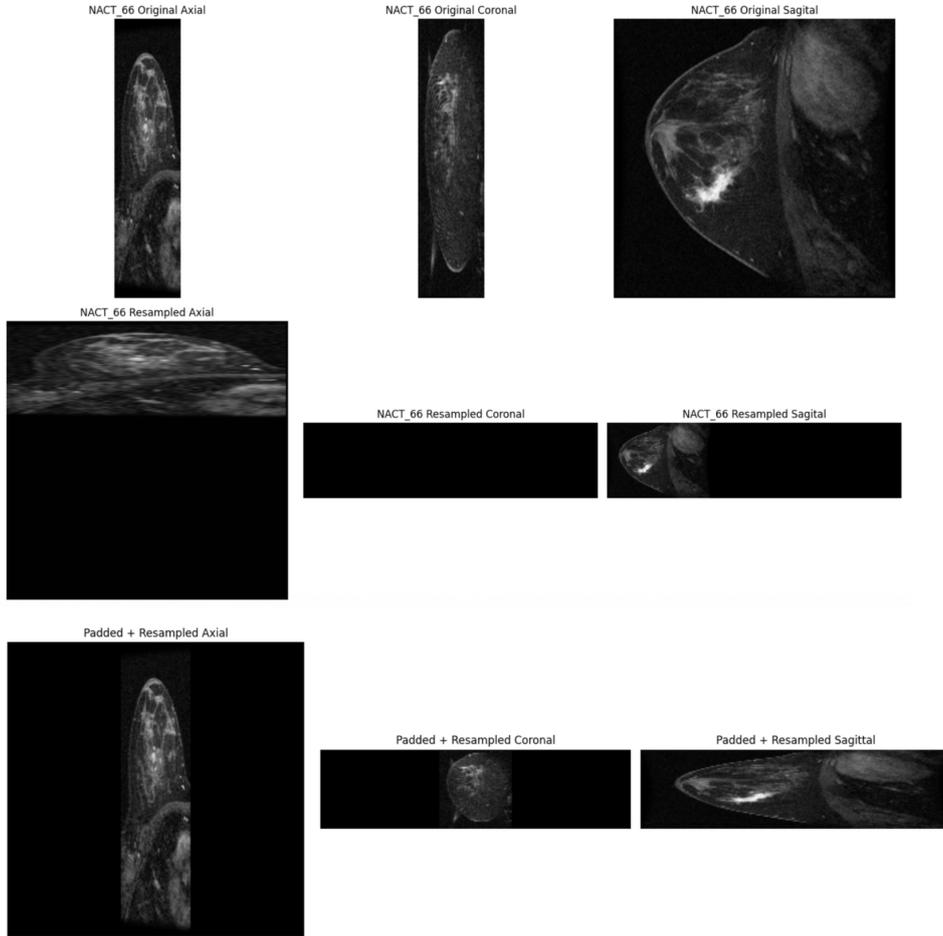


Figura 4.6: Comparación de los efectos en el rescalado de los planos de una imagen DCE-MRI del paciente NACT_66. En la primera fila se muestra la imagen original en orientación PSL. En la segunda fila, la imagen se rescaló directamente al tamaño de 256×256 . En la última fila, la imagen fue rellenada con padding para alcanzar una resolución de 256×256 y, finalmente, rescalada en profundidad.

Antes de extraer los cortes axiales, se aplica un rescalado de intensidades. Este paso es fundamental para reducir la variabilidad entre pacientes causada por diferencias en la adquisición de las imágenes. Al usar un rango simétrico como $[-1, 1]$, se facilita que los gradientes fluyan mejor y se evita el problema del desvanecimiento del gradiente en capas profundas.

Finalmente, en la extracción de los cortes se utiliza un CSV, `tumor_z_indices.csv`, que como se ha mencionado en la [Subsubsección 3.1.2](#), se emplea para garantizar que los cortes seleccionados contengan el tejido tumoral visible. Para la construcción, se aplica el preprocesamiento descrito para la segmentación tumoral, y posteriormente se utiliza la librería NumPy para obtener los índices Z que tienen tumor.

Para las máscaras de la ROI obtenidas con el algoritmo de Otsu y operaciones

morfológicas, se ha observado que, al extraer cortes pueden aparecer pequeñas estructuras aisladas. Como solución, se aplica un filtrado por conectividad -extrae la componente conexas de mayor tamaño- y, posteriormente, se realiza un relleno de huecos internos.

4.2.2. Imágenes sintéticas y Clasificación

En esta sección se describen dos *pipelines* principales desarrollados para abordar la tarea de predicción de la PCR. El objetivo es, por tanto, la clasificación de si un paciente ha alcanzado o no la PCR, utilizando como entrada una sola imagen axial correspondiente a la primera fase post-contraste.

Aunque el *pipeline* de preprocesamiento (Subsubsección 4.2.1) genera como salida una máscara segmentada del tumor junto con la imagen original, esta máscara únicamente se emplea durante el entrenamiento de la CycleGAN, y no se utiliza como entrada en la clasificación con el conjunto final. Además, durante el entrenamiento de todos los modelos se aplican técnicas de *data augmentation* para mejorar la capacidad de generalización y robustez. Se utilizan las transformaciones `RandomMotion` (véase Figura 4.7) y `RandomGhosting` (véase Figura 4.8) de la biblioteca TorchIO, cada una con una probabilidad del 50%, sobre las imágenes 3D DCE-MRI. Estas transformaciones tienen como objetivo simular artefactos comunes causados por el movimiento del paciente. Existen dos principales: los movimientos aleatorios, que generan imágenes ruidosas y borrosas, y los movimientos periódicos, que producen imágenes fantasma en la dirección de codificación de fase [53]. Para los cortes axiales, con sus respectivas máscaras, se aplica con una probabilidad del 50% una inversión horizontal (`torch.flip`). Esta transformación no altera la vista global del paciente tumbado boca arriba, pero invierte lateralmente la imagen, intercambiando las posiciones izquierda y derecha en términos anatómicos.

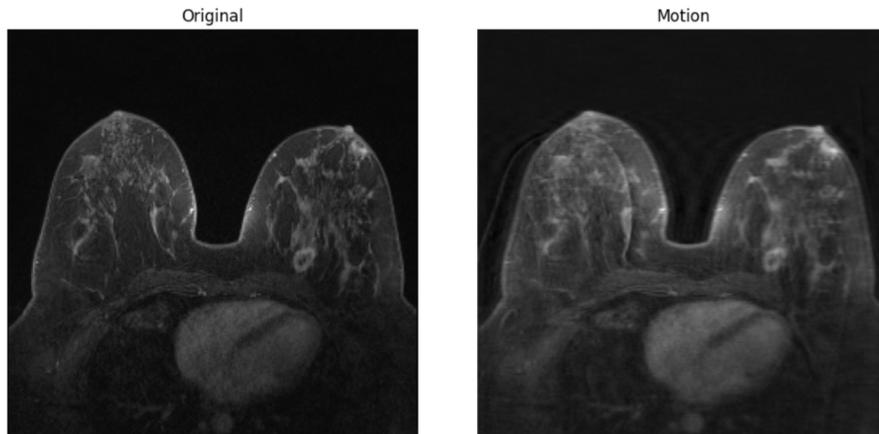


Figura 4.7: Aplicación del efecto de `RandomMotion` (derecha) sobre la imagen de corte axial del paciente DUKE_001 (izquierda).

A continuación se detallan los dos enfoques de clasificación utilizados.

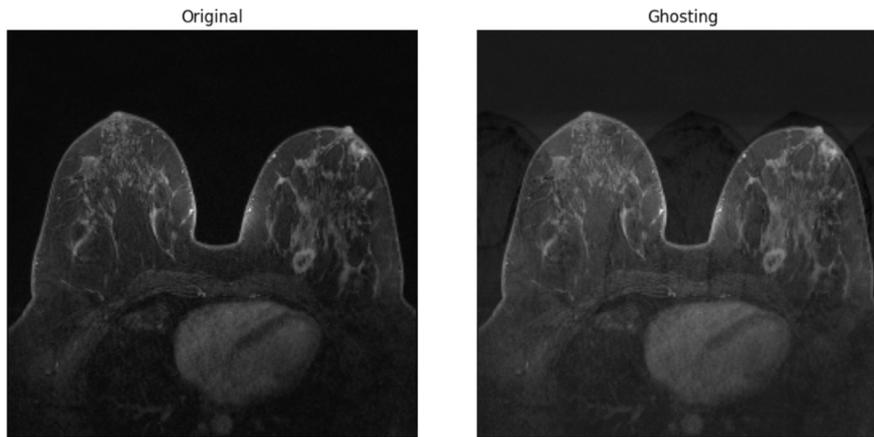


Figura 4.8: Aplicación del efecto de `RandomGhosting` (derecha) sobre la imagen de corte axial del paciente `DUKE_001` (izquierda).

