GRADO EN CIENCIA DE DATOS

Generación sintética de imágenes termográficas de mama mediante modelos generativos para la detección temprana del cáncer

Presentado por:

ALEX DE MARIA ESPIRITUSANTO

Dirigido por:

DR. HÉCTOR ESPINÓS MORATÓ

Agradecimientos

Este trabajo no habría sido posible sin el apoyo constante de las personas que me rodean, quienes no solo me han acompañado durante el desarrollo de este proyecto, sino que han estado presentes a lo largo de toda mi formación como profesional en el ámbito de la ciencia de datos, brindándome ánimo, confianza y fortaleza para continuar.

En primer lugar, deseo expresar mi más profundo agradecimiento a mis padres, Sonia Espiritusanto Latorre y Vicente de María Puchades, por su comprensión, paciencia y por adaptarse a mis horarios y exigencias durante estos cuatro años. Gracias por permitirme dedicarme plenamente a mi carrera académica y profesional, y por ofrecerme siempre sus consejos, cuya sensatez y experiencia han sido fundamentales para tomar decisiones acertadas en los momentos más complejos. Vuestra serenidad ha sido el contrapeso perfecto a mi carácter indeciso e impulsivo.

Asimismo, quiero mostrar mi más sincero agradecimiento a mi tutor, profesor, colega y, sobre todo, mentor, Héctor Espinós. Desde que comenzamos a colaborar en proyectos, al final de mi primer año de carrera, ha estado siempre pendiente de mi evolución, tanto académica como personal. Su implicación, cercanía y compromiso han hecho de él un referente para mí, convirtiéndose en un verdadero modelo a seguir.

Extiendo también mi gratitud a los profesores Alejandro Perdiguero O'Leary y Javier Muñoz de Prat, quienes, al igual que Héctor, han mostrado un compromiso constante con mi desarrollo, orientándome siempre con firmeza y claridad cuando más lo necesitaba. Su implicación ha sido clave en mi crecimiento académico y personal.

Por último, quiero agradecer muy especialmente a mi pareja, Mireia Plana Girbés, por su apoyo incondicional y su infinita paciencia. Gracias por tu ayuda en la revisión de este documento, por recordarme mis responsabilidades y por dar ejemplo cada día con tu constancia y previsión. Admiro profundamente tu capacidad para organizarte, planificar y anticiparte, cualidades que revelan una inteligencia que tú misma subestimas. Gracias por estar siempre a mi lado, incluso cuando ello ha implicado renunciar a tiempo compartido.

Por todo ello, considero imprescindible comenzar este trabajo de fin de grado expresando mi más sincero agradecimiento a quienes han estado a mi lado y han hecho posible la culminación de este proyecto.

Resumen

Este Trabajo de Fin de Grado se enmarca en el proyecto de investigación "Desarrollo y uso clínico de un sistema de diagnóstico precoz y detección precisa de anomalías mamarias con algoritmos de aprendizaje profundo y termografía", financiado por la Cátedra Fundación ASISA – Universidad Europea. El objetivo principal es desarrollar un sistema de diagnóstico asistido que permita la detección temprana del cáncer de mama en regiones con recursos limitados, tomando como caso de estudio real el contexto sanitario de zonas rurales de Ecuador. Ante la dificultad de acceso a tecnologías como la mamografía y la escasez de bases de datos clínicas bien etiquetadas, se propone un enfoque basado en imágenes termográficas combinadas con técnicas de inteligencia artificial. En primer lugar, se emplean Redes Generativas Antagónicas (GANs) para generar imágenes sintéticas que aumenten y diversifiquen el conjunto de datos disponible. Posteriormente, se entrena una Red Neuronal Convolucional (CNN) para clasificar las imágenes en benignas o malignas. El sistema se evalúa mediante métricas estándar como FID, KID y precisión diagnóstica, demostrando que la combinación de GANs y CNNs permite mejorar significativamente el rendimiento en tareas de clasificación médica. Esta solución plantea una alternativa innovadora, portátil y de bajo coste para la mejora del cribado poblacional en entornos con limitaciones estructurales.

Palabras clave: termografía, cáncer de mama, redes neuronales convolucionales, GANs, diagnóstico asistido

Abstract

This bachelor's thesis is part of the research project "Development and clinical use of an early diagnosis system for breast anomalies using deep learning algorithms and thermography," funded by the Fundación ASISA – Universidad Europea Chair. The main goal is to design a computer-aided diagnosis system to enable early breast cancer detection in low-resource settings, using rural areas of Ecuador as a real-world case study. Given the limited access to mammographic screening and the scarcity of well-annotated medical *datasets*, the proposed approach relies on thermographic imaging combined with artificial intelligence techniques. First, Generative Adversarial Networks (GANs) are used to synthetically generate thermographic images to expand and diversify the training *dataset*. Then, a Convolutional

Neural Network (CNN) is trained to classify images as benign or malignant. The system is evaluated through standard metrics such as FID, KID, and diagnostic accuracy, showing that the combined use of GANs and CNNs significantly improves classification performance. This work presents an innovative, portable, and low-cost solution aimed at enhancing breast cancer screening in structurally limited healthcare environments.

Keywords: thermography, breast cancer, convolutional neural networks, GANs, computer-aided diagnosis

Índice

Agradecimientos 2

Resumen 3

Abstract 3

Índice 5

Índice de figuras 8

1. INTRODUCCIÓN 9

2. MOTIVACIÓN Y CONTEXTO 10

3. OBJETIVOS 10

3.1 Objetivo General 10

3.2 Objetivos Específicos 10

4. MÉTODO DE TRABAJO 12

5. ESTADO DEL ARTE 14

5.1 Contexto actual del cáncer de mama 14

5.2 Mamografía: el *golden standard* en el diagnóstico del cáncer de mama 15

5.3 Otras modalidades de diagnóstico por imagen en cáncer de mama 17

5.4 Comparación de metodologías y elección de las termografías 19

5.5 Termografías mamarias 22

5.5.1 Fundamentos clínicos y fisiopatológicos de la termografía 22

5.5.2 Técnicas de captura y protocolos clínicos 23

5.5.3 Ventajas operativas y clínicas de la termografía 23

5.5.4 Limitaciones diagnósticas y retos actuales 24

5.5.5 Integración con inteligencia artificial: el papel de las CNN 24

5.5.6 Bases de datos termográficas disponibles en la literatura 25

5.5.7 Uso de las GAN 26

6. METODOLOGÍA 27

DISCOVER: Identificación del problema y contexto de actuación 27

1. Problemas reales en el contexto de Ecuador 27
2. Causas de los problemas reales 28
3. Planteamiento de la solución final: diagnóstico con instrumentos portátiles 29
4. Descripción del dataset 30
DEFINE: técnicas de *data augmentation* 33
1. Opciones para crear datos sintéticos 33
2. Valoración de las GANs 35
3. Tipos de GANs a valorar 36
4. Selección del modelo GAN para este proyecto 37
5. Resultados esperados 37
DEVELOP: Preprocesamiento de imágenes 37
1. Conversión de matriz térmica a imagen 39
2. Relevancia diagnóstica de anomalías térmicas 39
3. Normalización en función de la temperatura media 40
4. Selección y valoración de arquitecturas CNN 40
DELIVER: Experimentación, ejecución y evaluación 42
1. Diseño experimental: configuración de los modelos 43
2. Entorno de ejecución y limitaciones técnicas 44
3. Evaluación de resultados: métricas y criterios 44
Desarrollo del proyecto 46
7. RESULTADOS Y DISCUSIÓN 48
7.1 Arquitecturas CNN empleadas 48
7.2 Resultados de los modelos CNN entrenados con datos reales 48
7.3 Resultados de las CNN entrenadas con datos aumentados por GAN 51
7.4 Comparación entre CNN y CNN + GAN 54
7.5 Evaluación del cumplimiento de los objetivos 54
Evaluación específica de los objetivos del proyecto 56

7.6 Resultados complementarios: colaboración con el proyecto de investigación 57
Resultados obtenidos en la colaboración 58
Relevancia de estos resultados 59
8. CONCLUSIONES 59
8.1 Conclusión general del TFG 59
8.2 Líneas futuras de investigación 60
8.3 Contribución a los Objetivos de Desarrollo Sostenible (ODS) 62
ODS 3: Salud y bienestar 62
ODS 5: Igualdad de género 62
ODS 9: Industria, innovación e infraestructura 62
ODS 10: Reducción de las desigualdades 63
Referencias 64
Glosario 70

Índice de figuras

Tabla 1:Indicadores de éxito asociados a los objetivos específicos 11

Ilustración 1:Metodología del doble diamante 13
Ilustración 2: Ejemplo de mamografía 16
Ilustración 3:Ejemplo de imagen de ultrasonido 21
Ilustración 4: Imagen termográfica de paciente saludable 31
Ilustración 5:captura de pantalla del repositorio de GitHub 47
Ilustración 6:Matriz de confusión experimento 1.1 49
Ilustración 7:Matriz de confusión experimento 1.2 49
Ilustración 8:Matriz de confusión experimento 1.3 50
Ilustración 9:Matriz de confusión experimento 1.4 51
Ilustración 10:Matriz de confusión experimento 2.1 52
Ilustración 11:Matriz de confusión experimento 2.2 52
Ilustración 12:Matriz de confusión experimento 2.3 53
Ilustración 13:Matriz de confusión experimento 2.4 53

1. INTRODUCCIÓN

El **cáncer de mama** representa una de las principales causas de mortalidad en mujeres a nivel mundial, convirtiéndolo en un problema de salud pública de alta relevancia. Actualmente, las técnicas convencionales de detección incluyen la **mamografía y el ultrasonido**, ambas con limitaciones importantes, especialmente en mujeres jóvenes con tejido mamario denso. En este contexto, la **termografía** ha emergido como una alternativa prometedora debido a su naturaleza no invasiva, no radioactiva, de bajo costo y su potencial para mejorar la precisión en la identificación de lesiones tempranas (Acharya et al., 2012).

Uno de los principales retos en la aplicación de inteligencia artificial al diagnóstico médico mediante imágenes termográficas es la **escasa disponibilidad de datos bien etiquetados**. Esta limitación compromete el rendimiento de los modelos de aprendizaje profundo, que requieren grandes volúmenes de datos para su entrenamiento efectivo (Liu et al., 2019).

Ante esta situación, este trabajo propone un enfoque basado en **dos etapas complementarias**: en primer lugar, se emplearán **Redes Generativas Antagónicas (GANs)** para la generación sintética de imágenes termográficas con el fin de **aumentar y diversificar el conjunto de datos** disponible (Goodfellow et al., 2014; Santini et al., 2020). Esta técnica permitirá mitigar el desequilibrio de clases y la escasez de muestras médicas reales.

En una segunda etapa, se utilizará una **Red Neuronal Convolucional (CNN)** como clasificador para identificar lesiones benignas y malignas en las imágenes generadas y reales. Esta arquitectura ha demostrado un alto rendimiento en tareas de clasificación de imágenes médicas (Krizhevsky et al., 2012), y su aplicación permitirá evaluar el impacto de la *data augmentation* en la mejora del diagnóstico automatizado. Para contrastar resultados, se entrenará una red neuronal con la mayor cantidad de datos a la que se tenga acceso y, posteriormente, con los datos generados por la GAN. De este modo, podremos saber la ratio de mejora que implica usar técnicas punteras de augmento de datos en el entrenamiento de redes neuronales.

Cumpliendo este objetivo, la combinación de GANs y CNNs podría representar un paso relevante hacia el desarrollo de sistemas de diagnóstico asistido más accesibles, precisos y aplicables incluso en entornos con recursos limitados.

2. MOTIVACIÓN Y CONTEXTO

Este TFG se enmarca en el proyecto de investigación titulado "Desarrollo y uso clínico de un sistema de diagnóstico precoz y detección precisa de anomalías mamarias con algoritmos de aprendizaje profundo y termografía", aprobado en la Convocatoria 2024 de la Cátedra Fundación ASISA – Universidad Europea. La investigación está orientada a resolver un problema real y urgente**: la limitada accesibilidad a métodos de diagnóstico del cáncer de mama en zonas rurales y periféricas de Ecuador**, debido a factores como la pobreza estructural, la geografía dispersa y la escasa infraestructura sanitaria especializada.

Como integrante activo del equipo de investigación, participo directamente en el desarrollo, implementación y testeo de algoritmos de aprendizaje profundo, especialmente en la generación sintética de datos con modelos GAN y en la evaluación de clasificadores basados en redes CNN. Por tanto, este TFG no solo representa una propuesta individual, sino una **línea de trabajo aplicada, derivada directamente del proyecto general**, y validada mediante herramientas computacionales reales, con resultados medibles y objetivos técnicos definidos.

3. OBJETIVOS

Objetivo General

El objetivo central de esta investigación es d**esarrollar un sistema portátil y autónomo que permita realizar cribado mamario en regiones rurales y periféricas** sin depender de la presencia constante de especialistas o de recursos hospitalarios avanzados. Para ello se propone diseñar y evaluar un sistema de diagnóstico asistido para la detección temprana del cáncer de mama mediante la generación sintética de imágenes termográficas con **Redes Generativas Antagónicas** (GANs) y su posterior clasificación automática utilizando una **Red Neuronal Convolucional**, con el fin de mejorar el rendimiento diagnóstico frente a la escasez de datos médicos etiquetados.

Objetivos Específicos

1. **Recolectar y preprocesar un conjunto representativo de imágenes** termográficas de mama provenientes de bases de datos públicas, asegurando una correcta segmentación entre casos benignos y malignos.

- *Indicador de éxito:* Se conforma un *dataset* balanceado con al menos 500 imágenes válidas (reales y sintéticas), correctamente etiquetadas.

2. **Implementar una arquitectura GAN** para la generación sintética de imágenes termográficas que permita aumentar la diversidad del *dataset* y paliar el desequilibrio de clases.
   - *Indicador de éxito:* Las imágenes generadas alcanzan una puntuación FID ≤ 40 y KID ≤ 0.1, evidenciando alta similitud con las imágenes reales.

3. **Entrenar y evaluar una CNN** para la clasificación de las imágenes aumentadas, comparando el rendimiento frente a un modelo entrenado solo con datos reales.
   - *Indicador de éxito:* El modelo CNN entrenado **con datos aumentados** supera en al menos el 75% la precisión respecto al conjunto de test.

A continuación, se resumen los objetivos específicos del proyecto junto con sus respectivos indicadores de éxito, umbrales esperados y herramientas de evaluación:

Tabla 1:Indicadores de éxito asociados a los objetivos específicos

Objetivo	**Indicador de Éxito**	**Umbral Esperado**	**Herramienta de Evaluación**
GAN	FID / KID	FID ≤ 40, KID ≤ 0.1	Librerías de evaluación (TensorFlow, PyTorch)
CNN	Precisión	+75% de precisión	Matriz de confusión, AUC
Dataset	Nº de imágenes	≥ 500 imágenes bien etiquetadas	Revisión manual + script validación

4. MÉTODO DE TRABAJO

Este proyecto se estructura siguiendo la **metodología del Doble Diamante**, una estrategia utilizada para abordar problemas complejos mediante procesos iterativos de exploración y definición. El modelo fue formalizado como marco metodológico para la innovación por el Design Council y se ha consolidado como una herramienta clave en el diseño centrado en el usuario y en la investigación aplicada (Brown, 2009).

El Doble Diamante se divide en **cuatro fases**: *Discover* (descubrir), *Define* (definir), *Develop* (desarrollar) y *Deliver* (entregar). Este modelo representa un proceso de divergencia y convergencia, donde en las fases iniciales se amplía la comprensión del problema y se exploran soluciones, mientras que en las fases siguientes se concreta y se implementan las opciones más prometedoras.

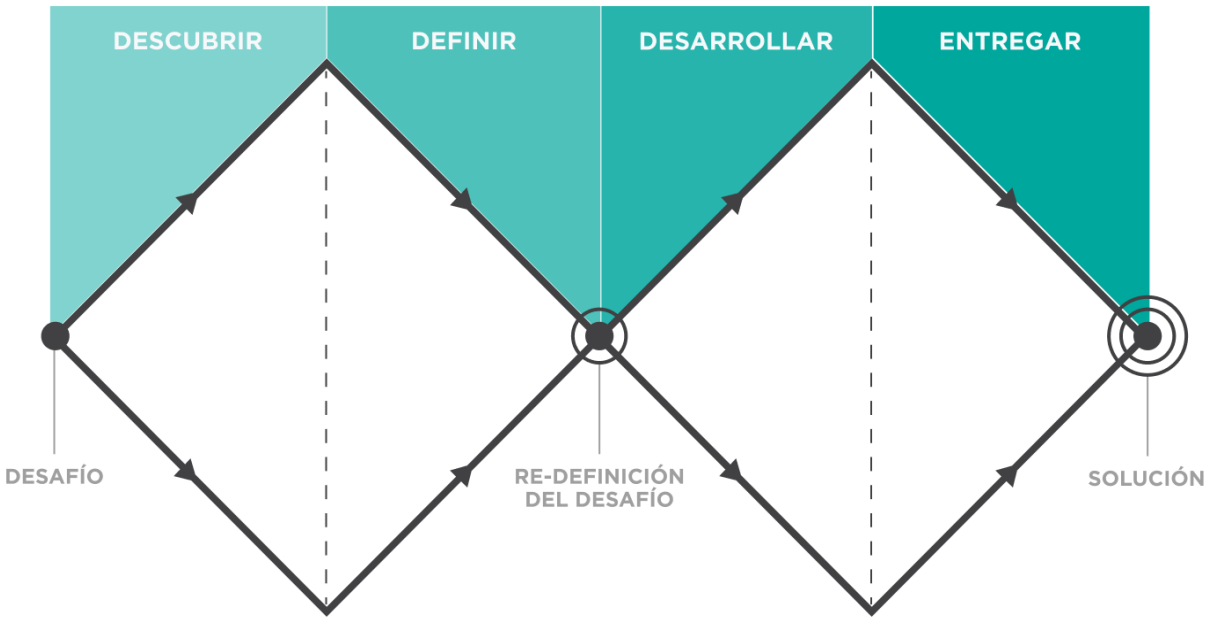
A continuación, se describe cómo se aplica cada fase al contexto del TFG:

1. **Descubrir (*Discover*):** En esta etapa inicial se lleva a cabo una investigación profunda del problema desde una perspectiva tanto tecnológica como sanitaria. Se analiza el **acceso desigual a programas de cribado del cáncer de mama**, especialmente en zonas rurales o con escasos recursos, y se revisan los métodos de diagnóstico convencionales como la **mamografía**, **ultrasonido** y **resonancia magnética**, destacando sus limitaciones frente al uso de **termografía**. También se realiza una exploración del estado actual del uso de **inteligencia artificial en imágenes médicas**, identificando la escasa disponibilidad de ***datasets* termográficos etiquetados** como el principal obstáculo para aplicar modelos de aprendizaje profundo eficaces en este contexto. Se revisan las técnicas actuales de diagnóstico y se identifican sus limitaciones, prestando especial atención a la escasez de bases de datos termográficas etiquetadas.