Pipeline del modelo base Este *pipeline* actúa como referencia para comparar los resultados obtenidos con los métodos basados en CycleGAN. En este caso, la entrada es la imagen axial preprocesada (véase [Figura 4.9](#)).

Se ha implementado la arquitectura descrita en [Sección 3.2.2](#) con la librería PyTorch.

Pipeline con AC-CycleGAN El segundo *pipeline* (véase [Figura 4.10](#)) está formado por un generador y un discriminador del modelo CycleGAN con clasificador auxiliar (AC-CycleGAN, por sus siglas en inglés). Como hemos visto en la [Sección 2.2.2](#), se traducen imágenes de un dominio X a un dominio B y viceversa. El generador utilizado corresponde al generador G que ha sido entrenado para traducir imágenes del campo de fuerza 1.5T a 3T y el discriminador consiste en el discriminador que discrimina las imágenes reales y falsas del dominio de 3T. El discriminador incluye dos funciones: la capacidad de distinguir imágenes reales de generadas y la capacidad de predecir la clase a la que pertenece la imagen. La arquitectura y parte de la implementación se han adaptado a partir de la versión original de CycleGAN disponibles públicamente, incluyendo como referencia el trabajo de Zhu et al. (2017) [27].

Se ha implementado la arquitectura del generador y el discriminador descrita en [Sección 3.2.2](#) con la librería PyTorch.

El entrenamiento del modelo AC-CycleGAN se realiza utilizando la imagen original y su clase como entrada, junto con la máscara del tumor como guía en la función de pérdida para preservar la coherencia estructural. Una vez entrenado, el generador se emplea para transformar imágenes entre dominios. Estas imágenes sintéticas, junto con las reales, se utilizan como entrada de un clasificador CNN, permitiendo así evaluar si la información generada por el modelo mejora la capacidad del clasificador para distinguir entre los pacientes con y sin PCR.

Este *pipeline* permite además explorar si el generador es capaz de aprender

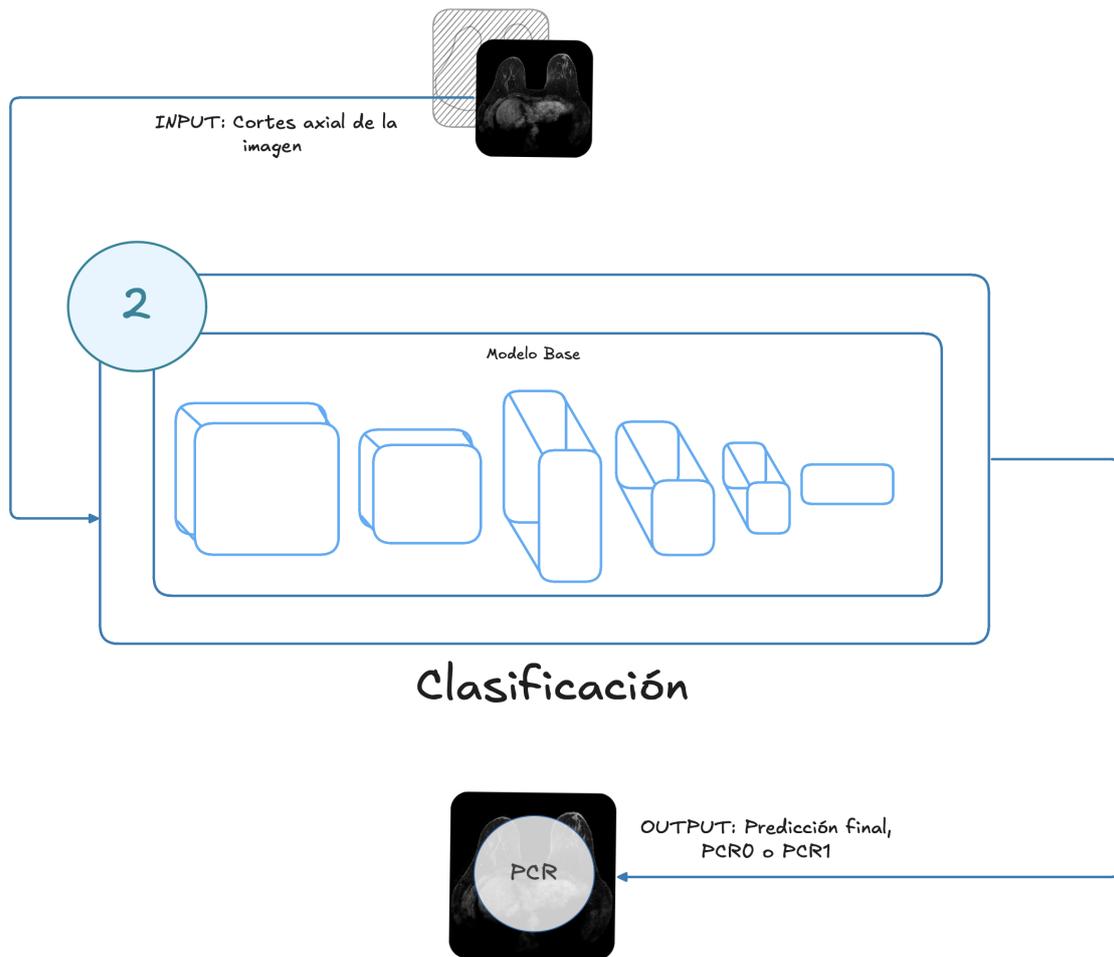


Figura 4.9: *Pipeline* de clasificación para un corte axial. El *pipeline* se compone de una CNN que emite una predicción de PCR.

implícitamente las distribuciones estadísticas de cada clase, y si dicha representación sintética contribuye a una mejor generalización del clasificador final. Para ello, en la implementación de la pérdida correspondiente al generador, se comprueba que las imágenes generadas sintéticamente conserven la etiqueta original mediante la cabeza de clasificación del discriminador. Por el contrario, para la implementación de la pérdida del discriminador relativa a la clasificación solo se tienen en cuenta las imágenes reales. Esto se debe a que en el entrenamiento el generador tiene como objetivo engañar al discriminador, produciendo imágenes sintéticas que imiten tanto la apariencia de la clase original como la del dominio objetivo. Por la parte del discriminador, se debe aprender a clasificar correctamente las imágenes reales para evitar introducir ruido en el entrenamiento.