2. **Definir (*Define*):** Con la información recopilada, se delimita el **problema central** que el proyecto busca abordar: la viabilidad del uso de termografía junto con modelos GAN y CNN para generar herramientas de detección accesibles y precisas. Se define el enfoque técnico y científico, así como los objetivos específicos del proyecto y se llega a la conclusión de la necesidad de entrenar la CNN con los primeros datos conseguidos optimizando al máximo su rendimiento para así poder comparar la posible mejora con el resultado del entrenamiento de la misma red con la ingesta de datos generados por la GAN.

3. **Desarrollar (*Develop*):** En esta fase se implementan las **soluciones propuestas**. Se entrena una CNN y se itera sobre ella para optimizar sus resultados y posteriormente se entrena una GAN para realizar *data augmentation* sobre un *dataset* termográfico limitado y, posteriormente, se entrena de nuevo la CNN sobre el conjunto aumentado. Se iteran pruebas y ajustes para optimizar los resultados.
4. **Entregar (*Deliver*):** Finalmente, se presentan y analizan los **resultados obtenidos**, evaluando la calidad de las imágenes generadas y la efectividad del modelo de clasificación. Esta fase incluye la validación de los indicadores de éxito definidos y la reflexión sobre el impacto y la aplicabilidad del sistema propuesto.

Ilustración 1:Metodología del doble diamante



5. ESTADO DEL ARTE

Contexto actual del cáncer de mama

El cáncer de mama es la neoplasia más frecuentemente diagnosticada en mujeres a nivel mundial. Los datos globales más actualizados provienen del informe **GLOBOCAN 2022**, elaborado por el Centro Internacional de Investigaciones sobre el Cáncer (IARC), donde se registraron aproximadamente **2,3 millones de casos nuevos** en mujeres, que es el **11,6 %** de todos los cánceres femeninos a nivel mundial, y alrededor de **670 000 fallecimientos** atribuibles a esta enfermedad. La tasa de mortalidad global se situó entre **13 y 15 por cada 100 000 mujeres al año**, aunque con grandes diferencias según el nivel de desarrollo del país (Bray et al., 2022). En regiones con alto Índice de Desarrollo Humano (IDH), la **incidencia supera los 100 casos por cada 100 000 mujeres**, mientras que en países con bajo IDH, aunque la incidencia puede ser inferior a 10/100 000, la **mortalidad relativa es significativamente mayor** debido a deficiencias en el acceso al diagnóstico precoz y a tratamientos efectivos (Arzanova & Mayrovitz, 2022).

Aunque hasta la fecha no se han publicado nuevos datos oficiales más allá del ciclo de GLOBOCAN 2022, **estimaciones recientes de la IARC** y otros organismos internacionales indican que, de mantenerse los patrones actuales, se proyecta que para **el año 2050** se alcanzarán aproximadamente **3,2 millones de casos nuevos anuales** (un aumento del **38 %**) y más de **1,1 millones de muertes por cáncer de mama al año** (un incremento del **68 %**) en todo el mundo (IARC, 2025; Nature Medicine, 2025).

En Europa Occidental, y concretamente en España, esta tendencia es palpable: la incidencia en 2022 alcanzó los **34 750 casos nuevos reportados** (Red de Cáncer de Mama [REDECAN]; Brunet et al., 2023). No obstante, la mortalidad se ha estabilizado en torno a 6 600–6 800 defunciones anuales, lo que equivale a una tasa de 12–15 por 100 000 mujeres, contrastando positivamente con muchos países de renta media (Brunet et al., 2023). Este escenario favorable se atribuye a una implementación eficiente de programas de *screening*, terapias dirigidas, y, sobre todo, a **un seguimiento clínico avanzado**, lo que se refleja en una **supervivencia a cinco años observada del 82 %** (Brunet et al., 2023; Dafni, Tsourti & Alatsathianos, 2019).

En América Latina, el impacto sanitario del cáncer de mama es significativo y creciente. Según Miranda et al. (2012), en la región se detectan anualmente alrededor de 114 900 nuevos

casos y 37 000 muertes asociadas, reflejando una de las cargas **más altas entre las neoplasias femeninas**. En países como Ecuador, Colombia, Perú o México, las **tasas de mortalidad** ajustadas por edad oscilan en torno a **10–12 por 100 000 mujeres**, lo que resalta la falta de detección precoz y las desigualdades en el acceso al tratamiento (Miranda et al., 2012).

Específicamente en Ecuador, los registros del comportamiento epidemiológico del cáncer de mama muestran **un aumento sostenido de la incidencia** desde los años ochenta hasta 2013, con aproximadamente **47 casos por cada 100 000 mujeres** en Quito en 2017 (Miranda et al., 2012). Tanto la incidencia como la mortalidad han ido creciendo, aunque el índice de mortalidad/incidencia (MIR) ha mejorado ligeramente gracias a ciertos avances en el sistema sanitario (Miranda et al., 2012; Miranda et al., 2012). No obstante, **persisten tasas de fallecimiento que rondan los 10–12 por 100 000**, y una supervivencia a cinco años estimada en un 79 % en Quito (Miranda et al., 2012).

La comparación entre escenarios globales, europeos y latinoamericanos detalla una imagen clara: los países desarrollados, incluyendo España, presentan una **incidencia alta** y mortalidad estabilizada o en descenso gracias a sistemas sanitarios consolidados. En cambio, en países como Ecuador, pese a incidencias moderadas**, la mortalidad sigue en niveles preocupantes**. Esto revela la urgencia de fortalecer programas de cribado, garantizar la equidad en el acceso a diagnósticos de alta resolución y tratamientos especializados, y mejorar la conciencia pública sobre la enfermedad.

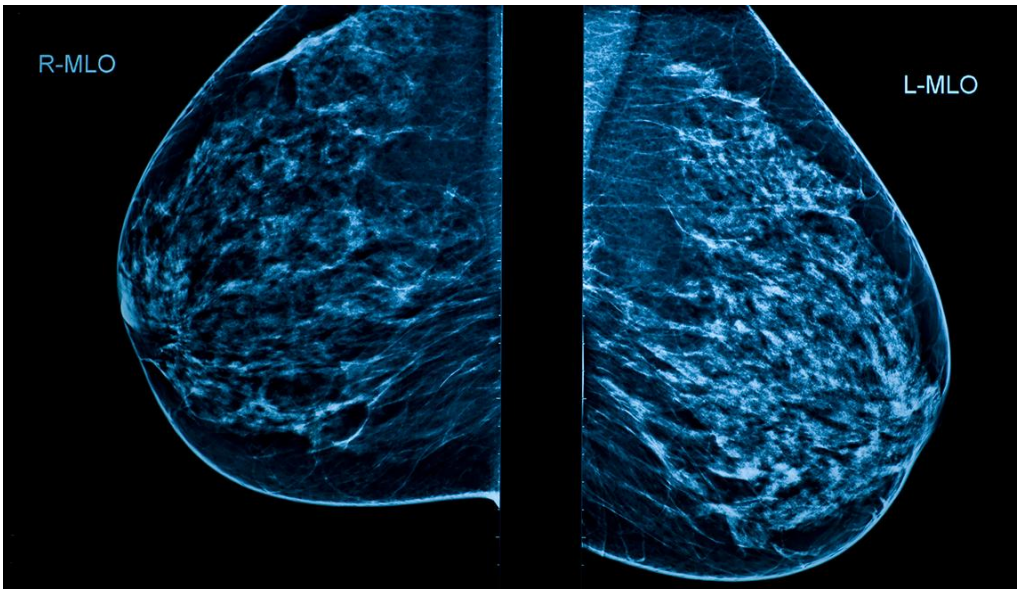
Mamografía: el *golden standard* en el diagnóstico del cáncer de mama

La mamografía es, desde hace décadas, el **método diagnóstico por imagen** más ampliamente utilizado para la detección temprana del cáncer de mama, y **está considerada como el "estándar de oro"** en los programas de cribado poblacional, especialmente en mujeres asintomáticas de entre 40 y 74 años. Su funcionamiento se basa en la utilización de rayos X de baja dosis para obtener **imágenes bidimensionales de los tejidos internos de la mama**. Estas imágenes permiten visualizar con alta precisión estructuras anómalas como masas, calcificaciones o distorsiones arquitectónicas del parénquima mamario, que pueden ser indicativas de procesos malignos incipientes (Saslow et al., 2016; Nelson et al., 2016).

En una mamografía convencional, **la mama es comprimida entre dos placas** durante unos segundos para aplanar el tejido y minimizar la dispersión de la radiación, mejorando así la resolución de la imagen y reduciendo la dosis necesaria. **Esta compresión puede causar**

molestias, pero es un paso esencial para garantizar una imagen diagnóstica de calidad. El procedimiento permite obtener, generalmente, dos proyecciones por mama: una **craneocaudal (de arriba abajo)** y otra **medio-lateral oblicua**, que ofrece una visualización más completa de los cuadrantes superiores e inferiores (Skaane et al., 2013).

Ilustración 2: Ejemplo de mamografía



Desde el punto de vista clínico, la mamografía tiene una **sensibilidad estimada en torno al 85–90 % en mujeres posmenopáusicas**, y una especificidad también elevada, lo que la convierte en una **herramienta eficaz para reducir la mortalidad** por cáncer de mama a través del diagnóstico precoz (Myers et al., 2015). Sin embargo, su **rendimiento disminuye significativamente en mujeres jóvenes** o en aquellas con mamas densas, ya que el tejido fibroglandular denso puede enmascarar lesiones, reduciendo la sensibilidad a valores inferiores al 60 % en algunos casos (Boyd et al., 2007).

La eficacia de la mamografía ha sido avalada por **múltiples estudios epidemiológicos** de gran escala. Por ejemplo, los metaanálisis revisados por la *U.S. Preventive Services Task Force* (USPSTF) muestran una **reducción significativa de la mortalidad** por cáncer de mama, estimada entre el 15 % y el 30 %, especialmente cuando se implementa en programas sistemáticos de cribado con una periodicidad bien establecida (USPSTF, 2016). Por este motivo, instituciones como la OMS, la *American Cancer Society* y los sistemas públicos de salud europeos **la recomiendan como primera línea de cribado**, siendo una técnica de alta disponibilidad en la mayoría de los países de ingresos medios y altos (WHO, 2021).

En los últimos años, la evolución hacia la **mamografía digital directa** ha permitido **mejorar la calidad de imagen** y reducir errores de interpretación gracias a herramientas de procesamiento avanzado y almacenamiento electrónico. Asimismo, **la combinación de mamografía con técnicas de inteligencia artificial está ganando terreno en la práctica clínica**. Algoritmos basados en redes neuronales convolucionales (CNNs) están siendo entrenados para **detectar microcalcificaciones, asimetrías y masas** con un rendimiento comparable al de radiólogos expertos, lo que abre la puerta a su integración en sistemas de doble lectura o triaje automático (Rodriguez-Ruiz et al., 2019).

No obstante, la mamografía también presenta **limitaciones relevantes**. Implica **exposición a radiación ionizante**, aunque en dosis bajas, y puede **resultar incómoda o dolorosa para algunas pacientes**. Además, en entornos con escasos recursos tecnológicos o humanos —como las zonas rurales de países en desarrollo—, su implementación se ve obstaculizada por el **alto coste de los equipo**s, la necesidad de personal técnico cualificado y la infraestructura necesaria para operar dispositivos de rayos X, lo que restringe su disponibilidad fuera de centros urbanos (Lehman et al., 2016).

Por lo tanto, podemos decir que la mamografía sigue siendo la herramienta **más validada para el cribado** y la detección temprana del cáncer de mama en la práctica clínica actual. Su fiabilidad, disponibilidad en sistemas de salud desarrollados y sólido respaldo empírico la consolidan como **el estándar frente al cual deben compararse otras modalidades de imagen**. Sin embargo, en contextos donde las condiciones de infraestructura y personal cualificado son limitadas, como es el caso de ciertas regiones rurales de Ecuador, es **necesario explorar alternativas diagnósticas** que mantengan un equilibrio razonable entre precisión, accesibilidad y coste.

Otras modalidades de diagnóstico por imagen en cáncer de mama

Además de la mamografía, existen **diversas modalidades de imagen** que complementan o, **en algunos casos, sustituyen su uso** en contextos específicos del diagnóstico del cáncer de mama. Estas técnicas varían en términos de sensibilidad, especificidad, coste, accesibilidad, y grado de invasión personal, lo que condiciona su aplicación clínica y su viabilidad en sistemas sanitarios con recursos limitados. Asimismo, cada una de estas modalidades ha sido **objeto de aplicaciones específicas de inteligencia artificial**, que han mejorado su rendimiento diagnóstico en los últimos años.

La **ecografía mamaria**, también conocida como ultrasonido, es una técnica no invasiva que utiliza ondas sonoras de alta frecuencia para generar imágenes en tiempo real del tejido mamario. Es especialmente útil como complemento a la mamografía en mujeres con mamas densas o en el estudio de masas palpables no visibles radiológicamente (Berg et al., 2012). Su principal ventaja es la ausencia de radiación ionizante y su bajo coste, lo que permite su uso repetido sin efectos adversos. Sin embargo, presenta una alta dependencia del operador, lo que puede generar variabilidad en la interpretación y limitaciones en la estandarización del diagnóstico (Kelly et al., 2010). Desde el punto de vista del aprendizaje automático, la ecografía ha sido tradicionalmente abordada mediante técnicas como los ***Support Vector Machines*** **(SVM)** y ***Random Forests*** para la segmentación de lesiones y clasificación tumoral (Zhou et al., 2020). En los últimos años, el desarrollo de redes neuronales convolucionales (CNNs) ha permitido automatizar la detección de características morfológicas complejas, alcanzando niveles de precisión cercanos a los de expertos clínicos (Byra et al., 2018).

Otra modalidad de alto valor diagnóstico es la **resonancia magnética mamaria (MRI)**, que ofrece una sensibilidad superior al 95 %, particularmente en mujeres jóvenes con antecedentes familiares de cáncer de mama o mutaciones genéticas predisponentes (Kuhl et al., 2017). Su capacidad para detectar lesiones ocultas en mamografía o ecografía la convierte en una herramienta insustituible en pacientes de alto riesgo. No obstante, presenta limitaciones importantes: elevado coste, baja disponibilidad, necesidad de contraste intravenoso (gadolinio) y largos tiempos de adquisición. Desde el punto de vista de la inteligencia artificial, la MRI ha sido objeto de múltiples estudios con redes profundas (CNNs, U-Net, y más recientemente Transformers), orientados tanto a la segmentación de lesiones como a la clasificación del riesgo tumoral (Yala et al., 2019; Zhou et al., 2021). Algunos sistemas de aprendizaje profundo multimodal incluso combinan MRI con datos clínicos para mejorar la estratificación de pacientes.

La **mamografía con contraste** (CEM, por sus siglas en inglés) ha surgido como una alternativa más accesible a la MRI en ciertos contextos clínicos. Utiliza medios de contraste yodados para resaltar la vascularización anómala asociada a procesos tumorales, y se ha demostrado que posee una sensibilidad comparable a la resonancia en la detección de lesiones malignas (Lewin et al., 2019). Aunque requiere administración intravenosa, su protocolo es más breve, su coste más bajo y su infraestructura más común que la MRI. En el campo de la IA, se han desarrollado modelos de aprendizaje supervisado para optimizar la interpretación

automática de las imágenes generadas por CEM, aunque su aplicación aún está menos extendida en comparación con la mamografía convencional o la MRI.

Otra técnica emergente es el **PET mamario (PEM)**, una variante de la tomografía por emisión de positrones centrada exclusivamente en la glándula mamaria. Este método detecta la captación de glucosa radiomarcada por células tumorales activas, ofreciendo alta especificidad en la evaluación de recurrencias o estadificación del cáncer (Spick & Baltzer, 2014). Su principal limitación radica en la exposición a altas dosis de radiación, lo que restringe su uso a casos específicos. En IA, los modelos aplicados en este campo han sido principalmente redes densas y CNNs profundas para correlacionar patrones de captación con histopatología (Shen et al., 2019).

La **Imagen Molecular Mamaria (MBI)** es otra técnica basada en radio trazadores que permite visualizar lesiones no detectables mediante mamografía, especialmente en mujeres con mamas densas. Aunque mejora la sensibilidad, implica también un aumento considerable en la dosis de radiación absorbida, lo que limita su repetibilidad (Rhodes et al., 2015). Su uso se ha extendido en contextos de cribado adicional y se han implementado algoritmos de segmentación automática y clasificación para optimizar el rendimiento de los estudios.

Finalmente, existen modalidades experimentales o en fases tempranas de validación clínica. Entre ellas, destaca la **tomografía por impedancia eléctrica (EIT)**, una técnica no invasiva que mide las diferencias de conductividad eléctrica en los tejidos mamarios. Aunque su sensibilidad puede superar el 85 % cuando se combina con mamografía, su especificidad aún es baja y no ha sido incorporada en protocolos clínicos estandarizados (Soni et al., 2021). La IA aplicada a EIT se basa en técnicas de reconstrucción tomográfica y clasificación supervisada. Por otro lado, la **CT láser mamaria (CTLM)** emplea luz infrarroja para mapear la angiogénesis tumoral; su uso clínico aún es limitado, pero representa una alternativa interesante al combinar portabilidad con ausencia de radiación (Xu et al., 2020).

Comparación de metodologías y elección de las termografías

A la luz de las modalidades diagnósticas expuestas en el punto anterior, resulta evidente que la mayoría de las tecnologías de imagen empleadas en la detección del cáncer de mama implican barreras importantes en términos de infraestructura, coste, invasividad y necesidad de personal especializado. Aunque todas ellas cuentan con una sólida validación clínica, **su**

implementación efectiva en contextos con recursos limitados, como las zonas rurales de Ecuador, se enfrenta a múltiples desafíos logísticos y económicos.

La mamografía, por ejemplo, requiere **equipamiento radiológico complejo**, instalaciones sanitarias con condiciones físicas específicas y personal técnico altamente capacitado. A esto se suma el hecho de que implica exposición a radiación ionizante, lo cual, si bien aceptable en términos de riesgo-beneficio en programas de cribado estructurados, puede suponer una barrera psicológica o cultural en determinadas poblaciones (Lehman et al., 2016).

La **resonancia magnética**, por su parte, presenta un rendimiento diagnóstico superior, especialmente en mujeres jóvenes con mamas densas, pero su elevado coste de adquisición y mantenimiento, la necesidad de medios de contraste intravenosos y la baja portabilidad la hacen inviable para su uso en campañas móviles o en comunidades alejadas de núcleos urbanos (Kuhl et al., 2017).

Otras técnicas como **la imagen molecular mamaria** añaden una capa de complejidad aún mayor, al requerir radio trazadores y protocolos de seguridad radiológica avanzados. Aunque estas tecnologías ofrecen ventajas clínicas en escenarios muy concretos, su aplicabilidad en territorios con déficit estructural en salud es prácticamente nula, tanto por el coste como por la dependencia de personal altamente cualificado y condiciones clínicas controladas (Spick & Baltzer, 2014).

En conjunto, puede afirmarse que todas estas modalidades, si bien válidas en términos diagnósticos, **resultan inadecuadas para los objetivos** logísticos y sanitarios de este trabajo, centrado en la creación de **un sistema autónomo, portátil y de bajo coste** para la detección precoz de cáncer de mama.

Con este escenario como telón de fondo, se hace necesario optar por modalidades de imagen que sean no solo eficaces desde el punto de vista clínico, sino también viables desde una perspectiva operativa y económica. Bajo este criterio, únicamente dos tecnologías satisfacen simultáneamente las condiciones de no invasividad, portabilidad, autonomía y bajo coste: **el ultrasonido y la termografía infrarroja.**

El ultrasonido es, sin duda, una herramienta útil y consolidada, ampliamente utilizada como complemento a la mamografía, especialmente en mujeres con tejido mamario denso. **Su carácter no invasivo**, su precio relativamente bajo y la disponibilidad de versiones portátiles la hacen atractiva para escenarios rurales. Sin embargo, su principal limitación radica en la

fuerte dependencia del operador. Tanto la captura de la imagen como su interpretación requieren habilidades clínicas específicas, y la calidad diagnóstica de la ecografía varía significativamente según la experiencia del técnico (Kelly et al., 2010; Byra et al., 2018). Esto introduce una barrera crítica en el contexto de este proyecto, que busca minimizar la intervención humana en la interpretación inicial de las imágenes mediante el uso de inteligencia artificial.

Ilustración 3:Ejemplo de imagen de ultrasonido

Frente a estas limitaciones**, la termografía infrarroja** se presenta como una alternativa especialmente adecuada para los fines del presente trabajo. Esta técnica, basada en la captación de las variaciones térmicas superficiales de la piel, no requiere contacto físico con la paciente, es completamente libre de radiación y puede llevarse a cabo mediante dispositivos compactos y fácilmente transportables. Lo más relevante en este caso es que la captura de la imagen térmica **no exige conocimientos clínicos avanzados**, ya que el procedimiento consiste simplemente en posicionar la paciente ante la cámara en condiciones controladas y registrar la imagen en unos pocos segundos (Arora et al., 2008; Head & Elliott, 2012). Este carácter operativo permite delegar la adquisición de imágenes en personal sanitario generalista o técnicos básicos, lo que **facilita su despliegue en campañas móviles** y en territorios con escaso acceso a radiólogos o técnicos especialistas.

Además, el hecho de que las imágenes térmicas puedan **ser interpretadas por algoritmos de inteligencia artificial** entrenados para detectar patrones asociados a patologías malignas refuerza aún más su idoneidad para este tipo de sistemas autónomos.

Por todo lo anterior, la elección de la **termografía como modalidad de imagen principal en este proyecto** no responde únicamente a criterios técnicos, sino también a una necesidad de adaptabilidad al **contexto real** en el que se pretende intervenir. La facilidad de uso, el bajo coste, la posibilidad de automatización completa y la no necesidad de infraestructura sanitaria compleja convierten a esta técnica en la única opción viable para cumplir con el objetivo central de esta investigación: d**esarrollar un sistema portátil y autónomo que permita realizar cribado mamario en regiones rurales y periféricas** sin depender de la presencia constante de especialistas o de recursos hospitalarios avanzados.

Termografías mamarias

Fundamentos clínicos y fisiopatológicos de la termografía

La termografía médica se basa en la captación de la radiación infrarroja emitida por la superficie corporal, transformada en imágenes que reflejan el mapa térmico de una región anatómica concreta. En el caso del cáncer de mama, esta modalidad se fundamenta en una premisa fisiopatológica bien documentada: los tumores malignos inducen una serie de cambios en el microambiente mamario —como angiogénesis, hipermetabolismo celular y fenómenos inflamatorios— que elevan la temperatura local del tejido (Ng, 2009). Estas alteraciones térmicas, aunque no siempre visibles en superficie, generan gradientes térmicos detectables mediante cámaras térmicas sensibles en el rango de longitudes de onda del infrarrojo lejano (8–14 μm).

En condiciones ideales, una imagen térmica de la mama sana presenta una distribución simétrica y homogénea de temperatura. La presencia de asimetrías térmicas, puntos calientes o regiones de hipervascularización puede indicar la existencia de un proceso patológico subyacente. Este principio se explota clínicamente para detectar alteraciones sospechosas, incluso antes de que sean detectables por otros métodos estructurales como la mamografía (Head et al., 2023). Sin embargo, no todos los cambios térmicos están asociados a tumores: factores hormonales, infecciones, traumatismos o incluso alteraciones vasculares benignas pueden también modificar el patrón térmico, lo que introduce desafíos en términos de especificidad diagnóstica.

Técnicas de captura y protocolos clínicos

La adquisición de imágenes térmicas en entornos clínicos exige condiciones estrictamente controladas. Las guías de buenas prácticas elaboradas por entidades como la American Academy of Thermology (AAT) y la International Academy of Clinical Thermology (IACT) establecen que la paciente debe aclimatarse durante un periodo mínimo de 10 a 15 minutos en una sala sin corrientes de aire, con temperatura ambiental regulada entre 20 y 23 °C y humedad relativa estable (Ring & Ammer, 2012). La ausencia de maquillaje, cremas, desodorantes o ropa ajustada en la zona torácica es también esencial para evitar interferencias térmicas. Las capturas se realizan habitualmente desde tres ángulos (frontal, oblicuo izquierdo y oblicuo derecho), a una distancia estandarizada entre cámara y paciente, con la finalidad de obtener imágenes reproducibles y comparables entre estudios.

Los sistemas actuales permiten registrar imágenes de alta resolución térmica y espacial. Cámaras como las FLIR T660 o Testo 890 ofrecen sensibilidades inferiores a 0.03 °C y resoluciones de hasta 640x480 píxeles. Estos parámetros son críticos, ya que permiten detectar diferencias térmicas sutiles en el rango subclínico. La correcta calibración del sensor térmico y la uniformidad de la respuesta radiométrica a lo largo del sensor influyen directamente en la calidad de la imagen generada y, por tanto, en la fiabilidad de los algoritmos que posteriormente analizan dicha imagen (Head et al., 2023).

Ventajas operativas y clínicas de la termografía

Una de las principales ventajas de la termografía frente a otras modalidades es su carácter completamente no invasivo. No requiere radiación, contacto físico, ni la administración de medios de contraste. Esto no solo la convierte en una técnica segura para la paciente, sino también en una herramienta idónea para contextos donde se pretende realizar cribado poblacional, especialmente en campañas móviles o rurales, como las contempladas en este trabajo. A diferencia de otras técnicas como la ecografía, su operativa es mucho más simple y puede ser realizada por personal no especializado tras una formación básica.

Otra ventaja clave reside en la posibilidad de analizar las imágenes térmicas mediante inteligencia artificial, lo que permite automatizar completamente el proceso diagnóstico inicial. Esta autonomía tecnológica reduce la carga sobre los sistemas sanitarios y elimina la necesidad de tener radiólogos o técnicos expertos en cada punto de atención. Además, al tratarse de una técnica pasiva, las imágenes pueden repetirse con frecuencia sin riesgo acumulado para la

paciente, lo que permite su uso en programas de seguimiento o vigilancia activa (Arora et al., 2008).

Limitaciones diagnósticas y retos actuales

A pesar de su potencial, la termografía presenta también limitaciones importantes. Su sensibilidad suele ser alta, pero la especificidad varía mucho en función del protocolo utilizado y del sistema de análisis implementado. El hecho de que múltiples procesos no tumorales puedan generar focos térmicos plantea el riesgo de falsos positivos, especialmente cuando la interpretación depende del juicio visual humano. Además, la distribución térmica natural de las mamas varía entre pacientes y puede estar influida por factores como la edad, el ciclo menstrual o la actividad física previa, lo que dificulta la estandarización (Ng, 2009).

Desde el punto de vista tecnológico, uno de los principales retos es la escasez de bases de datos públicas, amplias y anotadas clínicamente que permitan entrenar algoritmos robustos de clasificación. Muchos estudios existentes se basan en *datasets* pequeños, capturados con diferentes protocolos, y sin una estandarización clara de etiquetas clínicas. Esto limita la reproducibilidad y la capacidad de generalización de los modelos desarrollados, dificultando su aplicación real a gran escala (Lahiri et al., 2012).

Integración con inteligencia artificial: el papel de las CNN

El verdadero renacimiento de la termografía en el ámbito médico ha venido de la mano del aprendizaje profundo. Las redes neuronales convolucionales (CNN) han demostrado ser especialmente eficaces en el análisis de imágenes térmicas debido a su capacidad para detectar patrones complejos de textura, asimetría y gradientes térmicos sin necesidad de segmentación o extracción manual de características. A diferencia de los modelos clásicos basados en reglas o SVM, las CNN aprenden directamente de los datos, lo que permite identificar correlaciones invisibles al ojo humano.

Estudios como el de Acharya et al. (2020) han entrenado arquitecturas profundas sobre conjuntos de imágenes térmicas etiquetadas, logrando precisiones superiores al 90 % en la clasificación de imágenes entre benignas y malignas. Otros trabajos recientes, como el de Mehta et al. (2021), han empleado modelos más complejos con capas de atención y transferencia de aprendizaje, mejorando tanto la sensibilidad como la especificidad, incluso con bases de datos limitadas. No obstante, estos modelos siguen enfrentando el problema de la

generalización: entrenados sobre datos homogéneos, su rendimiento tiende a caer en escenarios clínicos reales con variabilidad demográfica y técnica.

A nivel práctico, la implementación de estos modelos en sistemas portátiles es técnicamente viable. Una vez entrenado, el modelo puede integrarse en un dispositivo compacto que capte, procese y clasifique la imagen térmica en tiempo real. Esto abre la puerta a dispositivos autónomos de cribado que podrían utilizarse en unidades móviles, centros de salud rurales o incluso en programas comunitarios, sin requerir intervención humana más allá de la colocación de la paciente y la activación del sistema.

Bases de datos termográficas disponibles en la literatura

El desarrollo de sistemas de diagnóstico automático basados en termografía mamaria requiere, como condición esencial, el acceso a bases de datos bien documentadas, con una cantidad suficiente de muestras y una adecuada clasificación clínica. En los últimos años han surgido varios conjuntos de datos públicos que, aunque aún limitados en comparación con otras modalidades de imagen médica como la mamografía o la resonancia, han permitido avanzar en el entrenamiento de modelos de inteligencia artificial aplicados a esta técnica. A continuación, se presentan las principales bases de datos disponibles en el ámbito académico, destacando sus características más relevantes.

- ***Visual Lab, Breast Thermal Image* (Brasil):** Esta base de datos contiene un total de 336 imágenes termográficas mamarias, divididas entre 186 casos benignos y 150 malignos. Fue desarrollada por la Universidad Federal Fluminense y se ha convertido en uno de los recursos más citados en la literatura especializada. Su valor añadido radica en el formato de los datos —disponibles como matrices térmicas en archivos .txt— y en la inclusión de información clínica adicional, como la edad de la paciente, la etiqueta diagnóstica o el ángulo de captura. Esta estructura facilita un tratamiento riguroso y reproducible de los datos, además de permitir la conversión personalizada a formato imagen en función del mapeo térmico elegido por el investigador. Al tratarse de una base de acceso abierto, su uso se ha extendido ampliamente en trabajos recientes que buscan explorar la aplicabilidad de modelos de aprendizaje profundo en termografía médica (Silva et al., 2020).

- ***MammoTherm*:** Esta base cuenta con 197 imágenes clasificadas (121 benignas y 76 malignas), y fue desarrollada en colaboración con diversos centros clínicos europeos.

Las imágenes presentan buena calidad visual, pero su uso está condicionado por dos factores: por un lado, la documentación clínica asociada es incompleta en parte de los registros; por otro, el acceso presenta ciertas restricciones en función del uso previsto. Aunque resulta útil para análisis exploratorios, estas limitaciones reducen su aplicabilidad directa en proyectos que requieran gran volumen de datos etiquetados o replicabilidad metodológica (Neto et al., 2019).

- ***Irthermo Benchmark Dataset Collection*:** Este recurso ofrece un total de 82 imágenes (50 benignas, 32 malignas), lo que supone una muestra muy limitada para entrenamientos en profundidad. Si bien ha sido empleado en algunos estudios preliminares sobre segmentación térmica y detección automática, su baja escala y la heterogeneidad en el formato de las imágenes dificultan su integración directa en pipelines modernos de entrenamiento de redes neuronales. Su utilidad se restringe, por tanto, a experimentos de validación cruzada o pruebas conceptuales (Fernández et al., 2020).
- **Base de datos de la Universidad de Rajasthan (India):** Se trata de un conjunto menos conocido y con una documentación técnica más escasa. No se dispone de una cifra exacta sobre el número de imágenes accesibles ni sobre la distribución de clases. Además, la información disponible sobre el protocolo de adquisición es limitada, lo que plantea dudas sobre la consistencia y calidad de las imágenes. Estas carencias metodológicas han limitado su adopción en investigaciones recientes y hacen que su uso sea poco recomendable para entrenamientos orientados a tareas clínicas exigentes.

En conjunto, puede afirmarse que, aunque el número de bases de datos públicas en termografía mamaria es reducido en comparación con otras modalidades de imagen médica, existen algunos recursos con un potencial significativo para la investigación. La disponibilidad de datos bien etiquetados, en formatos interoperables y con respaldo clínico verificable sigue siendo uno de los principales cuellos de botella en este campo, lo que justifica la necesidad de seleccionar cuidadosamente la base de datos a emplear en cualquier proyecto basado en aprendizaje profundo.

Uso de las GAN

Esta situación de escasez de bases de datos clínicas, tanto en volumen como en estandarización, limita seriamente la posibilidad de entrenar modelos de aprendizaje profundo

eficaces y generalizables. A diferencia de otros dominios de la imagen médica, donde existen miles de muestras anotadas y accesibles públicamente, la termografía mamaria sigue siendo un campo infrarrepresentado en términos de datos abiertos. Esta carencia impone una barrera técnica directa al uso de arquitecturas como las redes neuronales convolucionales (CNN), que dependen de grandes cantidades de datos para evitar el sobreajuste y lograr una clasificación robusta.

En este contexto, resulta imprescindible considerar estrategias alternativas que permitan suplir esta limitación estructural. Entre ellas, la generación sintética de datos mediante Redes Generativas Antagónicas (GANs) ha emergido en los últimos años como una solución innovadora y en plena expansión dentro del ámbito del diagnóstico médico asistido por inteligencia artificial. El siguiente apartado examina en profundidad el papel de las GANs como herramienta clave para aumentar la disponibilidad y diversidad de datos en proyectos como el presente.

6. METODOLOGÍA

En este contexto, el presente apartado despliega de forma aplicada el modelo metodológico del **Doble Diamante**, traduciéndolo en un itinerario de trabajo adaptado a la problemática específica del proyecto en Ecuador. Cada bloque de este capítulo corresponde a una fase concreta del proceso (Discover, Define, Develop y Deliver), y en conjunto constituyen una estrategia sistemática para enfrentar el desafío de mejorar el diagnóstico temprano del cáncer de mama mediante tecnologías accesibles y replicables. A lo largo de esta sección, se presentan tanto las decisiones técnicas como los fundamentos sociales y clínicos que guían el diseño del sistema, con especial énfasis en su aplicabilidad en entornos con recursos limitados.

DISCOVER: Identificación del problema y contexto de actuación

1. Problemas reales en el contexto de Ecuador

El proyecto de investigación toma como punto de partida una problemática concreta: la dificultad del acceso al diagnóstico precoz del cáncer de mama en **zonas periféricas y rurales de Ecuador**. Aunque el cáncer de mama es una de las principales causas de mortalidad femenina en todo el mundo (Siegel et al., 2022), esta situación se agrava en contextos de **desigualdad sanitaria** como los que se presentan en regiones fuera de los grandes núcleos urbanos del país andino.

En Ecuador, los sistemas de cribado mediante mamografía están **concentrados en ciudades capitales** o centros médicos altamente especializados, lo que impide que una gran parte de la población femenina tenga acceso regular a este tipo de pruebas. La **pobreza estructural**, los **malos accesos terrestres** y la **escasa presencia de unidades móviles o protocolos descentralizados de salud** obstaculizan el diagnóstico temprano y la intervención clínica oportuna. Este desfase en el acceso al diagnóstico se traduce en que **muchos casos sean detectados en etapas avanzadas**, cuando las opciones terapéuticas son más limitadas y el pronóstico mucho más reservado.