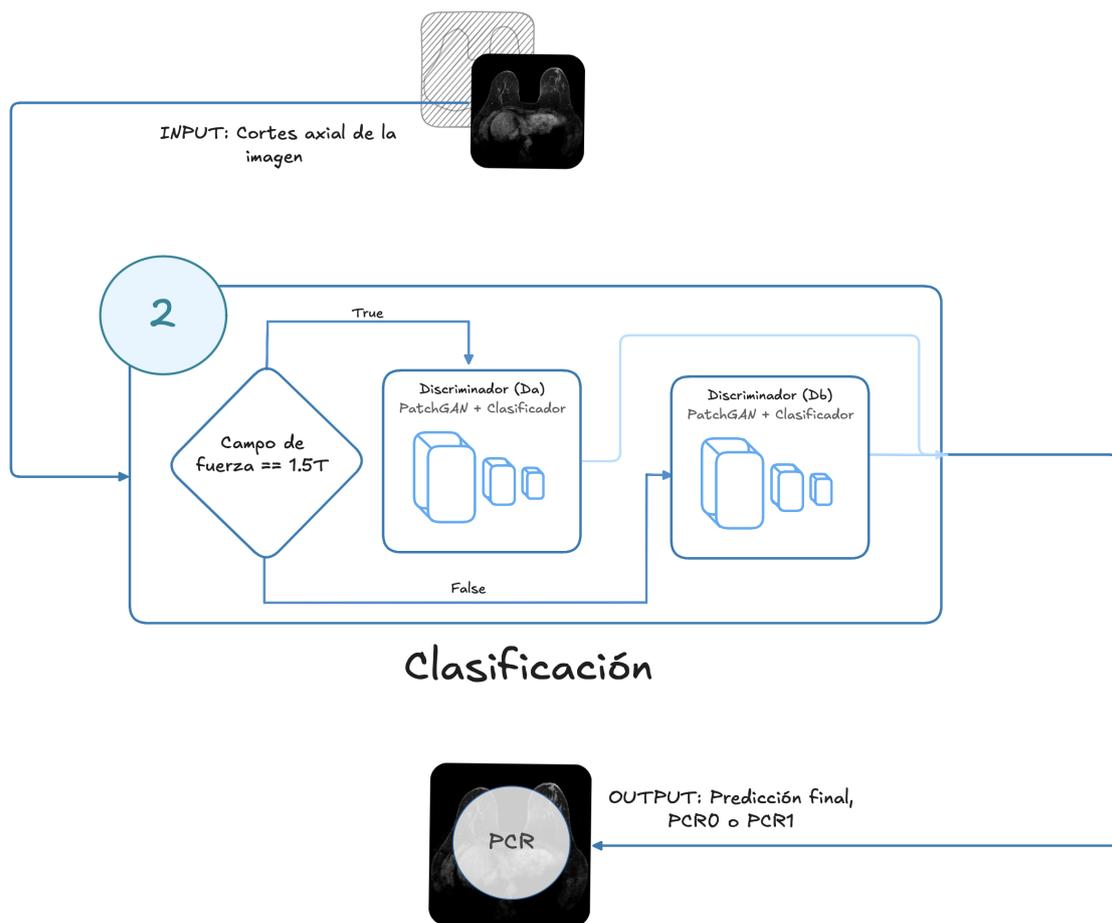


Figura 4.10: *Pipeline* de clasificación para un corte axial. El *pipeline* se compone de dos discriminadores para realizar la predicción de la PCR.

Capítulo 5

5. Análisis y discusión de resultados

Con el objetivo de contextualizar los resultados obtenidos con el modelo basado en CycleGAN, se han planteado principalmente dos experimentos.

Para evaluar los resultados, se han utilizado métricas clásicas de clasificación, como la precisión, curva ROC, AUC, *recall*, *f1-score* y *accuracy*, así como herramientas de interpretación como *Gradient-weighted Class Activation Mapping* (Grad-CAM) o *t-distributed stochastic neighbor embedding* (t-SNE), para analizar tanto las regiones de atención del modelo como la distribución de las características latentes en la discriminación entre clases.

En el primer experimento, además, se incluyen métricas de evaluación de imágenes sintéticas, como el MAE, PSNR y SSIM.

A continuación, se describen ambos experimentos, centrándose en los resultados obtenidos en cada etapa y en los ajustes realizados. Ambos experimentos se han llevado a cabo durante 100 épocas, utilizando una tasa de aprendizaje de 0.0002 y un parámetro $\beta_1 = 0,5$ para el optimizador.

5.1. Análisis de resultados del modelos basado en una CycleGAN

Todas las pruebas descritas en este apartado se realizaron durante 100 épocas, aplicando una reducción gradual de la tasa de aprendizaje a partir de la época 70 en intervalos de 10 iteraciones. Además, se establecieron los siguientes parámetros: $\lambda_{cyc} = 10,0$, $\lambda_{id} = 0,5\lambda_{cyc}$ y $\lambda_{mask} = 1,5$.

En un primer experimento, no se modificó la estructura descrita en [Sección 2.2.2](#), y se intentó discriminar entre clases de PCR. Este primer experimento reveló dos observaciones clave. En primer lugar, mediante el análisis de Grad-CAM aplicado al discriminador, se observó que la atención del modelo se centraba mayoritariamente en regiones no determinantes -fondo de la imagen- para la ejecución ([Figura 5.6b](#)), en lugar de áreas de interés como es la zona de la mama.

Para abordar este problema, se incorpora una máscara de atención que permite ponderar la pérdida $L1$, L_{mask} , utilizada tanto en las pérdidas de identidad como en las cíclicas (descritas en [Sección 3.2.2](#)). La máscara de la ROI del primer plano de la imagen [Figura 5.1a](#) permitió centrar la atención en esta área (véase [Figura 5.2](#)).

Aunque se corrigió la atención redirigiéndola hacia las ROI, el modelo mantuvo una capacidad limitada para diferenciar entre los dominios PCR0 y PCR1 (AUC = 0.5) con predicciones únicas en PCR0. Para comprobar si los generadores y discriminadores eran capaces de distinguir entre estas dos clases, y analizar la distribución resultante, se generaron diferentes mapas de características extraídas hasta la capa 19 del generador y hasta la capa 9 del discriminador. En esta representación de las

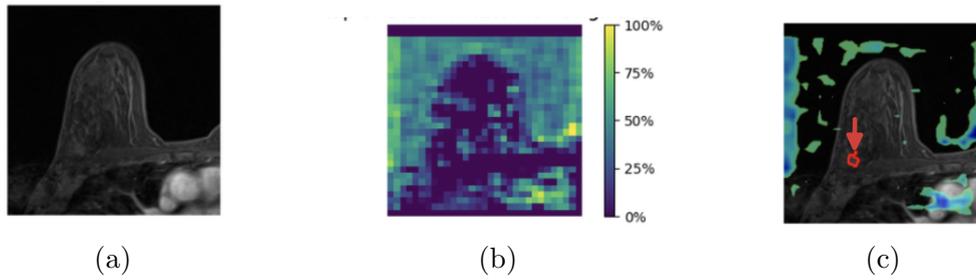


Figura 5.1: (a) Imagen original del dominio PCR1, (b) el mapa de atención generado por el discriminador del dominio B para dicha imagen, y (c) la superposición del mapa de atención sobre la imagen original. Nótese que la flecha señala el contorno bordeado en rojo, que corresponde a la máscara de segmentación del tumor, no al área de atención del mapa.

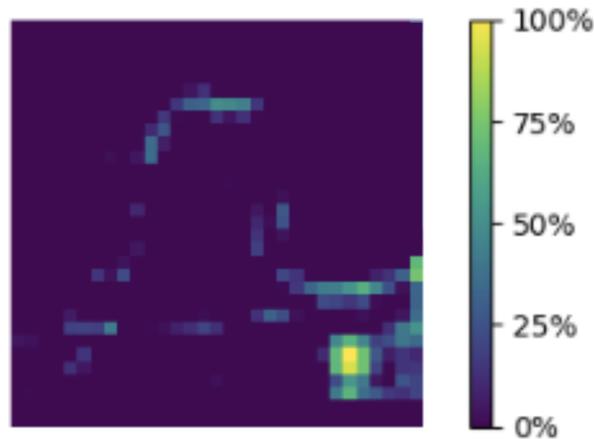


Figura 5.2: Mapa de atención de la imagen [Figura 5.1a](#) con ponderación por máscara.

distribuciones latentes, se observó que los modelos no eran capaces, por sí solos, de discriminar entre ambas clases (véase [Figura 5.3](#)). En el análisis t-SNE se realizó configurando los siguientes parámetros: número de componentes 2, perplexidad de 10, un máximo de 3000 iteraciones y usando la distancia euclidiana como métrica.

En un segundo experimento, se optó por la introducción de una clasificación explícita mediante el AC-CycleGAN. Este modelo, en específico, se entrenó en la traducción de las imágenes con campo magnético de 1.5T (dominio A) y de 3T (dominio B), ya que era necesario que ambos dominios contuvieran ejemplos de las clases PCR0 y PCR1. En la fase de evaluación (véase [Tabla 6](#)), el modelo mostró un rendimiento moderado para la clase PCR0, con una precisión del 66.77%, un recall del 66.99% y una puntuación F1 de 66.88%. Sin embargo, para la clase PCR1, el rendimiento fue inferior, con una precisión del 26.09%, un recall del 25.90% y una F1 de 25.99%. La precisión global del modelo fue del 54.24%. Esto demuestra la dificultad de generar representaciones suficientemente discriminativas para la clase PCR1 durante el proceso de traducción entre dominios.

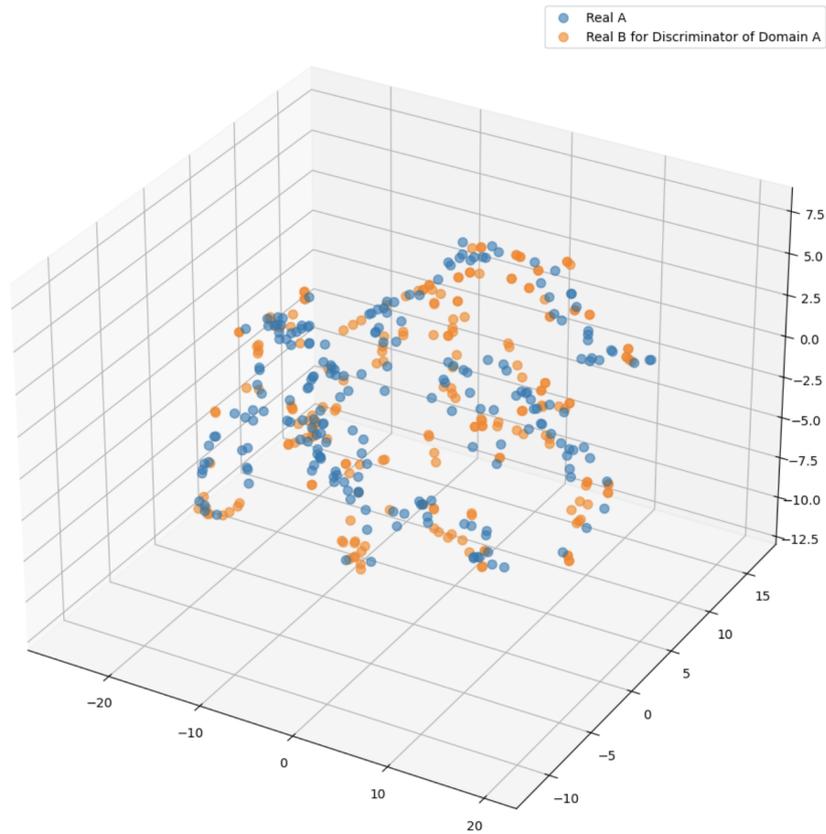


Figura 5.3: Representación tridimensional (t-SNE) de las características extraídas por el discriminador del dominio A de los datos de prueba, aplicadas a imágenes originales de los dominios A (azul) y B (naranja)

Tabla 6: Métricas de evaluación del modelo AC-CycleGAN sin mecanismos de atención por clase y medias globales.

Clase	Precisión	Recall	F1-score	Soporte
PCR0	0.67	0.67	0.67	306
PCR1	0.26	0.26	0.26	142
<i>Accuracy</i>			0.54	448
Macro promedio	0.46	0.46	0.46	448
Promedio ponderado	0.54	0.54	0.54	448

AUC: 0.44

Al redirigir la atención del modelo hacia las regiones de interés (ROI), mediante el uso de máscaras y la incorporación de una clasificación explícita con el modelo AC-CycleGAN, los resultados obtenidos mejoran levemente la precisión, especialmente para la atención en la región peritumoral (véase [Tabla 7](#)). En particular, se

observa un aumento en las métricas de la clase PCR0, con una precisión de 0.71 y un recall de 0.74, en comparación con la versión sin atención, donde estos valores eran 0.67 para ambas métricas (Tabla 6). Para la clase PCR1, aunque también hay un ligero incremento (precisión de 0.34 y recall de 0.30 frente a 0.26 y 0.26 respectivamente), el rendimiento sigue siendo limitado, especialmente para la clase PCR1. Al comparar estos resultados con las investigaciones del estado del arte (véase Subsección 2.3), se observa que dichas propuestas alcanzan métricas considerablemente más altas y una mejor aproximación al objetivo de discriminar entre las clases PCR0 y PCR1. La mayoría de estos trabajos incorpora variables clínicas, datos histopatológicos o información longitudinal, además de las imágenes, para mejorar la capacidad predictiva. Sin embargo, incluso considerando únicamente el mecanismo de atención aplicado a imágenes estáticas, era esperable observar una mejora en la discriminación considerable. En este caso, el modelo no logró aprender representaciones latentes suficientemente discriminativas, como lo evidencian los análisis t-SNE realizados (véase Figura 5.6a).

Tabla 7: Métricas de evaluación del modelo AC-CycleGAN con atención en la región peritumoral por clase y medias globales.

Clase	Precisión	Recall	F1-score	Soporte
PCR0	0.71	0.74	0.72	306
PCR1	0.34	0.30	0.32	142
<i>Accuracy</i>			0.61	448
Macro promedio	0.52	0.52	0.52	448
Promedio ponderado	0.59	0.60	0.60	448

AUC: 0.54

Para evaluar la calidad de las imágenes sintéticas generadas, al no disponer de pares de imágenes para realizar una comparación directa píxel a píxel entre la imagen real y la generada, se ha evaluado la calidad de las reconstrucciones. Para ello, se han comparado las métricas de MAE, PSNR y SSIM para el modelo CycleGAN y el AC-CycleGAN, con sus respectivos entrenamientos en las ROI. Esta evaluación se realizó en dos escenarios principales: la consistencia cíclica (véase Figura 5.4) y la consistencia de identidad (véase Figura 5.5).

En el escenario de la consistencia cíclica, el modelo CycleGAN tiene un rendimiento superior al AC-CycleGAN. En particular destacan en el entrenamiento sin ningún tipo de atención (véase Figura 5.4a) y con atención en regiones tumorales (véase Figura 5.4c) con valores de MAE bajos (0.028 y 0.034) y de PSNR por encima de 32.5. Para los modelos entrenados con atención en el primer plano (véase Figura 5.4b) el AC-CycleGAN muestra una competitividad significativa, incluso superando ligeramente al CycleGAN en ciertos aspectos, con valores MAE muy bajos (0.028 y 0.031) y PSNR por encima de 32.

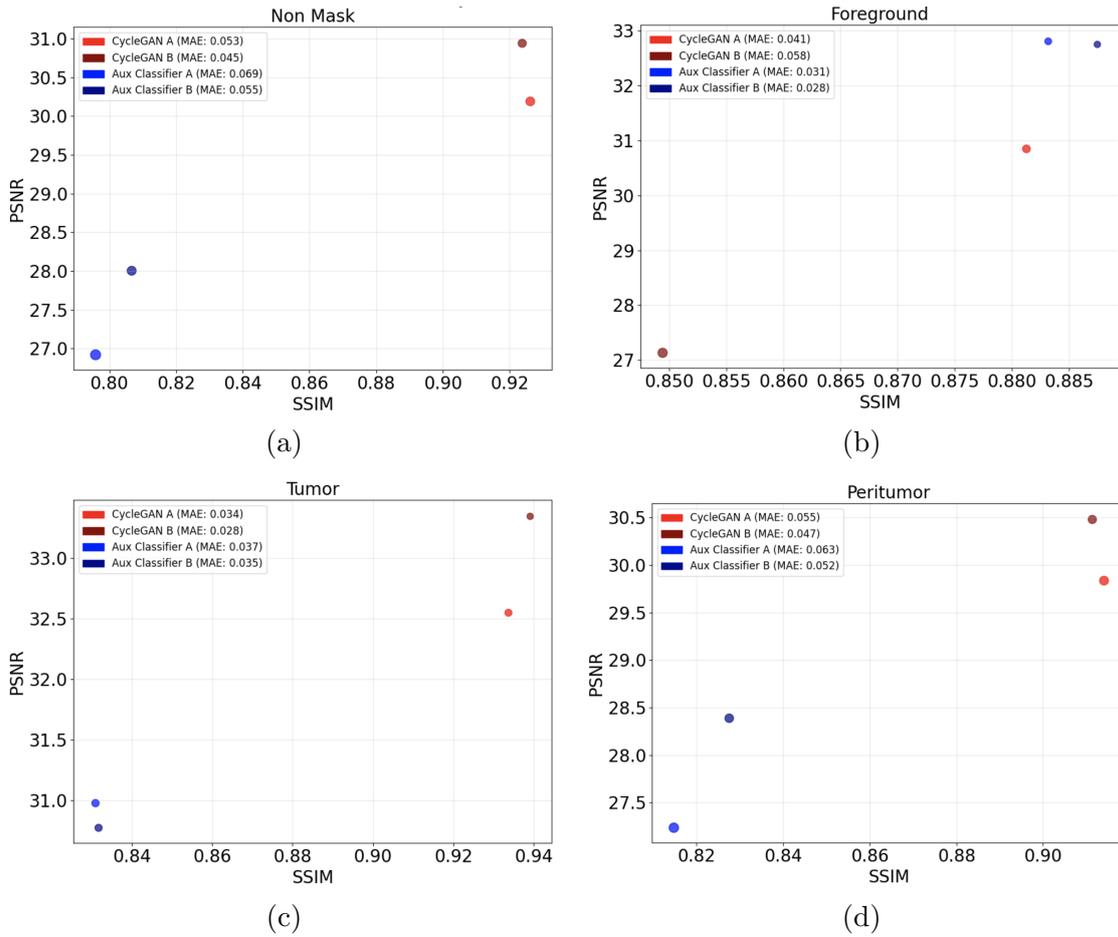


Figura 5.4: Comparativa del rendimiento de los generadores del modelo CycleGAN (rojo) y AC-CycleGAN (azul) en la tarea de consistencia de cíclica. Se evalúan los generadores A (dominio A a B) y B (dominio B a A) en cuatro regiones de interés: (a) sin máscara, (b) primer plano, (c) región tumoral y (d) región peritumoral. El tamaño de cada punto es proporcional al MAE.