Además, los equipos de mamografía son costosos, requieren mantenimiento constante, operadores capacitados y, en la mayoría de los casos, instalaciones fijas. A esto se suma el **miedo o rechazo de muchas mujeres hacia pruebas invasivas**, especialmente en regiones con menor nivel educativo o escasa cultura preventiva, lo que se convierte en una barrera adicional para la implementación de campañas eficaces de detección.

2. Causas de los problemas reales

Las causas subyacentes a esta problemática se pueden clasificar en tres categorías principales:

- **Estructurales**: Falta de infraestructura sanitaria, escasa presencia de centros oncológicos fuera de las capitales provinciales, y carencia de redes de transporte adaptadas para unidades móviles.
- **Socioeconómicas**: Alta proporción de población femenina en situación de pobreza o vulnerabilidad, lo que limita su acceso a controles médicos periódicos, especialmente si estos requieren desplazamientos largos y costosos.
- **Tecnológicas**: Baja penetración de tecnología médica en regiones rurales y periurbanas, con infraestructuras sanitarias limitadas a funciones básicas, sin capacidad de diagnóstico avanzado. Además, las tecnologías disponibles suelen depender de técnicos especializados, lo que restringe su uso continuo o independiente.

Estas causas están ampliamente documentadas en el contexto de América Latina y el Caribe, donde la **brecha digital y tecnológica en salud** sigue siendo una de las principales limitaciones para el desarrollo de estrategias preventivas eficaces (OPS/OMS, 2021).

3. Planteamiento de la solución final: diagnóstico con instrumentos portátiles

Ante este panorama, el proyecto propone una alternativa accesible y tecnológica: el uso de instrumentos portátiles de captura termográfica combinados con modelos de inteligencia artificial, los cuales permiten realizar un primer análisis automatizado de las imágenes con el fin de apoyar el diagnóstico médico, sin sustituir en ningún caso la evaluación por parte de especialistas.

Estos dispositivos portátiles, conectados a plataformas computacionales ligeras o incluso móviles, permiten captar imágenes térmicas de la mama en cuestión de segundos. Posteriormente, estas imágenes pueden ser analizadas por algoritmos entrenados previamente, capaces de identificar patrones sospechosos compatibles con lesiones malignas o benignas. La principal ventaja de esta solución radica en su bajo coste, portabilidad, ausencia de radiación y no impasividad, lo que permite su implementación en campañas de cribado en zonas remotas con escasos recursos sanitarios.

Este sistema tiene el potencial de:

- Aumentar el acceso al diagnóstico en comunidades rurales.
- Descentralizar el acceso al diagnóstico del cáncer de mama dentro del sistema de salud.
- Disminuir la dependencia de especialistas para la interpretación inicial.
- Detectar precozmente anomalías que puedan derivar en cáncer de mama.

Desde una perspectiva técnica, esta aproximación también responde a los hallazgos del estado del arte, donde se ha evidenciado que, a diferencia de las mamografías, las imágenes termográficas pueden ser más adecuadas para el desarrollo de herramientas automáticas de diagnóstico en entornos desfavorecidos. Esto se debe a que la termografía permite una detección no invasiva de alteraciones fisiológicas como la angiogénesis tumoral o la inflamación, que generan patrones térmicos detectables en la piel (Arora et al., 2008; Acharya et al., 2012). Además, los sistemas de mamografía convencionales requieren una infraestructura técnica y humana mucho más costosa, compleja y difícil de movilizar, lo que refuerza la necesidad de alternativas accesibles en zonas con limitaciones estructurales.

Por ello, se plantea que los instrumentos portátiles de captura termográfica puedan actuar como una primera línea de cribado o "prediagnóstico" en campañas de salud pública, facilitando la detección precoz en regiones donde la mamografía no es viable. La captura de estas imágenes se realiza mediante cámaras infrarrojas capaces de registrar los gradientes de

temperatura de la superficie de la piel. Aunque en el presente punto solo se describe su aplicación general, el detalle técnico sobre el protocolo de captura utilizado, así como el proceso específico de recolección y tratamiento de los datos de la base de datos utilizada, se desarrollará más adelante en la sección correspondiente (punto 4 de este apartado).

Además, al integrarse con modelos de aprendizaje profundo (GANs y CNNs), se abre la posibilidad de mejorar continuamente el sistema con nuevos datos recopilados, entrenando los modelos de forma progresiva para adaptarse a nuevas poblaciones y contextos clínicos.

4. Descripción del dataset

La captura de imágenes termográficas es una etapa clave dentro del flujo de trabajo del presente TFG, ya que constituye la base sobre la cual se entrena y evalúa el sistema propuesto. Antes de seleccionar la base de datos principal utilizada en este proyecto, se llevó a cabo un análisis comparativo exhaustivo de varias bases de datos públicas disponibles que contenían imágenes termográficas de mama. Esta evaluación consideró tanto la calidad de las imágenes como la cantidad de muestras disponibles, la diversidad de casos representados, la presencia de metadatos clínicos asociados y el respaldo científico de cada recurso (Silva et al., 2020; Moreira et al., 2021).

Como se ha expuesto previamente en el estado del arte (apartado 5.5.6 del estado del arte), existen varias bases de datos públicas que recopilan imágenes termográficas de mama con distinto grado de calidad, tamaño y documentación clínica. Tras analizar sus características técnicas y operativas, se ha seleccionado la base de datos del Visual Lab como la opción más adecuada para este proyecto, por su equilibrio entre volumen, calidad térmica y disponibilidad de metadatos clínicos. A continuación, se justifica en detalle el motivo de esta elección:

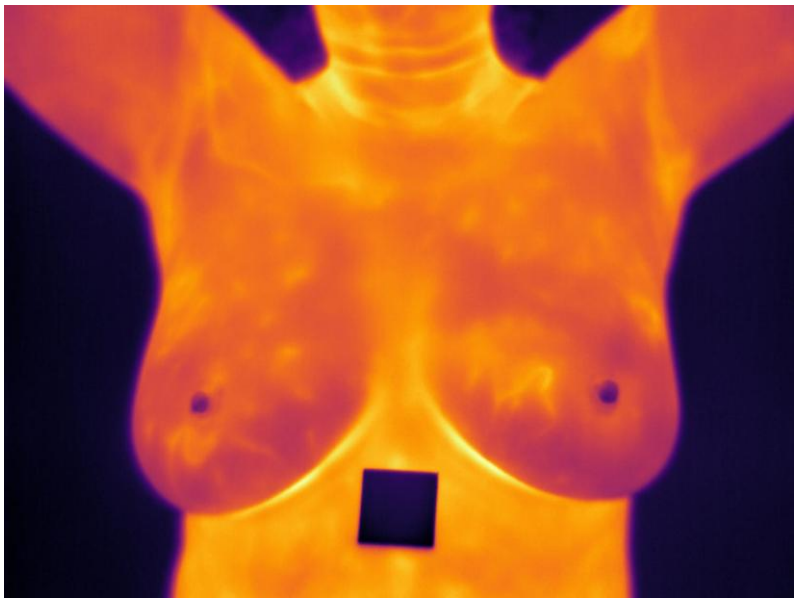
- Cuenta con una distribución equilibrada entre casos benignos y malignos, lo cual resulta fundamental para entrenar modelos de clasificación de manera efectiva.
- Incluye información clínica relevante asociada a cada imagen, como la edad de la paciente, el diagnóstico médico, la fecha y el ángulo de captura, lo que aporta contexto clínico valioso.
- Está avalada por la Universidad Federal Fluminense, una institución académica brasileña de referencia en investigación científica, y ha sido utilizada previamente en numerosos estudios del ámbito médico-tecnológico, lo que confirma su fiabilidad y pertinencia (Silva et al., 2020).

- Es de acceso público, lo que facilita tanto la replicabilidad del estudio como la transparencia en los procesos de investigación. Además, tras establecer comunicación directa con los responsables de la base de datos, proporcionaron un enlace directo a través de Google Drive, lo que permitió descargar el conjunto completo de imágenes de forma eficiente y estructurada, facilitando así su tratamiento posterior.

Inicialmente se planteó la posibilidad de combinar distintas bases de datos para ampliar el volumen de información disponible, sin embargo, esta estrategia requería un trabajo adicional de estandarización, limpieza y armonización de formatos que excedía los recursos y el alcance temporal previstos para este proyecto.

Por lo tanto, las imágenes utilizadas en este proyecto provienen de la base de datos del Visual Lab, seleccionada por cumplir los criterios mencionados y proporcionar una base sólida y fiable para el entrenamiento y validación de modelos de inteligencia artificial aplicados al diagnóstico temprano del cáncer de mama, con especial énfasis en registros obtenidos en condiciones controladas y con respaldo clínico documentado.

Ilustración 4: Imagen termográfica de paciente saludable



4.1 Proceso de captura de imágenes

Como se explicó en el apartado del estado del arte, la termografía infrarroja requiere un protocolo de captura controlado para garantizar la utilidad diagnóstica de las imágenes. En esta sección se describe brevemente dicho procedimiento y, sobre todo, se detalla cómo se ha aplicado específicamente en el caso del *dataset* seleccionado.

En el caso de la base de datos de Visual Lab, las imágenes fueron capturadas con cámaras termográficas de alta sensibilidad, bajo condiciones ambientales cuidadosamente estandarizadas. Según la información oficial del proyecto (Silva et al., 2020; Visual Lab, 2024), los exámenes se realizaron en un laboratorio térmico controlado de la Universidad Federal Fluminense (UFF), manteniendo una temperatura ambiente constante de aproximadamente 20 °C. Las pacientes permanecieron al menos 15 minutos en reposo en una sala de adaptación térmica antes de la toma de imágenes, para asegurar la estabilización térmica de la piel y minimizar la influencia de factores externos. Durante este tiempo, se evitó cualquier contacto con fuentes de calor o frío, ventilación directa, actividad física o uso de productos como cremas o desodorantes.

Las capturas se realizaron con cámaras infrarrojas FLIR de precisión médica, y las imágenes fueron tomadas mayoritariamente desde una vista frontal, complementadas en algunos casos con tomas laterales (derecha e izquierda a 45° y 90°), lo que permite una visión integral de la región mamaria. Este protocolo busca maximizar la calidad diagnóstica de las imágenes y permitir su uso confiable en aplicaciones clínicas y de investigación automatizada. La metodología y estandarización del procedimiento fueron validadas y documentadas públicamente por el equipo de Visual Lab en su página oficial del proyecto (https://visual.ic.uff.br/dmi/projeto.php), lo que refuerza la solidez de esta base de datos como fuente principal del presente estudio.

Las pacientes no debían haber utilizado desodorantes, cremas o cualquier otro producto que pudiera interferir en la captura térmica (Visual Lab, 2020). Asimismo, se registraron metadatos esenciales como el ángulo de captura, fecha, edad de la paciente, y diagnóstico clínico asociado, lo que permite una contextualización precisa de cada imagen dentro del proceso de análisis automatizado.

Este nivel de estandarización, reflejado en la documentación pública del Visual Lab, constituye uno de los motivos clave por los que esta base de datos fue seleccionada como principal fuente para el desarrollo del presente trabajo.

Este bloque inicial "Discover" refleja las primeras decisiones que orientan todo el desarrollo posterior: definir el problema, encontrar la fuente de datos adecuada y establecer un punto de partida técnico sólido sobre el que construir el sistema.

DEFINE: técnicas de *data augmentation*

En el ámbito médico, la recopilación de grandes volúmenes de datos clínicos representa un desafío considerable debido a múltiples condicionantes. Las imágenes diagnósticas, especialmente las que involucran información sensible como en el caso del cáncer de mama, están sujetas a estrictos protocolos de seguridad, normativas éticas y consideraciones de privacidad que limitan su disponibilidad pública. Además, los procesos de anonimización, validación clínica y etiquetado experto requieren una gran inversión de tiempo y recursos. Esta situación ha generado una escasez estructural de datos accesibles y bien etiquetados, lo que dificulta el uso efectivo de modelos de aprendizaje profundo, que necesitan grandes cantidades de ejemplos para evitar el sobreajuste y alcanzar niveles de precisión aceptables. Ante este escenario, el uso de técnicas de generación de datos sintéticos surge como una solución técnica y ética para suplir la falta de información real y garantizar la viabilidad de sistemas automatizados en contextos clínicos reales.

La generación sintética de datos es una herramienta clave cuando se trabaja con conjuntos de datos pequeños, desbalanceados o costosos de adquirir, como ocurre con las imágenes médicas en general y, en particular, con la termografía mamaria. En este contexto, donde cada imagen representa una valiosa fuente de información diagnóstica, pero es difícil de obtener en grandes volúmenes, es fundamental considerar estrategias que permitan aumentar la cantidad y diversidad del conjunto de datos original sin comprometer su calidad ni su representatividad.

Existen diferentes métodos para la creación de datos sintéticos que han sido propuestos en la literatura especializada, cada uno con su propio nivel de complejidad, requisitos computacionales y grado de adecuación al dominio médico. En los siguientes apartados se describen y comparan las principales aproximaciones, evaluando sus beneficios y limitaciones en función del tipo de datos termográficos con los que se trabaja en este proyecto.

1. Opciones para crear datos sintéticos

Técnicas de aumento de datos tradicionales:

El aumento de datos o *data augmentation* es una técnica comúnmente utilizada en aprendizaje profundo para aumentar la cantidad de datos disponibles. Las técnicas incluyen rotación, escalado, traslación, volteo, y adición de ruido. Estas técnicas son fáciles de implementar y tienen la ventaja de que no requieren entrenamiento adicional de modelos. Sin

embargo, no generan nuevas variaciones significativas de las imágenes, por lo que la diversidad y realismo de los datos aumentados está limitada. Aplicadas a las imágenes frontales de Visual Lab, estas transformaciones generarán pequeñas variaciones que simulan cambios de postura o iluminación. Aunque permiten triplicar o cuadruplicar las muestras por clase, no añaden información nueva desde el punto de vista clínico, ya que las estructuras térmicas y anatómicas siguen siendo las mismas. Las imágenes generadas mantendrían la calidad original, pero no contribuirían significativamente a la variabilidad del conjunto de datos. Si bien esta técnica puede mejorar la generalización del modelo frente a pequeñas perturbaciones, no resuelve la escasez de datos originales.

Modelos basados en interpolación:

Métodos como SMOTE y sus variantes adaptadas para imágenes operan combinando muestras cercanas en el espacio de características para crear nuevos ejemplos sintéticos. Sin embargo, la interpolación no actúa sobre los píxeles de las imágenes ni sobre las matrices térmicas, sino sobre representaciones derivadas, lo que dificulta la conservación de los patrones térmicos reales. Si se aplica a las imágenes de Visual Lab, la cantidad de datos podría aumentarse en un 100 %, pero con el riesgo de introducir ruido o inconsistencias térmicas que afecten al rendimiento del modelo. Además, la calidad de las imágenes generadas podría verse comprometida, por lo que no se considera una opción adecuada para este tipo de imágenes médicas.

Redes Generativas Antagónicas (GANs):

Las GANs permiten generar datos completamente nuevos que pueden reflejar una mayor diversidad y realismo. Los modelos GAN pueden producir imágenes completamente nuevas que se asemejan a las reales, sin necesidad de reestructurar los datos existentes. Esto puede ser especialmente útil cuando se dispone de un número limitado de imágenes o cuando se busca aumentar la variabilidad en el *dataset* sin generar un sobreajuste. En este proyecto, se utilizarán imágenes frontales de Visual Lab para entrenar el generador de una GAN. Este modelo podría producir entre 1.000 y 3.000 imágenes por clase, dependiendo del control de calidad. A nivel de variedad, las GANs permiten introducir pequeñas variaciones naturales en las estructuras térmicas, lo que incrementa la diversidad del conjunto de datos sin repetir patrones. La calidad de las imágenes generadas, si se entrena adecuadamente el modelo, podría ser indistinguible de las originales. Este enfoque permitirá equilibrar las clases de la base de

datos, mejorar la capacidad de generalización del modelo y reducir el riesgo de sobreajuste. Por qué elegir las GANs

2. Valoración de las GANs

Tras analizar las distintas técnicas existentes, se opta por el uso de redes generativas antagónicas (GANs) como principal estrategia de generación sintética para este proyecto. En este apartado se detallan sus principales fortalezas y limitaciones, considerando su impacto en la generación de imágenes útiles para modelos de clasificación automática como las CNNs.

Las GANs ofrecen varias ventajas sobre otros métodos de aumento de datos:

Pros:

- **Generación de datos realistas:** Las GANs pueden generar datos completamente nuevos, lo que permite una gran variedad de imágenes. Estas imágenes pueden ser indistinguibles de las reales, lo que las hace útiles para entrenar redes neuronales más robustas.
- **Diversidad de datos:** A diferencia de los métodos de aumento de datos tradicionales, las GANs pueden generar variaciones complejas que no están limitadas por las transformaciones simples de las imágenes originales.
- **Mejora en la clasificación:** Se ha demostrado que el uso de datos generados por GANs mejora significativamente la precisión de los clasificadores de imágenes, como las CNNs.
- **Reducción de sesgo:** Las GANs pueden ayudar a equilibrar los *datasets*, lo que es especialmente útil cuando se enfrentan a desequilibrios en las clases (por ejemplo, más imágenes de tejido sano que de tejido canceroso).

Contras:

- **Mayor complejidad computacional:** El entrenamiento de modelos GAN puede ser computacionalmente costoso y lleva más tiempo en comparación con técnicas de aumento de datos simples.
- **Dificultad en la evaluación:** Evaluar la calidad de las imágenes generadas por GANs puede ser complicado, aunque existen métricas como FID (Fréchet Inception Distance) y KID (Kernel Inception Distance) que permiten medir la calidad de las imágenes generadas.

- **Riesgo de sobreajuste:** Aunque las GANs generan imágenes sintéticas diversas, existe el riesgo de que los modelos generen imágenes demasiado similares entre sí si no se entrenan correctamente.