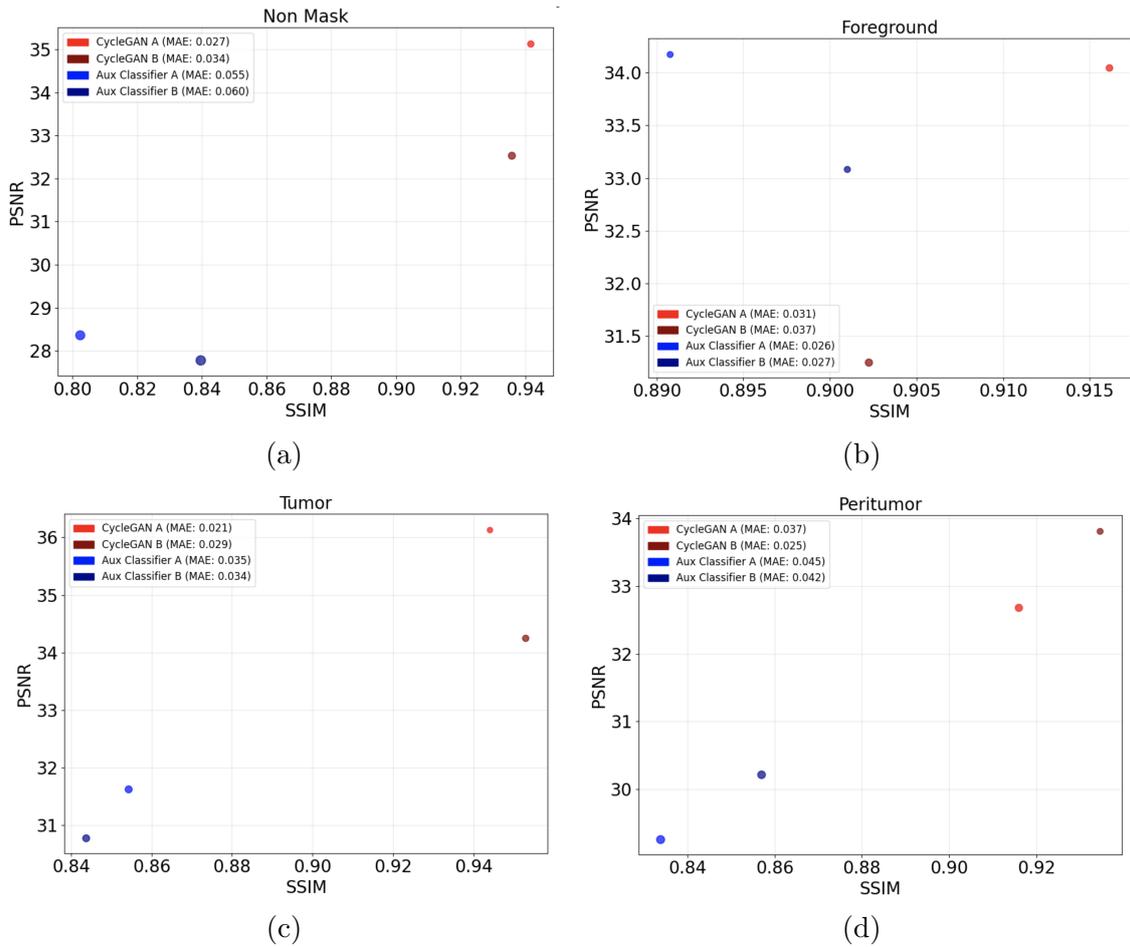


Figura 5.5: Comparativa del rendimiento de los generadores del modelo CycleGAN (rojo) y AC-CycleGAN (azul) en la tarea de consistencia de identidad. Se evalúan los generadores A (dominio A a B) y B (dominio B a A) en cuatro regiones de interés: (a) sin máscara, (b) primer plano, (c) región tumoral y (d) región peritumoral. El tamaño de cada punto es proporcional al MAE.

Por otro lado, en la consistencia de identidad, en su mayoría, el modelo CycleGAN presenta un rendimiento superior respecto al AC-CycleGAN. Vuelven a destacar en el entrenamiento sin atención (véase [Figura 5.5a](#)) y con atención en las regiones tumorales (véase [Figura 5.5c](#)). En general, la adición del clasificador auxiliar, en comparación con la CycleGAN, no contribuye positivamente a la capacidad de los generadores para conservar la imagen del dominio destino al que se aplica la transformación de identidad. Los peores resultados de consistencia de identidad se encuentran en el entrenamiento sin atención para el AC-CycleGAN (véase [Figura 5.5a](#)), y en la atención peritumoral (véase [Figura 5.5d](#)) para ambos modelos, aunque el CycleGAN sigue superando al AC-CycleGAN. Aún así, es importante destacar que, incluso en estas regiones, los resultados de identidad del CycleGAN son considerablemente buenos, lo que indica que el modelo mantiene una alta fidelidad en la reconstrucción directa.

Se observa que las métricas de identidad son mejores que las de consistencia cíclica, debido a que la tarea de reconstrucción de identidad es más sencilla que la reconstrucción de la imagen original después de un ciclo completo de traducción bidireccional.

5.2. Análisis de resultados del modelo base

Con el objetivo de evaluar el beneficio que aporta la traducción entre dominios mediante entrenamiento adversarial, se ha entrenado un modelo base de clasificación directa sobre las imágenes originales. En la [Tabla 8](#) se muestran las métricas de evaluación obtenidas por clase y sus promedios globales.

Los resultados muestran la incapacidad del modelo base para detectar correctamente los casos de PCR, con una exactitud del 55 %, un F1-score macro de 0.49 y un AUC de 0.52. El modelo presenta un fuerte sesgo hacia la clase PCR0 mayoritaria en el test, que presenta métricas notablemente superiores respecto a la clase PCR1. Aunque en el entrenamiento se cuenta con el mismo número de muestras de PCR0 y PCR1 (minoritaria en número de pacientes), esto se logra mediante la introducción de un mayor número de cortes por paciente en el caso de PCR1. Esta última, obtiene puntuaciones significativamente más bajas en precisión, recall y F1-score, lo que indica que el modelo no es capaz de aprender adecuadamente las características distintas.

Estos resultados son coherentes con los resultados investigados en la literatura (véase [Subsección 2.3](#)), donde, al utilizar imágenes completas sin mecanismos de atención ni la incorporación de variables clínicas, Chang et al. (2024) [37] obtuvieron un AUC en el rango de 0.45–0.54.

5.3. Comparativa entre modelos

Los resultados obtenidos con ambos modelos son insuficientes para una predicción efectiva de las clases PCR0 y PCR1. En ambos casos, la clase PCR1 presenta un rendimiento menor. Aunque se intentó evitar el desbalanceo entre clases igua-

Tabla 8: Métricas de evaluación del modelo base por clase y medias globales.

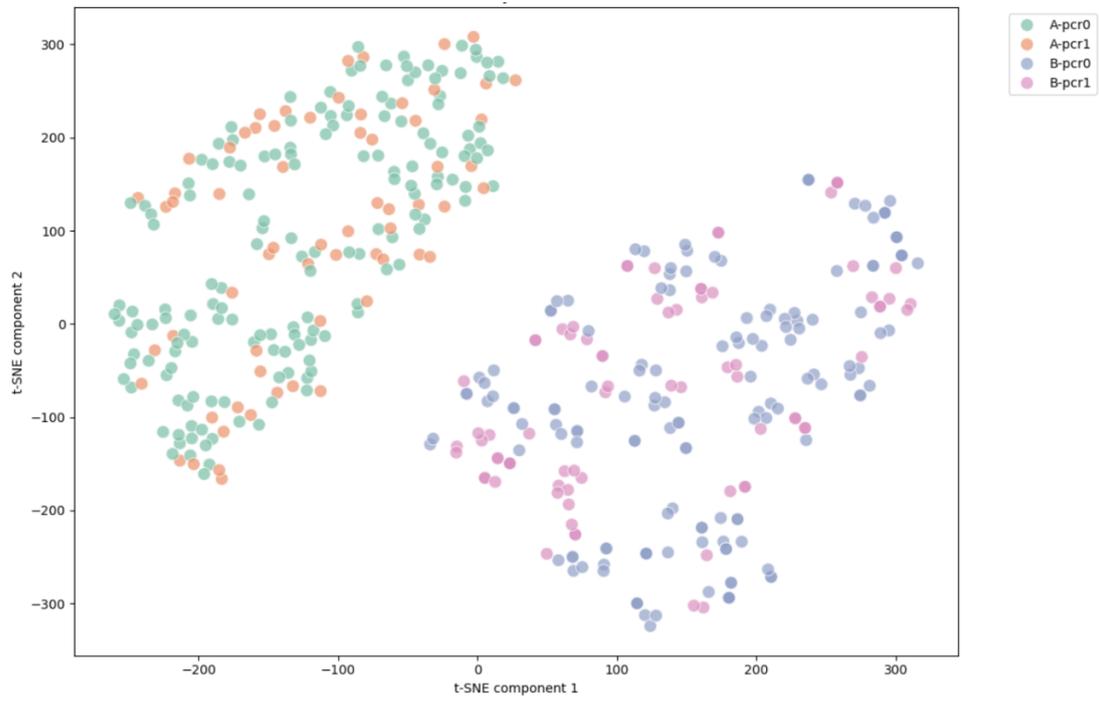
Clase	Precisión	Recall	F1-score	Soporte
PCR0	0.68	0.65	0.66	306
PCR1	0.31	0.34	0.32	142
<i>Accuracy</i>			0.55	448
Macro promedio	0.49	0.49	0.49	448
Promedio ponderado	0.56	0.55	0.55	448

AUC: 0.523

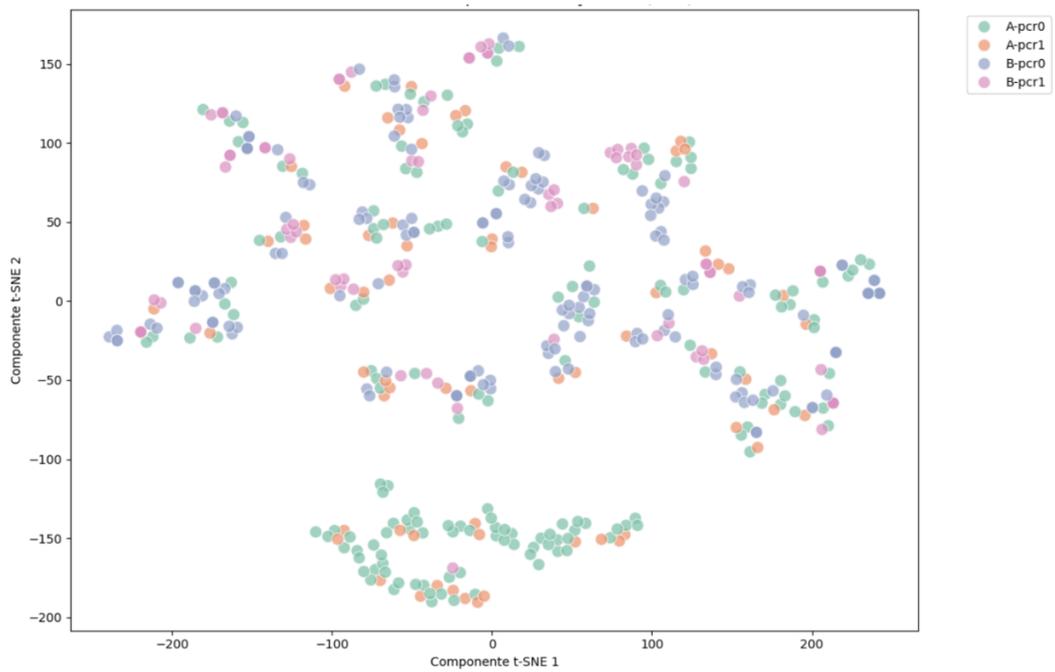
lando su número mediante la extracción de distintos cortes axiales para un mismo paciente (como se investigó en [Sección 2](#)).

Estas limitaciones pueden estar relacionadas con cuestiones más profundas, como la capacidad del modelo de aprender representaciones verdaderamente discriminativas. La predicción de la PCR no es una tarea sencilla, debido a la alta similitud visual entre las imágenes. Tanto en los mapas de Grad-CAM como los análisis t-SNE indican que las representaciones latentes aprendidas por los discriminadores (véase [Figura 5.6](#)) no muestran una separación clara entre las clases. Incluso después de redirigir la atención, sigue dispersándose en zonas irrelevantes (véase [Figura 5.2](#)).

En comparación con los resultados de la literatura (véase [Subsección 2.3](#)), son insuficientes. Este estudio sugiere que el éxito en la predicción de la PCR no depende solo de la arquitectura del modelo, sino también de la inclusión de variables clínicas, la precisión del preprocesamiento y la atención a las ROI.



(a)



(b)

Figura 5.6: Mapa t-SNE de las características extraídas de las imágenes del conjunto de prueba según la PCR, obtenidas (a) de los discriminadores de AC-CycleGAN y (b) del modelo base.

Capítulo 6

6. Conclusiones y trabajo futuro

Este trabajo ha explorado la viabilidad de utilizar una arquitectura basada en CycleGAN para la clasificación de imágenes DCE-MRI en términos de la PCR para pacientes que reciben NACT en cáncer de mama. Se planteó la hipótesis de que, mediante el entrenamiento del modelo para la *data augmentation* entre dos dominios, PCR0 y PCR1, la CycleGAN podría aprender implícitamente las dos distribuciones latentes y, con ello, discriminar las clases.

Los resultados obtenidos, sin embargo, muestran limitaciones significativas. Siempre teniendo en cuenta la restricción temporal y computacional que impidió explorar o ajustar los parámetros para nuevas configuraciones.

A pesar de emplear técnicas de atención, la capacidad de discriminar las clases fue baja, especialmente en la clase PCR1. El modelo base, con la arquitectura del discriminador, tampoco logró superar los resultados del modelo AC-CycleGAN. Esto sugiere que la dificultad para predecir las clases puede estar relacionada con la similitud visual entre ellas. Además, el hecho de que el modelo base y el discriminador, con estructuras similares, obtengan resultados comparables indica que el entrenamiento adversarial por sí solo no es suficiente para mejorar la discriminación entre las clases en este contexto.

Por último, hay que considerar la posible influencia del dominio de adquisición de los datos. Los modelos pueden captar diferencias entre el dominio técnico en lugar de los patrones clínicos, introduciendo un sesgo.

6.1. Trabajo futuro

Los resultados y conclusiones extraídas de este estudio abren la puerta a nuevas líneas de investigación futuras.

Por un lado, se podría explorar la integración de herramientas para una segmentación más precisa, así como mecanismos de atención avanzados.

Por otro lado, sería interesante incorporar variables clínicas, como los receptores hormonales (HR, *Hormone Receptor*) y HER2 (*Human Epidermal growth factor Receptor 2*), presentes en el conjunto de datos MAMA-MIA [41], para reforzar la capacidad de discriminación, además de separar la traducción entre dominios de la tarea de clasificación. También se puede profundizar en la generación de imágenes condicionales mediante el uso de etiquetas en el generador. Es decir, en la generación de imágenes con AC-CycleGAN, el generador recibe una etiqueta de clase (PCR0 o PCR1) para guiar la generación de manera condicional, incorporando una pérdida basada en la salida del discriminador para las imágenes generadas. Sin embargo, dado el bajo rendimiento del discriminador en la clasificación de estas clases, la señal de retroalimentación es errónea, lo que impide que la generación condicional sea efectiva. Una posible mejora sería explorar el uso de un clasificador externo

con buena precisión y combinarlo con el generador, de modo que la generación condicional esté guiada por una supervisión más fiable.

Estas nuevas líneas de investigación abren nuevas oportunidades para mejorar la eficacia de los modelos en escenarios clínicos. A lo largo del proyecto se ha observado que estos escenarios están caracterizados por la heterogeneidad, la escasez de datos y una elevada complejidad. Profundizar en estas líneas puede conducir al desarrollo de soluciones reales y adaptadas a las necesidades del ámbito médico.