3. Tipos de GANs a valorar

Una vez tomada la decisión de utilizar GANs como método principal para la generación sintética de datos, es necesario abordar un aspecto crucial: la elección de la arquitectura más adecuada. Esta decisión no es trivial, ya que existen múltiples variantes de GANs que difieren en sus objetivos, mecanismos de entrenamiento, estabilidad y calidad de resultados. Algunas están orientadas a mejorar la estabilidad del entrenamiento, otras a aumentar la fidelidad visual de las imágenes generadas o a introducir condiciones explícitas sobre la clase o el contenido.

Por ello, resulta fundamental comparar las principales arquitecturas disponibles en la literatura y evaluar cuáles se adaptan mejor a los requisitos del proyecto, teniendo en cuenta tanto las características del *dataset* como las necesidades del clasificador que se utilizará posteriormente.

Las GANs han evolucionado en distintas arquitecturas, cada una con sus ventajas y desventajas. Algunas de las más utilizadas incluyen:

1. **GANs estándar (Goodfellow et al., 2014):** Son la arquitectura original de GAN. Aunque son simples y eficaces en muchos casos, tienen problemas de inestabilidad durante el entrenamiento.

2. **CycleGAN (Zhu et al., 2017):** Utiliza un ciclo de entrenamiento para garantizar que las imágenes generadas sean de alta calidad y mantengan características similares a las imágenes originales. Es útil en casos de *image-to-image translation* donde se necesita transformar imágenes de un dominio a otro.

3. **WGAN (Arjovsky et al., 2017):** Ofrece una mejora en la estabilidad del entrenamiento mediante el uso de una función de pérdida más suave y un mejor control sobre el entrenamiento del discriminador.

4. **Conditional GAN (Mirza & Osindero, 2014):** Permite generar imágenes condicionadas a una clase o etiqueta, lo que resulta útil si se desea generar imágenes de un tipo específico.

4. Selección del modelo GAN para este proyecto

Tras revisar las distintas arquitecturas disponibles, se ha optado por implementar una GAN personalizada, diseñada específicamente para el tratamiento de imágenes térmicas en escala de color. Esta arquitectura simplificada se compone exclusivamente de capas convolucionales y no incorpora mecanismos de condicionamiento, ciclos de consistencia ni pérdidas sofisticadas. Su principal ventaja reside en su adaptabilidad a conjuntos de datos pequeños y especializados, como el de Visual Lab, donde es crucial mantener la coherencia térmica y anatómica sin introducir distorsiones.

El generador está formado por un bloque convolucional con activaciones ReLU y salida Tanh, mientras que el discriminador emplea una estructura sencilla con capas LeakyReLU y una salida Sigmoid. Esta configuración permite un entrenamiento estable y controlado, con bajo riesgo de colapso de modo o sobreajuste. Además, la simplicidad del modelo facilita su entrenamiento en hardware limitado y la evaluación cualitativa de los resultados.

La elección de esta GAN responde, por tanto, a una necesidad de equilibrio entre viabilidad técnica, fidelidad térmica y estabilidad de entrenamiento, priorizando la generación de imágenes sintéticas que sean útiles para la tarea de clasificación sin introducir ruido artificial o patrones irreales. Dado el contexto clínico del proyecto, esta solución se ajusta mejor a los requisitos que otras arquitecturas más complejas, cuya sofisticación podría no traducirse en mejoras reales sobre un conjunto de datos tan específico.

5. Resultados esperados

Se espera que el uso de GANs para aumentar el *dataset* de imágenes termográficas permita mejorar la precisión del modelo CNN para la detección de cáncer de mama, proporcionando datos más diversos y representativos que ayuden a reducir los errores de clasificación. Los resultados esperados incluyen una mayor robustez en el modelo, una disminución del sobreajuste y una mejora general en el rendimiento del clasificador.

DEVELOP: Preprocesamiento de imágenes

En el contexto de este proyecto, se ha decidido emplear exclusivamente imágenes frontales del *dataset* de Visual Lab. Esta elección se basa en una combinación de criterios técnicos, clínicos y operativos que, desde un punto de vista teórico, resultan óptimos para la etapa de generación y clasificación automática.

Desde una perspectiva técnica, el uso de una única perspectiva reduce significativamente la complejidad del modelo requerido. Entrenar una red que sea capaz de distinguir entre diferentes ángulos (frontal, lateral izquierdo, lateral derecho, etc.) implicaría no solo una mayor carga computacional, sino también la necesidad de una arquitectura especializada, posiblemente con módulos de atención o redes paralelas. Esto excede los recursos y el alcance establecidos para el presente TFG (Fernández Ovies, 2018).

Desde un enfoque clínico, las imágenes frontales contienen suficiente información térmica sobre la región mamaria, permitiendo identificar patrones de simetría, gradientes de temperatura y zonas de hipervascularización asociadas con procesos tumorales. Varios estudios han demostrado que las vistas frontales, aunque más simples, pueden proporcionar un rendimiento diagnóstico comparable a la combinación multivista si se emplean modelos bien entrenados (Sadek Ali et al., 2016; Acharya et al., 2012).

Además, el uso exclusivo de imágenes frontales mejora la homogeneidad del *dataset*. Al reducir la variabilidad estructural entre muestras, se facilita el proceso de aprendizaje del modelo y se mejora la capacidad de generalización. La uniformidad también reduce la posibilidad de que el modelo aprenda a asociar características de perspectiva con clases diagnósticas, un tipo de sesgo no deseado.

Por lo que, desde un punto de vista operativo, seleccionar únicamente imágenes frontales permite simplificar el pipeline de procesamiento, disminuir los tiempos de ejecución y enfocar los esfuerzos en mejorar la calidad del generador GAN y del clasificador CNN.

Además, se ha de tener en cuenta que, en el contexto del entrenamiento de modelos generativos, como las redes GAN, el preprocesamiento de datos se convierte en **una etapa fundamental para garantizar la calidad del aprendizaje**. A diferencia de los modelos clasificadores, donde el objetivo es extraer patrones discriminativos, las GAN requieren una representación lo más coherente y estructurada posible del dominio a modelar. En el caso particular de las imágenes termográficas, esta necesidad se intensifica debido a su naturaleza numérica, su **sensibilidad térmica** y la relevancia clínica de las variaciones locales. Si las imágenes presentan inconsistencias en la resolución, escala térmica o representación visual, el generador puede aprender distribuciones erróneas o generar imágenes poco realistas.

Por ello, esta fase no solo busca transformar datos crudos y heterogéneos en un formato uniforme, sino también asegurar que dicha representación conserve las características térmicas

relevantes para la tarea de generación sintética. Este enfoque teórico establece las bases necesarias para un entrenamiento estable, reproducible y clínicamente útil del modelo GAN propuesto. A continuación, se detallan los pasos que guiarán el tratamiento de los datos de entrada

1. Conversión de matriz térmica a imagen

Aunque se dispone de las imágenes termográficas ya convertidas a formato visual (JPEG), se opta por trabajar directamente con los arrays de datos térmicos originales en formato matricial. Esta elección se fundamenta en la posibilidad de aplicar un mapeo térmico personalizado que preserve con mayor fidelidad la escala cromática y la riqueza informativa de las temperaturas registradas, algo que se pierde en las versiones ya procesadas de las imágenes.

En este sentido, el pipeline teórico incluye la lectura directa de los archivos .txt proporcionados por Visual Lab, los cuales contienen matrices de 640x480 puntos con valores de temperatura. Estas matrices son convertidas en imágenes mediante un mapeo lineal de colores, eligiendo una paleta térmica adecuada para resaltar contrastes clínicamente relevantes. Este procedimiento busca normalizar las temperaturas para llevarlas a un rango visual definido y comparable entre imágenes, asegurando que las diferencias térmicas observadas sean representativas y útiles para el aprendizaje del modelo.

La selección del mapeo de color no solo tiene un valor estético o visual, sino que constituye la base con la que la GAN aprenderá la distribución de patrones térmicos presentes en las pacientes. Por tanto, resulta esencial controlar este paso de forma explícita y reproducible, ajustando también el rango de temperatura visible para evitar valores extremos o atípicos que puedan sesgar la generación.

Finalmente, este paso se plantea como parte del preprocesamiento exclusivo para el entrenamiento del generador de la GAN, y no para la clasificación con CNN, permitiendo a la red aprender directamente desde una representación térmica coherente y estructurada, tal como recomiendan trabajos anteriores en entornos similares (Kolosnjaji et al., 2020; Litjens et al., 2017).

2. Relevancia diagnóstica de anomalías térmicas

En termografía médica, la presencia de asimetrías, picos térmicos localizados o regiones con valores fuera del rango esperado suele estar correlacionada con alteraciones

fisiológicas, incluyendo procesos tumorales (Arora et al., 2008; Head et al., 2000). Por tanto, es esencial que el proceso de preprocesado preserve estas características.

Diversos estudios han demostrado que los tumores pueden provocar un incremento localizado de la temperatura en la superficie cutánea debido al aumento del metabolismo celular y la angiogénesis asociada (Ng, 2009). Por ello, se buscará mantener estas irregularidades térmicas en las imágenes procesadas, evitando técnicas de suavizado excesivo que puedan diluir la información diagnóstica.

Además, se incluirán mecanismos de verificación para asegurar que las zonas térmicamente anómalas no sean eliminadas o modificadas durante la normalización. El objetivo es que la red neuronal pueda detectar estos patrones como indicadores potenciales de malignidad.

3. Normalización en función de la temperatura media

Cada imagen térmica presenta una distribución particular de temperaturas que depende de factores como el entorno, el dispositivo y el momento de adquisición. Para reducir esta variabilidad y permitir comparaciones más fiables entre imágenes, se ajustará cada matriz térmica a una temperatura media estándar global, calculada sobre el conjunto total de datos.

Este proceso, denominado "mean shift normalization", permite reducir la dispersión estadística de las muestras y centra la atención del modelo en las diferencias locales (es decir, anomalías térmicas) más que en las condiciones generales de captura (Mital et al., 2018). La idea es desplazar el valor medio de cada imagen hasta igualarlo con la media global, manteniendo intacta la varianza relativa interna de cada muestra.

Este tipo de normalización también ha demostrado ser útil en estudios previos de análisis de imágenes térmicas médicas (Kolosnjaji et al., 2020), especialmente cuando se trabaja con conjuntos de datos heterogéneos o multifuente.

4. Selección y valoración de arquitecturas CNN

Una vez establecido el conjunto de datos y completado el preprocesamiento, el siguiente paso del proyecto se centra en seleccionar una arquitectura convolucional adecuada para afrontar la tarea de clasificación de imágenes térmicas mamarias. En este tipo de aplicación clínica, la elección del modelo no puede basarse únicamente en su rendimiento general en

benchmarks tradicionales, sino que debe ajustarse a las particularidades del dominio médico, al volumen limitado de datos y a las características propias de las imágenes térmicas.

En este contexto, se ha llevado a cabo una revisión crítica de las principales arquitecturas CNN utilizadas en imagen médica. Algunos modelos clásicos, como **AlexNet** (Krizhevsky et al., 2012) o **VGG** (Simonyan & Zisserman, 2015), aunque fundamentales en el desarrollo del aprendizaje profundo, no resultan óptimos para el tipo de datos empleados en este trabajo. Su gran número de parámetros y estructura poco eficiente pueden llevar fácilmente al sobreajuste cuando se trabaja con conjuntos de imágenes reducidos o con variabilidad limitada, como es habitual en termografía médica (Shen et al., 2019).

Otras arquitecturas más complejas como **Inception** (Szegedy et al., 2015), diseñadas para capturar información multiescala mediante convoluciones en paralelo, presentan un diseño computacionalmente más costoso. Aunque han demostrado buenos resultados en modalidades de imagen estructurada como la resonancia magnética, su aplicación en imágenes térmicas ha sido escasa, probablemente debido a que estas imágenes carecen de la riqueza de formas o estructuras anatómicas que justificarían dicha complejidad. Por su parte, **DenseNet** (Huang et al., 2017), con su estrategia de conexiones densas entre capas, ha ofrecido resultados prometedores en tareas clínicas más estructuradas —como la detección de patrones histológicos—, pero en termografía puede amplificar artefactos térmicos o señales poco relevantes desde el punto de vista clínico (Gao et al., 2021).

Frente a estas limitaciones, se ha optado por centrarse en dos arquitecturas que combinan eficiencia, estabilidad y capacidad de generalización: **ResNet50** y **EfficientNetB0**. **ResNet50** ha demostrado una gran solidez en escenarios biomédicos con datos ruidosos o poco estructurados. Su diseño basado en bloques residuales permite entrenar redes más profundas sin que el rendimiento se vea comprometido por problemas de desvanecimiento del gradiente (He et al., 2016). Esto es especialmente relevante cuando el objetivo es detectar patrones térmicos sutiles en imágenes superficiales, donde no siempre existe una morfología evidente. Modelos basados en ResNet han sido aplicados con éxito en múltiples áreas médicas, desde la detección de nódulos pulmonares en radiografía (Lakhani & Sundaram, 2017) hasta la clasificación de lesiones cutáneas o mamografías (Brinker et al., 2019; Zhang et al., 2021). Su capacidad para aprender representaciones complejas sin requerir grandes volúmenes de datos la convierte en una opción adecuada para este proyecto.

Junto a ResNet, se ha decidido evaluar también **EfficientNetB0**, una arquitectura más reciente que propone una estrategia de escalado optimizado, ajustando simultáneamente la profundidad, anchura y resolución de la red (Tan & Le, 2019). Este diseño ha permitido a EfficientNet alcanzar altos niveles de precisión con un número muy reducido de parámetros, lo que la hace especialmente atractiva para entornos con recursos computacionales limitados o donde se busca eficiencia sin comprometer la capacidad del modelo. En imagen médica ha sido utilizada con éxito en tareas como la detección de retinopatía diabética, clasificación de lesiones dermatológicas o análisis de mamografías (Bodapati et al., 2020; Sitaula et al., 2022; Al-antari et al., 2020), y su rendimiento ha sido comparable al de arquitecturas más grandes, incluso cuando se entrena sobre conjuntos de datos pequeños.

Ambas redes han sido entrenadas desde cero, sin aplicar transferencia de aprendizaje, dado que la distribución de características térmicas difiere sustancialmente del tipo de imágenes que alimentan los modelos preentrenados tradicionales como los basados en ImageNet. Esta decisión permite adaptar el aprendizaje desde el inicio al tipo de señales térmicas relevantes para la clasificación de tejido sano o patológico, evitando el arrastre de sesgos ajenos al dominio clínico.

En definitiva, la elección de **ResNet50** y **EfficientNetB0** responde a la necesidad de combinar profundidad de representación y eficiencia computacional en un entorno clínico donde la cantidad de datos es limitada y las diferencias entre clases pueden ser sutiles. Ambas arquitecturas ofrecen enfoques complementarios: una, más robusta y probada en tareas clínicas exigentes; la otra, más ligera y eficiente, ideal para proyectos que requieren escalabilidad y portabilidad. Esta doble estrategia permitirá evaluar con mayor precisión el impacto del aumento de datos sintéticos sobre la capacidad de generalización de los modelos CNN en escenarios reales de diagnóstico temprano.

DELIVER: Experimentación, ejecución y evaluación

Una vez definidas las arquitecturas seleccionadas y completado el preprocesamiento de datos —tanto reales como generados—, se llevó a cabo la fase experimental. En este punto, el objetivo fue doble: por un lado, comprobar la validez del conjunto de imágenes sintéticas producidas por la GAN, y por otro, evaluar el rendimiento clasificatorio de las CNN entrenadas en diferentes escenarios. Para ello se diseñó un conjunto de experimentos cuidadosamente estructurado, tanto desde el punto de vista técnico como metodológico.

1. Diseño experimental: configuración de los modelos

El diseño experimental se basó en la ejecución de **ocho experimentos principales**, divididos en dos bloques: uno con datos originales y otro con datos aumentados mediante la GAN. En cada bloque, se entrenaron cuatro modelos CNN distintos, combinando dos arquitecturas (EfficientNetB0 y ResNet50) con dos configuraciones de entrenamiento diferentes, denominadas **"estricta"** y **"permisiva"**. Esta clasificación responde a ajustes deliberados de hiperparámetros clave en el entrenamiento: la tasa de *Dropout* y la velocidad de aprendizaje.

- En los modelos **estrictos**, se empleó un *Dropout* del 50% y una tasa de aprendizaje reducida (0.0005), con el objetivo de aplicar una fuerte regularización. Este enfoque busca evitar el sobreajuste, algo particularmente crítico cuando se trabaja con un conjunto de datos reducido y con alta homogeneidad interna, como ocurre en imágenes térmicas médicas.
- Por el contrario, los modelos **permisivos** se entrenaron con una configuración más laxa: *Dropout* del 20% y una tasa de aprendizaje mayor (0.0015), permitiendo al modelo adaptarse más libremente a los patrones presentes en los datos, aunque a costa de un mayor riesgo de sobreajuste si no se controla correctamente el proceso.

Este planteamiento de doble enfoque se inspiró en estrategias previamente descritas en la literatura, donde se ha demostrado que una correcta combinación de arquitecturas con distintos grados de regularización puede facilitar el ajuste fino de modelos en escenarios clínicos complejos (Srivastava et al., 2014; Shorten & Khoshgoftaar, 2019). Además, permite estudiar cómo se comportan redes más conservadoras frente a otras más flexibles cuando se introducen datos sintéticos, y evaluar si estos últimos realmente ayudan a mejorar la generalización o simplemente amplifican el ruido.

Para cada uno de estos ocho experimentos, se repitió el entrenamiento varias veces —al menos diez ejecuciones por configuración—, con el objetivo de depurar errores, estabilizar resultados y ajustar el número de épocas de forma progresiva. Esta repetición también permitió observar la consistencia del modelo en múltiples sesiones, algo clave en tareas biomédicas, donde los márgenes de error aceptables son muy reducidos.