Referencias

- [1] *Breast cancer. WHO Fact Sheets*. Disponible en <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>.
- [2] World Health Organization (WHO) & International Agency for Research on Cancer. *Rate of cancer deaths among females worldwide in 2022, by type of cancer (per 100,000 population)*. Statista, (29 de abril de 2024). Disponible en <https://www.statista.com/statistics/1031301/cancer-death-rate-females-worldwide-by-type/>.
- [3] International Agency for Research on Cancer. *Distribución porcentual de las muertes por cáncer más frecuentes entre la población mundial en 2022, por tipo*. Statista, (27 de julio de 2024). Disponible en <https://www.statista.com/estadisticas/1390420/tipos-de-cancer-causantes-de-mas-muertes/>.
- [4] Kim, J., Harper, A., McCormack, V., Sung, H., Houssami, N., Morgan, E., Mutebi, M., Garvey, G., Soerjomataram, I., Fidler-Benaoudia, M. M. *Global patterns and trends in breast cancer incidence and mortality across 185 countries*. *Nature Medicine*, (2025). Disponible en <https://doi.org/10.1038/s41591-025-03502-3>.
- [5] Iqbal, J., Shafi, A. A., Alharthi, B. N. *Neoadjuvant chemotherapy in locally advanced breast cancer*. *PubMed*, (1 noviembre 2014). Disponible en <https://pubmed.ncbi.nlm.nih.gov/25404445/>.
- [6] OMS, International Agency for Research on Cancer. *Tipos de cáncer con mayor número de nuevos casos detectados a nivel mundial en 2022 y 2050*. Statista, (12 de julio de 2024). Disponible en <https://www.statista.com/estadisticas/636254/cancer-nuevos-casos-a-nivel-mundial-por-tipo/>.
- [7] Antonini, M., Mattar, A., Pereira, T. M., Oliveira, L. L., Teixeira, M. D., Amorim, A. G., Ferraro, O., De Oliveira, L. C., Ramos, M. D. N. M., Cavalcante, F. P., Zerwes, F., Madeira, M., Sorares, L. R., De Camargo Millen, E., Frasson, A. L., Brenelli, F. P., Facina, G., Fenile, R., Arakelian, R., . . . Gebrim, L. H. *Pathologic Complete Response and Breast Cancer Survival Post-Neoadjuvant Chemotherapy: A Systematic Review and Meta-Analysis of Real-World Data*. *Heliyon*, (2025). Disponible en <https://doi.org/10.1016/j.heliyon.2025.e43069>
- [8] Losurdo, L., Basile, T. M. A., Fanizzi, A., Bellotti, R., Bottigli, U., Carbonara, R., Dentamaro, R., Diacono, D., Didonna, V., Lombardi, A., Giotta, F., Guaragnella, C., Mangia, A., Massafra, R., Tamborra, P., Tangaro, S., La Forgia, D. *A Gradient-Based Approach for Breast DCE-MRI Analysis*. *BioMed Research International*, 2018 (2018), p. 1-10. Disponible en <https://doi.org/10.1155/2018/9032408>.
- [9] Chatterji, M., Mercado, C. L., Moy, L. *Optimizing 1.5-Tesla and 3-Tesla Dynamic Contrast-Enhanced Magnetic Resonance Imaging of the Breasts*. *Magnetic*

- Resonance Imaging Clinics of North America*, 18(2) (2010), p. 207-224. Disponible en <https://doi.org/10.1016/j.mric.2010.02.011>.
- [10] Mann, R. M., Cho, N., Moy, L. *Breast MRI: State of the Art. Radiology*, 292(3) (2019), p. 520-536. Disponible en <https://doi.org/10.1148/radiol.2019182947>.
- [11] Atasever, S., Azginoglu, N., Terzi, D. S., Terzi, R. *A comprehensive survey of deep learning research on medical image analysis with focus on transfer learning. Clinical Imaging*, 94 (2022), p. 18-41. Disponible en <https://doi.org/10.1016/j.clinimag.2022.11.003>.
- [12] Sharma, P. S., Saindane, A. M. *Standardizing Magnetic Resonance Imaging Protocols Across a Large Radiology Enterprise: Barriers and Solutions. Current Problems in Diagnostic Radiology*, 49(5) (2020), pp. 312-316. Disponible en <https://doi.org/10.1067/j.cpradiol.2020.01.012>.
- [13] Furman-Haran, E., Feinberg, M. S., Badikhi, D., Eyal, E., Zehavi, T., Degani, H. *Standardization of Radiological Evaluation of Dynamic Contrast Enhanced MRI: Application in Breast Cancer Diagnosis. TCRT Express* (2013). Disponible en <https://doi.org/10.7785/tcrtexpress.2013.600263>.
- [14] Guan, H., Liu, M. *Domain Adaptation for Medical Image Analysis: A Survey. IEEE Transactions On Biomedical Engineering*, 69(3) (2021), p. 1173-1185. Disponible en <https://doi.org/10.1109/tbme.2021.3117407>.
- [15] Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., Azim, M. A. *Transfer learning: a friendly introduction. Journal of Big Data*, 9(1) (2022). Disponible en <https://doi.org/10.1186/s40537-022-00652-w>.
- [16] ImageNet. *ImageNet*. Disponible en <https://www.image-net.org/>.
- [17] Alzubaidi, L., Fadhel, M. A., Al-Shamma, O., Zhang, J., Santamaría, J., Duan, Y., Oleiwi, S. R. *Towards a Better Understanding of Transfer Learning for Medical Imaging: A Case Study. Applied Sciences*, 10(13) (2020). Disponible en <https://doi.org/10.3390/app10134523>
- [18] Alzubaidi, L., Al-Amidie, M., Al-Asadi, A., Humaidi, A. J., Al-Shamma, O., Fadhel, M. A., Zhang, J., Santamaría, J., Duan, Y. *Novel Transfer Learning Approach for Medical Imaging with Limited Labeled Data. Cancers*, 13(7) (2021), p. 1590. Disponible en <https://doi.org/10.3390/cancers13071590>.
- [19] Alzubaidi, L., Fadhel, M. A., Al-Shamma, O., Zhang, J., Santamaría, J., Duan, Y., Oleiwi, S. R. *Towards a Better Understanding of Transfer Learning for Medical Imaging: A Case Study. Applied Sciences*, 10(13) (2020), p. 4523. Disponible en <https://doi.org/10.3390/app10134523>.
- [20] Alzubaidi, L., Fadhel, M. A., Al-Shamma, O., Zhang, J., Duan, Y. *Deep Learning Models for Classification of Red Blood Cells in Microscopy Images to Aid in Sickle Cell Anemia Diagnosis. Electronics*, 9(3) (2020), p. 427. Disponible en <https://doi.org/10.3390/electronics9030427>.

- [21] Alzubaidi, L., Al-Shamma, O., Fadhel, M. A., Farhan, L., Zhang, J., Duan, Y. *Optimizing the Performance of Breast Cancer Classification by Employing the Same Domain Transfer Learning from Hybrid Deep Convolutional Neural Network Model*. *Electronics*, 9(3) (2020), p. 445. Disponible en <https://doi.org/10.3390/electronics9030445>.
- [22] Heker, M., Greenspan, H. *Joint Liver Lesion Segmentation and Classification via Transfer Learning*. arXiv.org, 26 de abril de 2020. Disponible en <https://arxiv.org/abs/2004.12352>.
- [23] Isola, P., Zhu, J., Zhou, T., Efros, A. A. *Image-to-Image Translation with Conditional Adversarial Networks*. arXiv (Cornell University), 2016. Disponible en <https://doi.org/10.48550/arxiv.1611.07004>.
- [24] Bank, D., Koenigstein, N., Giryes, R. *Autoencoders*. arXiv (Cornell University), 2020. Disponible en <https://doi.org/10.48550/arxiv.2003.05991>.
- [25] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. *Generative adversarial networks*. arXiv (Cornell University), 2014. Disponible en <https://doi.org/10.48550/arxiv.1406.2661>.
- [26] Garrucho, L., Delegue, E., Osuala, R., Kessler, D., Kushibar, K., Díaz, O., Lekadir, K., Igual, L. *Fat-Suppressed Breast MRI Synthesis for Domain Adaptation in Tumour Segmentation*. En *Lecture Notes in Computer Science*, pp. 202-211 (2025). Disponible en https://doi.org/10.1007/978-3-031-77789-9_20.
- [27] Zhu, J., Park, T., Isola, P., Efros, A. A. *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*. arXiv, 30 de marzo de 2017. Disponible en <https://arxiv.org/abs/1703.10593>.
- [28] Modanwal, G., Vellal, A., Mazurowski, M. A. *Normalization of breast MRIs using cycle-consistent generative adversarial networks*. *Computer Methods and Programs in Biomedicine*, 208 (2021), p. 106225. Disponible en <https://doi.org/10.1016/j.cmpb.2021.106225>.
- [29] Muramatsu, C., Nishio, M., Goto, T., Oiwa, M., Morita, T., Yakami, M., Kubo, T., Togashi, K., Fujita, H. *Improving breast mass classification by shared data with domain transformation using a generative adversarial network*. *Computers in Biology and Medicine*, 119 (2020), p. 103698. Disponible en <https://doi.org/10.1016/j.compbmed.2020.103698>.
- [30] Odena, A., Olah, C., Shlens, J. *Conditional image synthesis with auxiliary classifier GANs*. arXiv (Cornell University) (2016). Disponible en <https://doi.org/10.48550/arxiv.1610.09585>.