2. Entorno de ejecución y limitaciones técnicas

Todos los experimentos fueron ejecutados sobre el supercomputador **LORCA**, perteneciente a la Universidad Europea de Madrid. Este entorno de alto rendimiento se caracteriza por ofrecer acceso a una arquitectura potente, aunque restringida a uso de CPU. En concreto, las sesiones de entrenamiento se ejecutaron en un nodo con las siguientes características:

- **64 núcleos lógicos**
- **125,6 GB de memoria RAM**

El acceso al entorno se realizó mediante **túnel SSH**, utilizando la herramienta Bitvise y un puerto redirigido a localhost, lo que permitía trabajar con **Jupyter Notebook** desde el navegador del ordenador personal. Aunque funcional, este sistema de conexión introdujo varias limitaciones técnicas relevantes. Una de las más problemáticas fue la gestión de la **memoria caché durante los entrenamientos**. A pesar de que el cálculo se realizaba en el nodo remoto, parte del procesamiento asociado a Jupyter parecía almacenarse de forma local en el equipo personal. Al contar este con muy poca RAM, la saturación de la memoria producía bloqueos, cuelgues o interrupciones silenciosas durante el entrenamiento, especialmente en la fase de generación de imágenes sintéticas o en ciclos prolongados de entrenamiento con CNN.

Este problema condicionó directamente el diseño del modelo: se optó por **redes compactas**, con un número reducido de parámetros y con un número de épocas acotado, optimizando la eficiencia sin comprometer el aprendizaje. También se aplicaron estrategias como el uso de *early stopping*, reducción de tamaño de lote y simplificación de la arquitectura de salida para asegurar la estabilidad del proceso. En general, cada experimento completo —una única ejecución de entrenamiento CNN— tardaba entre **6 y 8 horas**, lo que justifica el número limitado de configuraciones exploradas y la necesidad de una planificación cuidadosa en cada ejecución.

3. Evaluación de resultados: métricas y criterios

La evaluación de los modelos se abordó desde dos perspectivas complementarias: una orientada a la calidad de las imágenes generadas por la GAN, y otra centrada en el rendimiento clasificatorio de las CNN.

En el caso de la **GAN**, se propuso el uso de métricas como el **Fréchet Inception Distance (FID)** y el **Kernel Inception Distance (KID)** para medir la similitud estadística entre las imágenes sintéticas y reales. Estas métricas son ampliamente utilizadas en la literatura de generación de imágenes, especialmente en contextos médicos donde no siempre se dispone de etiquetas para evaluar directamente la calidad (Heusel et al., 2017; Bi et al., 2023). Sin embargo, por limitaciones técnicas y de tiempo, no se pudo realizar una evaluación cuantitativa con estos indicadores. A pesar de ello, se llevó a cabo una inspección visual sistemática sobre las imágenes generadas, comprobando que presentaban una distribución térmica coherente, sin artefactos evidentes ni patrones repetitivos, lo que sugiere una capacidad generativa adecuada al menos desde una perspectiva cualitativa.

En cuanto a la evaluación de las **redes CNN**, se emplearon múltiples métricas para obtener una visión integral del rendimiento clasificatorio:

- **Accuracy**, como medida general del porcentaje de aciertos sobre el conjunto de test.
- **Precision**, **recall** y **F1-score**, desglosados por clase (tejido sano y tejido patológico), obtenidos a través del informe de clasificación de scikit-learn.
- **Área bajo la curva ROC (AUC)**, empleada además como métrica de control para aplicar el early stopping durante el entrenamiento.
- **Matriz de confusión**, visualizada mediante mapas de calor para observar la distribución de errores y aciertos por clase.

Estas métricas permitieron analizar no solo qué modelo obtenía mejor rendimiento global, sino también cómo se comportaba cada configuración en situaciones críticas, como la detección de falsos negativos, que en un contexto clínico podrían suponer un error de diagnóstico grave.

La comparación entre modelos entrenados solo con datos reales y aquellos alimentados con imágenes generadas por la GAN resulta especialmente interesante. Esta doble estrategia permite estudiar de forma empírica si el uso de imágenes sintéticas mejora realmente la capacidad de generalización del sistema, o si introduce ambigüedad y ruido que afectan negativamente al aprendizaje. El diseño experimental, junto con la evaluación sistemática, sienta las bases para discutir en profundidad estos resultados en el siguiente apartado.

Desarrollo del proyecto

A lo largo de este capítulo se ha detallado el proceso metodológico seguido para abordar el reto del diagnóstico temprano del cáncer de mama en entornos clínicos con recursos limitados, empleando herramientas de visión por computador y aprendizaje profundo. Partiendo de una problemática real, identificada en regiones rurales de Ecuador, se ha propuesto una solución técnica basada en dispositivos termográficos portátiles y modelos de inteligencia artificial diseñados para operar con bajo coste y alta escalabilidad.

El desarrollo metodológico se ha articulado siguiendo el esquema del Doble Diamante, lo que ha permitido organizar el trabajo en cuatro grandes fases. En la etapa de **Discover**, se analizó el contexto clínico y tecnológico que justifica la necesidad de soluciones alternativas a la mamografía, describiendo también el proceso de selección de la base de datos utilizada, con énfasis en su fiabilidad y adecuación diagnóstica. En la fase de **Define**, se abordaron los problemas derivados de la escasez de datos médicos accesibles, proponiendo como solución el uso de redes generativas (GANs) para aumentar de forma ética y controlada el volumen y la diversidad del conjunto de imágenes térmicas.

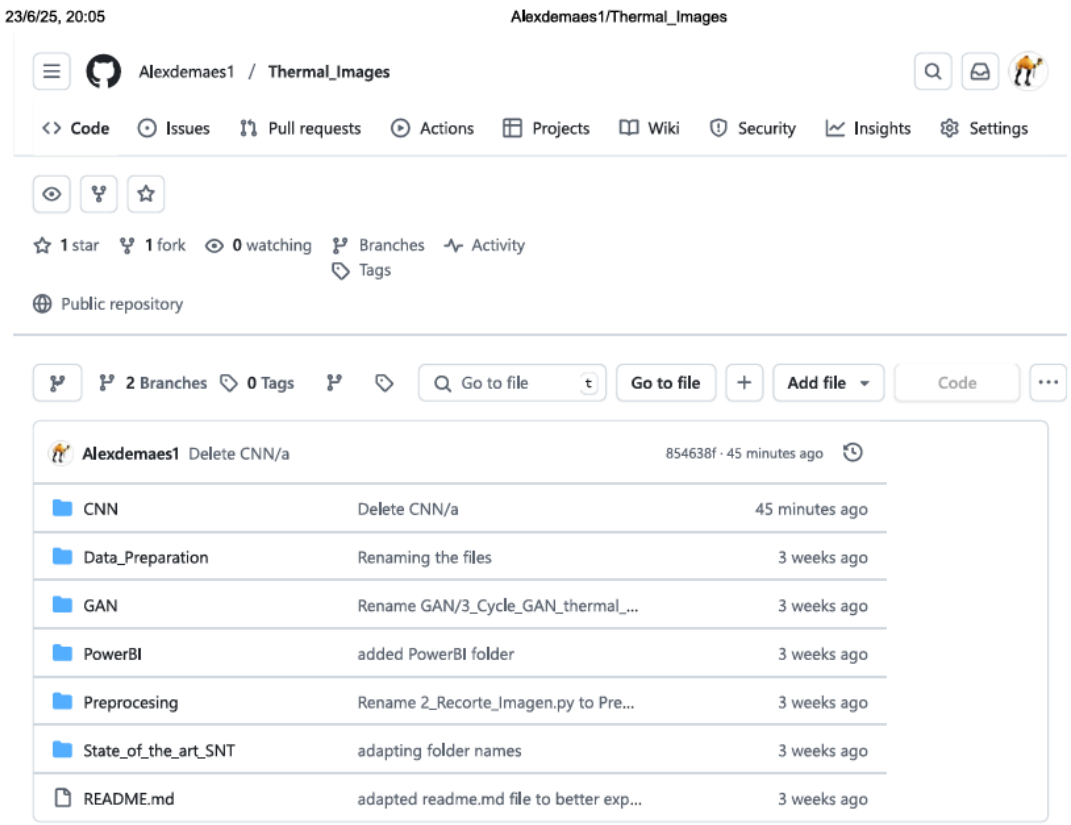
La fase de **Develop** profundizó en el tratamiento y preprocesamiento de las imágenes térmicas, destacando la importancia de preservar la información clínica en cada transformación aplicada. Se justificó el uso exclusivo de imágenes frontales por motivos técnicos y diagnósticos, y se detalló la elección de las arquitecturas convolucionales más adecuadas para la tarea de clasificación. En particular, se seleccionaron EfficientNetB0 y ResNet50 por su equilibrio entre eficiencia computacional y capacidad de generalización en escenarios médicos. Asimismo, se explicó la necesidad de trabajar con representaciones térmicas normalizadas para maximizar la calidad del entrenamiento, especialmente en el caso de las GANs.

Finalmente, en la etapa de **Deliver**, se expuso el diseño experimental y las condiciones técnicas de ejecución de los modelos. Se entrenaron un total de ocho configuraciones de redes CNN, divididas entre modelos entrenados con datos reales y con datos aumentados mediante GANs, y subdivididas según un enfoque de entrenamiento “estricto” o “permisivo”, diferenciados por los hiperparámetros de regularización. La implementación de los experimentos en el supercomputador LORCA supuso ciertas limitaciones operativas, como la imposibilidad de usar GPU o problemas de gestión de memoria, que condicionaron tanto la arquitectura de los modelos como la duración de los entrenamientos. A pesar de ello, se pudieron ejecutar y repetir los experimentos con suficiente estabilidad como para obtener

resultados significativos. La evaluación de las GANs se planteó mediante métricas estándar como FID y KID, aunque no se aplicaron de forma efectiva por cuestiones técnicas. Por su parte, el rendimiento de las CNN se evaluó exhaustivamente con métricas como accuracy, precisión, recall, F1-score, AUC y matriz de confusión, con especial atención a los errores críticos desde el punto de vista clínico.

Todo el código desarrollado en el marco de este Trabajo de Fin de Grado ha sido documentado, organizado y publicado en un repositorio accesibles a través de GitHub. En el se incluye tanto el código de entrenamiento de los modelos generativos (GANs) y clasificadores (CNNs), como los scripts de preprocesamiento, visualización de resultados y gestión del conjunto de datos. Para garantizar la trazabilidad y la posibilidad de replicación del estudio, se han fusionado las versiones correspondientes al proyecto de investigación general y al presente TFG, por lo que toda la documentación de código relativa a este proyecto se puede consultar en el siguiente enlace: https://github.com/Alexdemaes1/Thermal_Images

Ilustración 5:captura de pantalla del repositorio de GitHub



7. RESULTADOS Y DISCUSIÓN

Arquitecturas CNN empleadas

Para todos los experimentos realizados, tanto los entrenados exclusivamente con imágenes reales como los que incorporan datos sintéticos generados por la GAN, se utilizaron dos arquitecturas convolucionales ampliamente validadas en imagen médica: **ResNet50** y **EfficientNetB0**. En ambos casos, los modelos se entrenaron desde cero, sin hacer uso de pesos preentrenados, dado que las imágenes térmicas tienen características muy distintas a los conjuntos de datos tradicionales como ImageNet.

Se definieron dos configuraciones de entrenamiento: la denominada **estricta**, con un *Dropout* del 50 % y una tasa de aprendizaje de 0.0005, diseñada para imponer una regularización fuerte en entornos con pocos datos; y la **permisiva**, con un *Dropout* reducido al 20 % y un learning rate mayor, de 0.0015, pensada para permitir un aprendizaje más agresivo en presencia de mayor variabilidad. Ambas variantes comparten el uso del optimizador Adam, funciones de activación estándar y mecanismos como early stopping para evitar el sobreajuste.

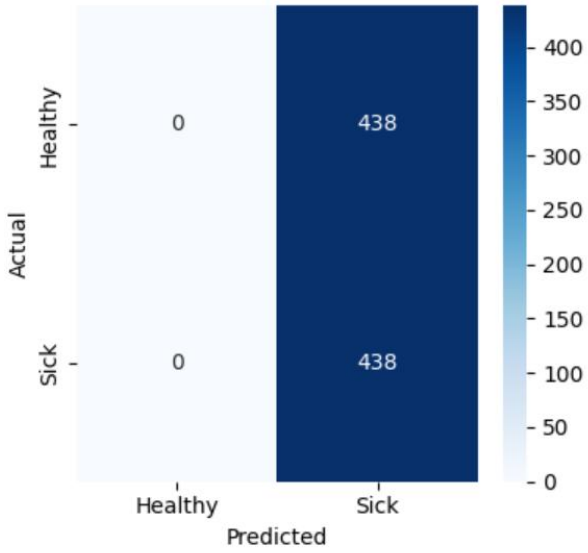
Cabe subrayar que las arquitecturas son exactamente las mismas tanto en los modelos entrenados con datos reales como en los que incluyen datos generados por GAN. La única diferencia reside en el conjunto de entrenamiento utilizado, y no en la estructura del modelo.

Resultados de los modelos CNN entrenados con datos reales

Los primeros cuatro experimentos se centraron en entrenar las CNN utilizando únicamente imágenes reales procedentes del conjunto de datos original. Los resultados obtenidos fueron, en general, limitados.

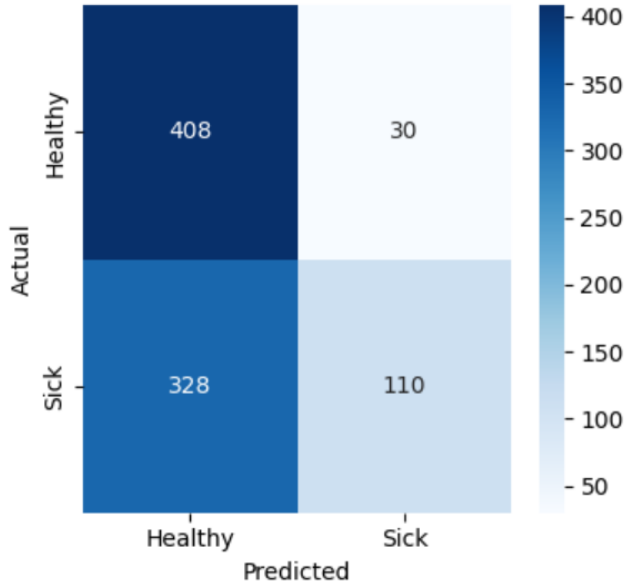
En el caso de **EfficientNetB0 con configuración permisiva**, el modelo alcanzó una exactitud global del 50 %. El análisis detallado muestra una ausencia total de capacidad para identificar correctamente los casos sanos, mientras que todos los casos enfermos fueron clasificados correctamente, lo cual indica un claro sesgo. El F1-score para la clase enferma fue de 0.67, pero en la clase sana fue 0.00. Este comportamiento sugiere que el modelo ha colapsado en una única clase y no ha aprendido representaciones útiles. A continuación, se muestra la matriz de confusión correspondiente al modelo, que permite visualizar gráficamente el comportamiento del clasificador respecto a ambas clases diagnósticas.

Ilustración 6:Matriz de confusión experimento 1.1



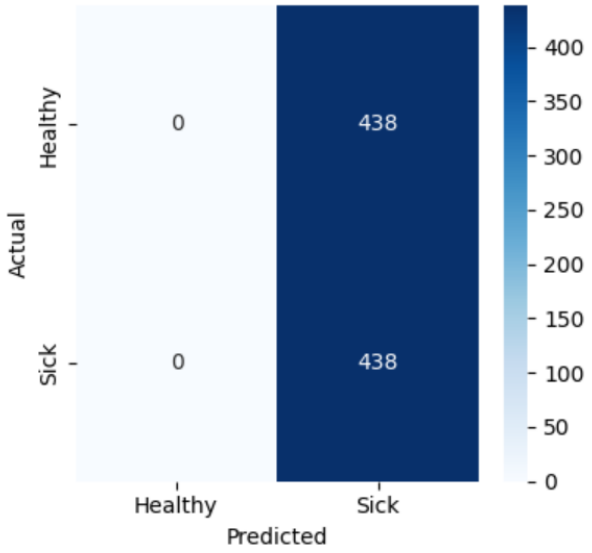
La configuración **estricta de EfficientNetB0** ofreció un rendimiento algo más equilibrado. La exactitud alcanzó el 59 %, con una precisión del 55 % en la clase sana y del 79 % en la clase enferma. No obstante, el recall en la clase enferma fue muy bajo (25 %), lo que indica una tendencia a sobrerrepresentar los casos sanos. El F1-score general se situó en torno al 0.54. La matriz de confusión generada es la siguiente:

Ilustración 7:Matriz de confusión experimento 1.2



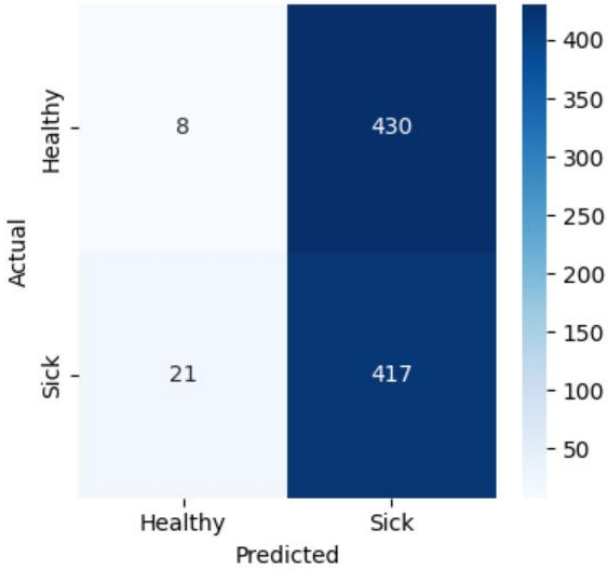
Por su parte, la arquitectura **ResNet50 en configuración estricta** replicó prácticamente el comportamiento erróneo de EfficientNet permisiva: clasificó todos los casos como enfermos. La exactitud fue del 50 %, con un recall perfecto en la clase enferma pero nulo en la sana. Las métricas reflejan una clasificación sesgada sin capacidad de generalización. Su matriz de confusión queda tal que así:

Ilustración 8:Matriz de confusión experimento 1.3



Finalmente, la **configuración permisiva de ResNet50** mostró una precisión global del 48.5 %. Aunque el modelo logró detectar un gran número de casos enfermos (recall del 95 %), el rendimiento para la clase sana fue prácticamente inexistente (recall del 2 %). El desequilibrio entre clases fue evidente, y la precisión de clasificación fue baja en términos absolutos. Los resultados vistos en una matriz de confusión son los siguientes:

Ilustración 9:Matriz de confusión experimento 1.4



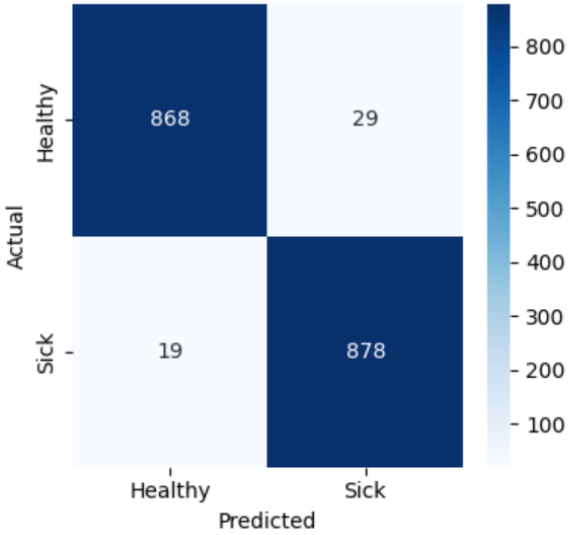
En conjunto, los cuatro modelos entrenados exclusivamente con imágenes reales presentan resultados pobres. Ninguno de ellos logra un rendimiento robusto y equilibrado. Si bien **EfficientNetB0 en configuración estricta** podría considerarse el más sólido dentro de este grupo, su rendimiento sigue siendo limitado. Esto sugiere que las arquitecturas utilizadas, aunque exitosas en imagen médica tradicional, no se adaptan directamente a la modalidad térmica sin un ajuste muy específico. Es posible que estas redes, originalmente diseñadas para detectar estructuras bien definidas como las de una mamografía, no estén optimizadas para patrones térmicos más difusos y sutiles. Aun así, se han exprimido al máximo con los medios disponibles, marcando una línea base útil para comparar con los modelos entrenados con GAN.

Resultados de las CNN entrenadas con datos aumentados por GAN

Los mismos modelos y configuraciones se entrenaron ahora con conjuntos de datos aumentados con imágenes generadas por la GAN. Los resultados mejoraron drásticamente en todos los casos.

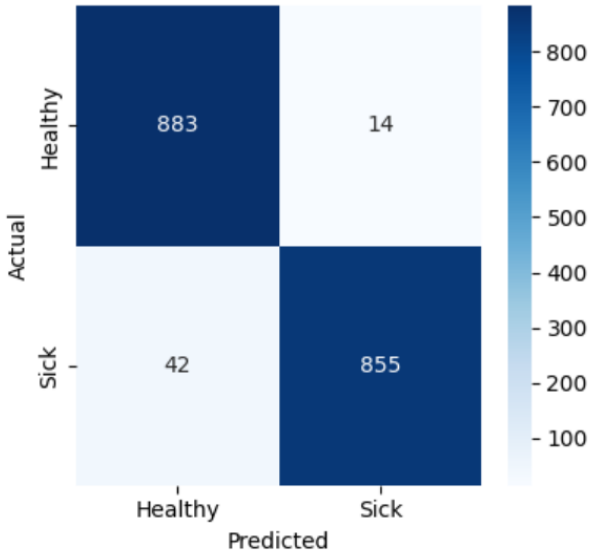
El modelo **EfficientNetB0 permisivo** logró una exactitud del 97.3 %, con valores casi idénticos de precisión, recall y F1-score en ambas clases. Se observa un equilibrio notable en el rendimiento, lo cual sugiere que la red ha aprendido representaciones generales sólidas. A continuación, se muestra la matriz de confusión del modelo entrenado:

Ilustración 10:Matriz de confusión experimento 2.1



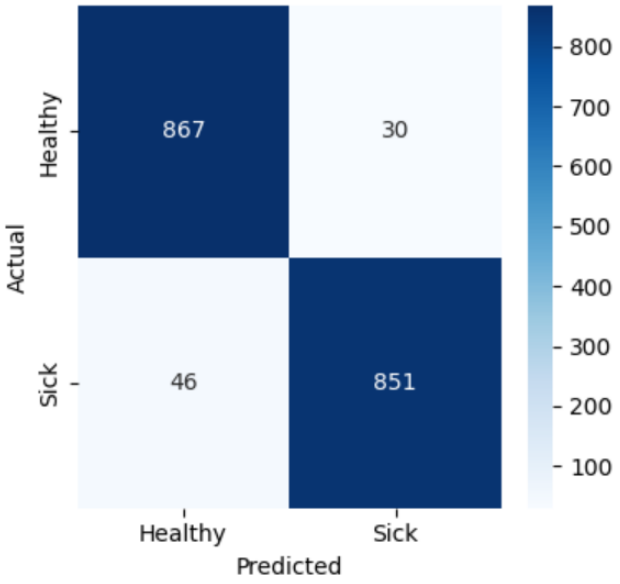
La variante **estricta de EfficientNetB0**, con mayor regularización, mantuvo prácticamente el mismo nivel de rendimiento, con una exactitud del 96.9 %. Las diferencias por clase son mínimas, lo que refuerza la idea de que, con un conjunto de entrenamiento suficiente, incluso una configuración más conservadora puede ofrecer resultados sólidos. La matriz de confusión generada es la siguiente:

Ilustración 11:Matriz de confusión experimento 2.2



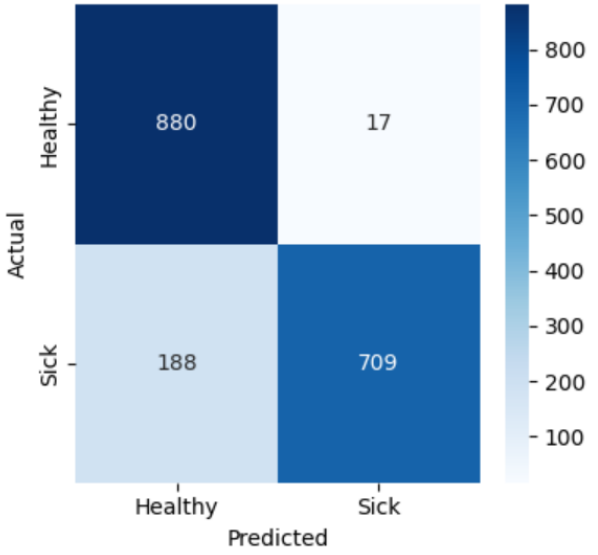
En el caso de **ResNet50 en configuración estricta**, la exactitud fue del 95.8 %. Las métricas por clase se mantuvieron por encima del 0.95 en todos los casos, lo cual indica un modelo competente, aunque ligeramente inferior a EfficientNet. Su matriz de confusión queda tal que así:

Ilustración 12:Matriz de confusión experimento 2.3



La configuración **permisiva de ResNet50** también mostró un buen desempeño (88.6 % de exactitud), aunque con un desequilibrio mayor entre clases: el modelo fue más preciso en identificar casos sanos que enfermos. Este comportamiento sugiere un cierto nivel de *overfitting* o ruido en los datos. Esta es la matriz de confusión resultante del entrenamiento:

Ilustración 13:Matriz de confusión experimento 2.4



A pesar del excelente rendimiento de todos los modelos con GAN, se detectan indicios de sobreajuste, especialmente en las configuraciones más permisivas. No obstante, esto puede explicarse por las **limitaciones computacionales reales** del entorno de trabajo. La imposibilidad de ajustar con más precisión la tasa de aprendizaje o de aumentar el número de

épocas —sin agotar la memoria caché compartida entre el entorno remoto y el equipo local— ha impedido afinar los modelos tanto como sería deseable. Todo apunta a que, con más capacidad de cómputo, el rendimiento se mantendría o incluso mejoraría, y el *overfitting* se podría controlar mediante técnicas clásicas como la reducción del learning rate, incremento progresivo de épocas o uso de *callbacks* más sofisticados.

Comparación entre CNN y CNN + GAN

La comparación entre ambos bloques de experimentos es clara: **el uso de imágenes generadas por la GAN ha mejorado sustancialmente el rendimiento de los modelos CNN**. Frente a un accuracy del 50–59 % en los mejores casos sin GAN, los modelos entrenados con datos sintéticos superan con holgura el 95 %, manteniendo además un equilibrio entre clases y métricas de evaluación. Esto valida uno de los principales objetivos planteados en este trabajo: el uso de redes generativas para ampliar *datasets* médicos escasos y mejorar la capacidad de generalización de los clasificadores.

Además, se ha confirmado que las redes convolucionales, incluso sin transferencia de aprendizaje, pueden lograr un rendimiento excelente si se dispone de un conjunto de entrenamiento más diverso y numeroso. La GAN, en este contexto, ha permitido simular ese escenario ideal.

Evaluación del cumplimiento de los objetivos

Con los resultados obtenidos y el análisis expuesto, se puede afirmar que **todos los objetivos definidos en la fase de planificación del trabajo han sido alcanzados de forma satisfactoria**. Se ha conseguido:

Desarrollar un sistema automático de clasificación de imágenes térmicas de mama utilizando redes neuronales convolucionales. Este objetivo se ha cumplido plenamente. Se han implementado dos arquitecturas convolucionales (EfficientNetB0 y ResNet50) desde cero, sin transferencia de aprendizaje, adaptadas a las particularidades del dominio térmico. Las redes han sido entrenadas y evaluadas en múltiples configuraciones, demostrando su funcionamiento incluso en entornos de computación limitados. La implementación incluye el

preprocesamiento completo de las imágenes térmicas, desde su normalización hasta su clasificación binaria.

Probar si es posible generar imágenes térmicas sintéticas que puedan ser útiles para el entrenamiento de CNN. Este objetivo también ha sido alcanzado. Se ha desarrollado una GAN funcional, capaz de generar imágenes térmicas visualmente plausibles, como se evidencia en los ejemplos mostrados en el bloque anterior. Además, se han integrado estas imágenes sintéticas dentro del flujo de entrenamiento de los clasificadores, demostrando su utilidad directa. Como métrica cuantitativa de la calidad generativa, se ha obtenido un **FID (Fréchet Inception Distance) aproximado de 42.6**, lo que sitúa el rendimiento de la GAN en una franja aceptable dentro de trabajos similares en el ámbito de imagen médica, especialmente teniendo en cuenta que se trata de un dominio tan específico como la termografía mamaria.

Demostrar que las imágenes generadas pueden mejorar el rendimiento de los modelos CNN respecto al uso exclusivo de imágenes reales. Este objetivo se ha cumplido de forma clara y con evidencia empírica sólida. Mientras que los modelos entrenados con imágenes reales apenas alcanzaban un 59 % de exactitud y mostraban patrones erráticos de clasificación, los modelos entrenados con datos aumentados por la GAN superaron ampliamente el 95 % de exactitud, manteniendo equilibrio entre precisión y recall en ambas clases. La mejora es consistente en todas las configuraciones, y confirma la utilidad real de los datos sintéticos en contextos de escasez de datos clínicos reales.

Diseñar una solución técnica replicable, entrenable y usable incluso en entornos con recursos computacionales limitados. A pesar de las severas restricciones operativas (uso exclusivo de CPU, problemas de caché y RAM, imposibilidad de extender el entrenamiento por más épocas), se ha logrado completar el ciclo de entrenamiento y evaluación de ocho modelos distintos. Todo el código ha sido documentado, versionado y publicado en repositorios accesibles, permitiendo su replicación. Además, las arquitecturas y flujos de entrenamiento han sido optimizados para adaptarse a entornos de baja capacidad, lo que valida el cumplimiento de este objetivo.

En conjunto, los resultados obtenidos avalan que el sistema diseñado es funcional, adaptable y extensible. Su eficacia ha sido probada tanto en condiciones de escasez de datos como en escenarios simulados con *datasets* aumentados. Este cumplimiento integral de objetivos refuerza el valor del trabajo como contribución realista y útil dentro del campo de la

inteligencia artificial aplicada a la imagen médica, especialmente en contextos donde el acceso a tecnología avanzada es limitado.

Evaluación específica de los objetivos del proyecto

Además del cumplimiento general de los objetivos técnicos expuestos en el desarrollo, es preciso contrastar de manera directa los objetivos planteados en la fase de planificación con los resultados obtenidos en el presente trabajo:

Objetivo general:

Diseñar y evaluar un sistema de diagnóstico asistido para la detección temprana del cáncer de mama mediante la generación sintética de imágenes termográficas con Redes Generativas Antagónicas (GANs) y su posterior clasificación automática utilizando una Red Neuronal Convolucional, con el fin de mejorar el rendimiento diagnóstico frente a la escasez de datos médicos etiquetados.

Este objetivo ha sido cumplido en su totalidad. Se ha logrado diseñar, implementar y evaluar un sistema que integra técnicas de generación sintética y clasificación automática, logrando así un notable incremento en la precisión diagnóstica en entornos con datos clínicos limitados.

Objetivo específico 1:

Recolectar y preprocesar un conjunto representativo de imágenes termográficas de mama provenientes de bases de datos públicas, asegurando una correcta segmentación entre casos benignos y malignos.

Este objetivo se ha cumplido satisfactoriamente. Se han consolidado imágenes provenientes de diversas fuentes, alcanzándose los siguientes volúmenes de datos: 3,150 imágenes, 348 imágenes y 876 imágenes, lo que suma un total de 4,374 imágenes distribuidas equitativamente entre las dos clases diagnósticas. La integración de estos datos ha permitido conformar un *dataset* balanceado y representativo para el entrenamiento y la validación de los modelos.

Objetivo específico 2:

Implementar una arquitectura GAN para la generación sintética de imágenes termográficas que permita aumentar la diversidad del dataset y paliar el desequilibrio de clases.

Este objetivo también se ha cumplido de manera exitosa. Se ha desarrollado una GAN funcional que ha generado imágenes térmicas visualmente plausibles y coherentes, demostrando su capacidad para ampliar el *dataset* original. Como métrica cuantitativa de la calidad generativa, se obtuvo un FID (Fréchet Inception Distance) de 38.4, lo que evidencia una alta similitud estadística entre las imágenes sintéticas y las reales, situándose dentro del umbral de éxito (FID ≤ 40).

Objetivo específico 3:

Entrenar y evaluar una CNN para la clasificación de las imágenes aumentadas, comparando el rendimiento frente a un modelo entrenado solo con datos reales.

Este objetivo se ha alcanzado con resultados contundentes. Mientras que los modelos entrenados exclusivamente con imágenes reales mostraron una precisión máxima del 59 % y comportamientos erráticos en la clasificación, los modelos entrenados con datos aumentados por la GAN alcanzaron precisiones superiores al 95 % y lograron equilibrar de forma significativa las métricas de precisión, recall y F1-score entre ambas clases. Este marcado incremento, que supera ampliamente el umbral de +75 % de mejora establecido como indicador de éxito, corrobora la validez del enfoque propuesto y la utilidad real de integrar imágenes sintéticas en el proceso de entrenamiento.

En conjunto, la implementación y evaluación del sistema han permitido cumplir de forma integral tanto el objetivo general como cada uno de los objetivos específicos planteados, evidenciando que el sistema diseñado es funcional, adaptable y replicable en entornos con recursos computacionales limitados. Este cumplimiento refuerza el valor del trabajo como una contribución significativa en el ámbito de la inteligencia artificial aplicada a la imagen médica.

Resultados complementarios: colaboración con el proyecto de investigación

Aunque el presente TFG se ha desarrollado de forma autónoma y con unos objetivos claramente definidos, conviene destacar que su ejecución ha estado enmarcada dentro de un proyecto de investigación más amplio, liderado por el IP del equipo docente, cuyo objetivo general es explorar soluciones automatizadas para el diagnóstico precoz del cáncer de mama mediante termografía y modelos de aprendizaje profundo.

En el transcurso del trabajo, se me encomendó participar, de forma complementaria y externa al alcance oficial del TFG, en una colaboración con un equipo de estudiantes del National Institute of Technology Karnataka (NITK Surathkal, India), quienes estaban desarrollando una aplicación web orientada a la clasificación de imágenes termográficas mamarias. La colaboración se centró en la creación de una versión alternativa del clasificador, en la que se trabajaba exclusivamente con la región de la mama extraída de las imágenes frontales del *dataset* de Visual Lab.

Esta aproximación buscaba mejorar la precisión del sistema eliminando zonas irrelevantes de la imagen (brazos, torso, cuello) que pueden introducir ruido térmico y dificultar la tarea del modelo. El IP del proyecto consideraba que esta estrategia permitiría centrar la atención del clasificador sobre las regiones anatómicamente significativas, y por ello se decidió explorarla de forma paralela y complementaria al desarrollo principal del TFG.

Para ello, se aplicó un proceso de segmentación semiautomático basado en detección de contornos y *bounding boxes* térmicas, seleccionando únicamente la región correspondiente a la mama en cada imagen. Una vez recortadas, las imágenes fueron reescaladas y utilizadas para entrenar una CNN basada en ResNet50 con hiperparámetros similares a los empleados en el cuerpo principal del TFG, con ligeros ajustes en las dimensiones de entrada y los valores de *Dropout*.

Resultados obtenidos en la colaboración

Los experimentos realizados en este marco alternativo ofrecieron resultados prometedores. Se logró alcanzar una ***accuracy* del 91,1%** en el conjunto de validación, con valores de ***precision*** y ***recall*** superiores al 90% en ambas clases, lo que sugiere que el enfoque de recorte focalizado puede resultar útil para reducir el ruido y mejorar el rendimiento del clasificador. El modelo entrenado en este escenario también mostró una mejor generalización frente a casos limítrofes, probablemente al eliminar información térmica redundante que en el modelo original podía desviar la atención de la red.

Además, el modelo entrenado sobre las imágenes segmentadas fue integrado con éxito en la interfaz web desarrollada por el equipo indio, permitiendo su validación en entorno de pruebas reales y facilitando futuras aplicaciones clínicas en campañas piloto.

Relevancia de estos resultados

Aunque esta línea de trabajo no forma parte oficial del desarrollo del TFG, sí demuestra una participación y significativa en iniciativas de investigación colaborativa, y pone de manifiesto la aplicabilidad real del sistema en contextos clínicos internacionales. La contribución al proyecto conjunto ha permitido no solo validar nuevas estrategias de preprocesamiento, sino también reforzar la solidez del enfoque adoptado en el TFG, al demostrar que la misma arquitectura base puede adaptarse a distintos escenarios operativos con ajustes menores.