- [31] Li, Y., Fan, Y., Xu, D., Li, Y., Zhong, Z., Pan, H., Huang, B., Xie, X., Yang, Y., Liu, B. *Deep learning radiomic analysis of DCE-MRI combined with clinical characteristics predicts pathological complete response to neoadjuvant chemotherapy in breast cancer. Frontiers in Oncology*, 12 (2023). Disponible en <https://doi.org/10.3389/fonc.2022.1041142>.
- [32] Syed, A., Adam, R., Ren, T., Lu, J., Maldjian, T., Duong, T. Q. *Machine learning with textural analysis of longitudinal multiparametric MRI and molecular subtypes accurately predicts pathologic complete response in patients with invasive breast cancer. PLoS ONE*, 18, 1 (2023). Disponible en <https://doi.org/10.1371/journal.pone.0280320>.
- [33] Huang, Y., Zhu, T., Zhang, X., Li, W., Zheng, X., Cheng, M., Ji, F., Zhang, L., Yang, C., Wu, Z., Ye, G., Lin, Y., Wang, K. *Longitudinal MRI-based fusion novel model predicts pathological complete response in breast cancer treated with neoadjuvant chemotherapy: a multicenter, retrospective study. EClinicalMedicine*, 58 (2023). Disponible en <https://doi.org/10.1016/j.eclinm.2023.101899>.
- [34] Sutton, E. J., Onishi, N., Fehr, D. A., Dashevsky, B. Z., Sadinski, M., Pinker, K., Martinez, D. F., Brogi, E., Braunstein, L., Razavi, P., El-Tamer, M., Sacchini, V., Deasy, J. O., Morris, E. A., Veeraraghavan, H. *A machine learning model that classifies breast cancer pathologic complete response on MRI post-neoadjuvant chemotherapy. Breast Cancer Research*, 22(1) (2020). Disponible en <https://doi.org/10.1186/s13058-020-01291-w>.
- [35] Guerrero, R. E. D., Carvalho, L., Bocklitz, T., Popp, J., Oliveira, J. L. *A Data Augmentation Methodology to Reduce the Class Imbalance in Histopathology Images. Deleted Journal*, 37(4) (2024), p. 1767-1782. Disponible en <https://doi.org/10.1007/s10278-024-01018-9>.
- [36] Mao, N., Dai, Y., Zhou, H., Lin, F., Zheng, T., Li, Z., Yang, P., Zhao, F., Li, Q., Wang, W., Liang, Y., Xie, H., Ma, H., Zhang, L., Guo, Y., Song, X., Zhang, H., Lu, J. *A multimodal and fully automated system for prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer. Science Advances*, 11(18) (2025). Disponible en <https://doi.org/10.1126/sciadv.adr1576>.
- [37] Kim, S., Lee, J., Cho, N., Kim, Y. (2024). Deep-learning based discrimination of pathologic complete response using MRI in HER2-positive and triple-negative breast cancer. *Scientific Reports*, 14(1). Disponible en <https://doi.org/10.1038/s41598-024-74276-w>.
- [38] Xu, Z., Zhou, Z., Son, J. B., Feng, H., Adrada, B. E., Moseley, T. W., Candelaria, R. P., Guirguis, M. S., Patel, M. M., Whitman, G. J., Leung, J. W. T., Le-Petross, H. T. C., Mohamed, R. M., Panthi, B., Lane, D. L., Chen, H., Wei, P., Tripathy, D., Litton, J. K., ... Ma, J. *Deep Learning Models Based on*

Pretreatment MRI and Clinicopathological Data to Predict Responses to Neoadjuvant Systemic Therapy in Triple-Negative Breast Cancer. Cancers, 17, 6 (2025), p. 966. Disponible en <https://doi.org/10.3390/cancers17060966>.

- [39] D'Anna, A., Aranzulla, C., Carnaghi, C., Caruso, F., Castiglione, G., Grasso, R., Gueli, A. M., Marino, C., Pane, F., Pulvirenti, A., Stella, G. *Comparative analysis of machine learning models for predicting pathological complete response to neoadjuvant chemotherapy in breast cancer: An MRI radiomics approach. Physica Medica*, 131 (2025). Disponible en <https://doi.org/10.1016/j.ejmp.2025.104931>.
- [40] Lv, M., Zhao, B., Mao, Y., Wang, Y., Su, X., Zhang, Z., Wu, J., Gao, X., Wang, Q. *Deep learning model for the early prediction of pathologic response following neoadjuvant chemotherapy in breast cancer patients using dynamic contrast-enhanced MRI. Frontiers In Oncology*, 15 (2025). Disponible en <https://doi.org/10.3389/fonc.2025.1491843>.
- [41] Universitat de Barcelona. *MAMA-MIA - Colección de datos*. Disponible en <https://www.ub.edu/mama-mia/>.
- [42] Newitt, D., Hylton, N., en nombre de I-SPY 1 Network y ACRIN 6657 Trial Team. *Multi-center breast DCE-MRI data and segmentations from patients in the I-SPY 1/ACRIN 6657 trials*. Disponible en <https://doi.org/10.7937/K9/TCIA.2016.HdHpgJLK>.
- [43] Li, W., Newitt, D. C., Gibbs, J., Wilmes, L. J., Jones, E. F. *I-SPY 2 Breast Dynamic Contrast Enhanced MRI Trial (ISPY2)*. Disponible en <https://doi.org/10.7937/TCIA.D8Z0-9T85>.
- [44] Saha, A.; Harowicz, M. R.; Grimm, L. J.; Weng, J.; Cain, E. H.; et al.: *Dynamic contrast-enhanced magnetic resonance images of breast cancer patients with tumor locations*, Disponible en <https://doi.org/10.7937/TCIA.e3sv-re93>
- [45] Saha, A., Harowicz, M. R., Grimm, L. J., Weng, J., Cain, E. H., et al. *Dynamic contrast-enhanced magnetic resonance images of breast cancer patients with tumor locations*. Disponible en [<https://doi.org/10.7937/TCIA.e3sv-re93>]
- [46] Garrucho, L., Kushibar, K., Reidel, C., Joshi, S., Osuala, R., Tsiirikoglou, A., Bobowicz, M., Del Riego, J., Catanese, A., Gwoździewicz, K., Cosaka, M., Abo-Elhoda, P. M., Tantawy, S. W., Sakrana, S. S., Shawky-Abdelfatah, N. O., Salem, A. M. A., Kozana, A., Divjak, E., Ivanac, G., ... Lekadir, K. *A large-scale multicenter breast cancer DCE-MRI benchmark dataset with expert segmentations. Scientific Data*, 12(1) (2025). Disponible en <https://doi.org/10.1038/s41597-025-04707-4>.
- [47] Otsu, N. *A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions On Systems Man And Cybernetics*, 9(1) (1979), p. 62-66. Disponible en <https://doi.org/10.1109/tsmc.1979.4310076>

- [48] Raid, A., Khedr, W., El-Dosuky, M., Aoud, M. *Image Restoration Based on Morphological Operations. International Journal Of Computer Science Engineering And Information Technology*, 4(3) (2014), p. 9-21. Disponible en <https://doi.org/10.5121/ijcseit.2014.4302>.
- [49] Bredies, K., Lorenz, D. *Mathematical image processing. Applied and numerical harmonic analysis*, (2018). Disponible en <https://doi.org/10.1007/978-3-030-01458-2>.
- [50] Braman, N. M., Etesami, M., Prasanna, P., Dubchuk, C., Gilmore, H., Tiwari, P., Plecha, D., Madabhushi, A. *Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast Cancer Research*, 19(1) (2017). Disponible en <https://doi.org/10.1186/s13058-017-0846-1>.
- [51] Kaggle. *Your Machine Learning and Data Science Community*. Disponible en <https://www.kaggle.com>
- [52] DICOM. *DICOM. DICOM Standard*. Disponible en <https://www.dicomstandard.org/>.
- [53] TorchIO. *Augmentation. Torchio Documentation*. Disponible en <https://torchio.readthedocs.io/transforms/augmentation.html>