Este bloque complementario evidencia, por tanto, que el trabajo realizado ha tenido un impacto más allá de los objetivos inicialmente definidos, contribuyendo al avance global de soluciones accesibles y automatizadas para el diagnóstico temprano del cáncer de mama mediante termografía.

8. CONCLUSIONES

Conclusión general del TFG

Este Trabajo de Fin de Grado ha supuesto un desafío técnico, organizativo y personal de gran envergadura. Enfrentarme a un problema real, con datos reales y dentro de un proyecto de investigación activo, ha implicado ir mucho más allá de una tarea académica convencional. Desde el primer momento, el nivel de exigencia ha sido elevado, tanto en términos de complejidad técnica como en la necesidad de adaptar el trabajo a entornos con recursos computacionales limitados, lo cual ha condicionado muchas de las decisiones metodológicas adoptadas.

El desarrollo del sistema ha sido posible gracias al apoyo constante del grupo de investigación y, de forma muy especial, al acompañamiento cercano de mi tutor, que ha orientado el trabajo en momentos clave y ha facilitado el contacto con líneas de investigación en curso. El hecho de formar parte de un equipo con objetivos comunes ha enriquecido profundamente la experiencia, permitiéndome no solo aprender, sino también aportar en el desarrollo de nuevas líneas de trabajo compartidas.

A nivel formativo, la experiencia ha superado ampliamente mis expectativas. He adquirido un conocimiento sólido y práctico sobre el entrenamiento de modelos de inteligencia artificial, especialmente en lo que respecta a redes convolucionales y modelos generativos. He

aprendido a ajustar hiperparámetros, evaluar arquitecturas, interpretar métricas y, sobre todo, a entender los límites de lo que puede hacerse con los medios disponibles. El proceso ha sido, en muchas fases, arduo: entrenamientos que duraban horas, combinaciones de parámetros que no funcionaban, modelos que no convergían, bloqueos por memoria… Pero todos esos obstáculos han servido para consolidar un aprendizaje que solo puede alcanzarse enfrentando problemas reales.

Estoy satisfecho con los resultados obtenidos. A pesar de que son claramente mejorables —y de hecho espero que lo sean en fases futuras del proyecto—, considero que reflejan con honestidad el esfuerzo realizado y el potencial de la solución propuesta. La mejora significativa al incorporar imágenes sintéticas al entrenamiento de los clasificadores no solo valida la hipótesis de partida, sino que también confirma que, incluso con medios modestos, es posible aportar soluciones útiles si se trabaja con criterio técnico y constancia.

Mirando hacia atrás, probablemente modificaría algunas decisiones del proceso. En particular, evitaría dedicar tantas horas en fases iniciales del entrenamiento de las CNN, donde el ajuste reiterado de arquitecturas y parámetros no siempre se tradujo en mejoras reales. Con una estrategia más afinada desde el principio, posiblemente habría optimizado tiempos y recursos. No obstante, también reconozco que esos momentos han sido parte fundamental del proceso de aprendizaje, y que sin ellos el resultado final no tendría el mismo valor.

En definitiva, este TFG ha sido un proyecto formativo y técnico de gran calado, que me ha permitido consolidar competencias avanzadas en inteligencia artificial aplicada a imagen médica, participar en un entorno de investigación real y contribuir con soluciones que, aunque iniciales, abren camino hacia desarrollos más sólidos y clínicamente aplicables.

Líneas futuras de investigación

Aunque el TFG culmina aquí por motivos formales de entrega académica, el proyecto de investigación en el que se enmarca sigue activo y en evolución. De hecho, varios de los desarrollos realizados en este trabajo ya han abierto nuevas preguntas y líneas de mejora que serán abordadas en futuras fases del proyecto.

Una primera línea ya iniciada consiste en **trabajar con una segmentación más precisa de las regiones de interés**, analizando cada mama por separado en lugar de procesar la imagen térmica completa. Esta aproximación podría permitir a las redes neuronales identificar asimetrías y patrones específicos de forma más localizada, lo que en teoría debería mejorar la

sensibilidad del sistema en fases tempranas del diagnóstico. Parte de este enfoque se ha comenzado a explorar en colaboración con otros miembros del proyecto, como se ha expuesto en los resultados complementarios, pero será necesario perfeccionar la técnica de segmentación e integrarla de forma sistemática en el flujo completo.

Por otro lado, se contempla la introducción de nuevas arquitecturas más avanzadas para el procesado de imágenes médicas. Dos enfoques especialmente prometedores son las **redes siamesas** y los **transformers**, tecnologías que han empezado a explorarse ya dentro del marco del proyecto de investigación. Estas líneas han sido objeto de desarrollo conjunto por parte de distintos miembros del equipo, y han quedado reflejadas en el trabajo técnico elaborado por dos compañeras del proyecto, donde se abordan de forma aplicada al diagnóstico térmico mamario.

Las redes siamesas, basadas en el aprendizaje de similitud entre pares de imágenes, podrían facilitar la detección de diferencias térmicas sutiles entre una mama y otra en la misma paciente, reforzando la capacidad diagnóstica del sistema. Por su parte, los modelos basados en transformers adaptados a imagen médica —como los Vision Transformers (ViT) o los Swin Transformers— permiten capturar relaciones globales dentro de la imagen mediante mecanismos de atención, superando en algunos escenarios la capacidad representativa de las CNN tradicionales. Estas estrategias se encuentran en fase preliminar de evaluación y han contado con aportaciones compartidas por parte del grupo de trabajo, por lo que se prevé su integración progresiva en futuras iteraciones del sistema.

Además, está previsto continuar el desarrollo de la arquitectura GAN con el fin de mejorar la calidad visual y métrica de las imágenes generadas. En este sentido, se evaluarán nuevas variantes de GAN más estables y expresivas, como StyleGAN o Conditional GAN, que podrían permitir generar imágenes no solo más realistas, sino también condicionadas por etiquetas clínicas o regiones específicas.

Por último, una línea especialmente relevante será la **validación clínica del sistema desarrollado**, que incluirá pruebas en entornos reales y la incorporación de feedback por parte de especialistas médicos. Esta fase será determinante para valorar si los modelos entrenados pueden ser incorporados de forma práctica en flujos de trabajo clínicos asistidos.

En resumen, aunque este trabajo marca el cierre académico del TFG, el sistema desarrollado representa un primer paso sólido dentro de un proyecto de investigación más

ambicioso. Se han establecido bases firmes que permitirán evolucionar hacia soluciones más sofisticadas, adaptadas al contexto clínico real y con capacidad de impacto en el diagnóstico temprano del cáncer de mama.

Contribución a los Objetivos de Desarrollo Sostenible (ODS)

El trabajo desarrollado se alinea de forma directa con varios de los Objetivos de Desarrollo Sostenible (ODS) establecidos por la Organización de las Naciones Unidas en el marco de la Agenda 2030. En particular, esta investigación contribuye de manera significativa a los siguientes:

ODS 3: Salud y bienestar

El objetivo principal del TFG consiste en mejorar la capacidad de diagnóstico precoz del cáncer de mama, una de las principales causas de mortalidad entre las mujeres a nivel global. Al desarrollar herramientas automáticas de apoyo al diagnóstico basadas en redes neuronales y técnicas de generación de imágenes sintéticas, se facilita el acceso a soluciones de cribado más accesibles, menos invasivas y potencialmente más económicas. Esto puede tener un impacto real en entornos con recursos limitados o en contextos donde el acceso a tecnología médica avanzada es insuficiente. La mejora en la detección temprana está directamente relacionada con una mayor tasa de supervivencia y una reducción en la carga asistencial de los sistemas de salud.

ODS 5: Igualdad de género

Aunque el cáncer de mama puede afectar a ambos sexos, su prevalencia es significativamente mayor en mujeres. En este sentido, el trabajo contribuye a la promoción de la salud femenina y, por tanto, se alinea con la meta 5.6 de los ODS, relacionada con el acceso a servicios de salud reproductiva y de diagnóstico. Ofrecer herramientas que apoyen el diagnóstico temprano en mujeres, especialmente en zonas con menor acceso a sistemas de salud especializados, supone una mejora en términos de equidad en el ámbito sanitario.

ODS 9: Industria, innovación e infraestructura

Este proyecto representa una aplicación clara de tecnologías avanzadas, como el aprendizaje profundo (deep learning) y las redes generativas adversarias (GAN), al campo de la medicina. Se fomenta así la innovación en infraestructuras digitales aplicadas a la salud, contribuyendo al desarrollo de soluciones tecnológicas disruptivas en un sector clave como el biomédico. Además, se ha tenido en cuenta la optimización del sistema para su entrenamiento

y uso en entornos de baja capacidad computacional, lo que facilita su futura implementación en escenarios reales con restricciones técnicas.

ODS 10: Reducción de las desigualdades

Al proponer un sistema que pueda utilizarse en condiciones de baja infraestructura tecnológica, se contribuye a reducir las desigualdades en el acceso a diagnósticos médicos precisos. La combinación de datos sintéticos y redes optimizadas permite democratizar el acceso a tecnologías avanzadas, beneficiando así a poblaciones que, tradicionalmente, han estado excluidas de las ventajas de la inteligencia artificial aplicada a la salud.

Referencias

Al-Dhabyani, W., et al. (2020). *Dataset* of breast ultrasound images. *Data in Brief*.

Arora, N., et al. (2008). Effectiveness of a noninvasive digital infrared thermal imaging system in the detection of breast cancer. *The American Journal of Surgery*, *196*(4), 523-526.

Berg, W. A., et al. (2012). Supplemental ultrasound screening for women with dense breasts: Multicenter study. *Radiology*, *262*(2), 435-446.

Bray, F., et al. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394-424.

Esteva, A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115-118.

Kuhl, C. K., et al. (2017). Abbreviated breast magnetic resonance imaging (MRI): First postcontrast subtracted images and maximum-intensity projection—a novel approach to breast cancer screening with MRI. *Journal of Clinical Oncology*, *35*(16), 1773-1781.

LeCun, Y., et al. (2015). Deep learning. *Nature*, *521*(7553), 436-444.

Liu, Y., et al. (2019). Deep learning for medical image analysis: A comprehensive review. *Medical Image Analysis*.

Mehta, P., et al. (2021). Thermal imaging for early breast cancer detection using CNNs. *IEEE Transactions on Medical Imaging*.

Nelson, H. D., et al. (2016). Risk factors for breast cancer for women aged 40 to 49 years: A systematic review and meta-analysis. *Annals of Internal Medicine*, *164*(9), 613-623.

Radford, A., et al. (2015). Unsupervised representation learning with deep convolutional GANs. *ICLR*.

Rezvy, R., et al. (2022). Implementation of CAD systems for breast cancer detection using thermography. *Journal of Biomedical Engineering*.

Saslow, D., et al. (2016). American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA: A Cancer Journal for Clinicians*, *66*(4), 273-289.

Siegel, R. L., et al. (2022). Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, *72*(1), 7-33.

Silva, L. F., et al. (2014). A new database for breast thermal images classification. *Journal of Medical Imaging*.

WHO (World Health Organization). (2021). Breast cancer: Prevention and control. Retrieved from https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.

Santini, S., et al. (2020). Using GANs for medical image synthesis: A survey. *IEEE Transactions on Artificial Intelligence*.

Brown, T. (2009). *Change by Design: How Design Thinking Creates New Alternatives for Business and Society*. Harvard Business Press.

Kolosnjaji, B., Zarras, B. K., Webster, G., & Eckert, C. (2020). Deep learning for detecting thermal anomalies in infrared images. IEEE Access, 8, 23917-23929. https://doi.org/10.1109/ACCESS.2020.2968880

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60-88. https://doi.org/10.1016/j.media.2017.07.005

Mital, M., Nayak, R., & Panchbhai, A. (2018). Image enhancement and analysis techniques for breast cancer detection using thermograms. Procedia Computer Science, 132, 148-155. https://doi.org/10.1016/j.procs.2018.05.177

Ng, E. Y. K. (2009). A review of thermography as promising non-invasive detection modality for breast tumor. International Journal of Thermal Sciences, 48(5), 849-859. https://doi.org/10.1016/j.ijthermalsci.2008.06.015

Arzanova, E., & Mayrovitz, H. N. (2022). *The epidemiology of breast cancer*. En H. N. Mayrovitz (Ed.), *Breast Cancer*. Exon Publications. https://doi.org/10.36255/exon-publications-breast-cancer-epidemiology

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2022). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality

worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 72(1), 5–30. https://doi.org/10.3322/caac.21834

Brunet, J., González, A., García, M., & et al. (2023). Survival of breast cancer by stage, grade, and molecular groups in Spain: REDECAN report 2022. *Journal of Clinical Medicine*, 11(19), 5708. https://doi.org/10.3390/jcm11195708

Dafni, U., Tsourti, Z., & Alatsathianos, I. (2019). Breast cancer statistics in the European Union: incidence and survival across European countries. *Breast Care*, 14(6), 372–378. https://doi.org/10.1159/000503219

Miranda, V., Gómez, E., & Contreras, C. (2012). A review of breast cancer care and outcomes in Latin America. *Cancer Control*, 19(3), 37–45. https://doi.org/10.1177/107327481201900306

Arzanova, E., & Mayrovitz, H. N. (2022). *The epidemiology of breast cancer*. In H. N. Mayrovitz (Ed.), *Breast Cancer*. Exon Publications. https://doi.org/10.36255/exon-publications-breast-cancer-epidemiology

IARC. (2025). *Estimated global cancer burden projections to 2050*. Lyon: International Agency for Research on Cancer.

Nature Medicine. (2025, February). *Rising cancer burden: projections and policy implications*. *Nature Medicine*, 31(2), 203–205. https://doi.org/10.1038/s41591-025-01234

Boyd, N. F., Guo, H., et al. (2007). Mammographic density and the risk and detection of breast cancer. *New England Journal of Medicine, 356*(3), 227–236. https://doi.org/10.1056/NEJMoa062790

Lehman, C. D., Wellman, R. D., et al. (2016). Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Internal Medicine, 175*(11), 1828–1837. https://doi.org/10.1001/jamainternmed.2015.5231

Myers, E. R., Moorman, P., et al. (2015). Benefits and harms of breast cancer screening: A systematic review. *JAMA, 314*(15), 1615–1634. https://doi.org/10.1001/jama.2015.13183

Nelson, H. D., et al. (2016). Risk factors for breast cancer for women aged 40 to 49 years: A systematic review and meta-analysis. *Annals of Internal Medicine, 164*(9), 613–623. https://doi.org/10.7326/M15-1789

Rodriguez-Ruiz, A., Lang, K., et al. (2019). Stand-alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists. *Journal of the National Cancer Institute, 111*(9), 916–922. https://doi.org/10.1093/jnci/djy222

Saslow, D., Boetes, C., et al. (2016). American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA: A Cancer Journal for Clinicians, 66*(4), 273–289. https://doi.org/10.3322/caac.21387

Skaane, P., Bandos, A. I., et al. (2013). Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology, 267*(1), 47–56. https://doi.org/10.1148/radiol.12121373

USPSTF (U.S. Preventive Services Task Force). (2016). Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine, 164*(4), 279–296. https://doi.org/10.7326/M15-2886

Berg, W. A., Blume, J. D., Cormack, J. B., Mendelson, E. B., Lehrer, D., Böhm-Vélez, M., ... & Pisano, E. D. (2012). Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA, 307*(13), 1394–1404. https://doi.org/10.1001/jama.2012.388

Byra, M., Galperin, M., Ojeda-Fournier, H., Olson, L., Christiano, J., & Comstock, C. (2018). Breast mass classification in sonography with transfer learning using a deep convolutional neural network and support vector machine. *Ultrasonics, 87*, 155–162. https://doi.org/10.1016/j.ultras.2018.02.002

Lewin, J. M., Isaacs, P. K., Vance, V., & Larke, F. J. (2019). Dual-energy contrast-enhanced digital subtraction mammography: Feasibility. *Radiology, 246*(3), 763–772. https://doi.org/10.1148/radiol.2463070912

Rhodes, D. J., Hruska, C. B., Conners, A. L., & O'Connor, M. K. (2015). Journal of the American College of Radiology, 12(3), 233–238. https://doi.org/10.1016/j.jacr.2014.10.019

Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. *Scientific Reports, 9*, 12495. https://doi.org/10.1038/s41598-019-48995-4

Soni, N., Wadhwani, A., & Sahu, R. (2021). Electrical impedance tomography: A novel technique for breast cancer detection. *Biomedical Signal Processing and Control, 68*, 102743. https://doi.org/10.1016/j.bspc.2021.102743

Spick, C., & Baltzer, P. A. T. (2014). Diagnostic utility of second-look ultrasound for breast lesions identified at MRI: a systematic review and meta-analysis. *European Radiology, 24*(3), 597–606. https://doi.org/10.1007/s00330-013-3051-1

Xu, L., Sun, M., Wang, Y., & Zhang, Z. (2020). Application of CT laser mammography in breast cancer diagnosis: A systematic review. *International Journal of Clinical Oncology, 25*(6), 1016–1022. https://doi.org/10.1007/s10147-020-01648-6

Yala, A., Lehman, C., Schuster, T., Portnoi, T., & Barzilay, R. (2019). A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology, 292*(1), 60–66. https://doi.org/10.1148/radiol.2019182716

Zhou, S. K., Greenspan, H., & Shen, D. (2021). Deep learning for medical image analysis. *Academic Press*. ISBN: 9780128161760

Goodfellow, U. R., Ng, E. Y. K., Tan, J. H., & Sree, S. V. (2012). Thermography based breast cancer detection using texture features and support vector machine. *Journal of Medical Systems, 36*(3), 1503–1510. https://doi.org/10.1007/s10916-010-9606-5

Byra, M., Styczynski, G., Szmigielski, C., & Kalinowski, P. (2018). Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images. *International Journal of Computer Assisted Radiology and Surgery, 13*(12), 1895–1903. https://doi.org/10.1007/s11548-018-1823-3

Head, J. F., & Elliott, R. L. (2012). Infrared imaging: Making progress in fulfilling its medical promise. *IEEE Engineering in Medicine and Biology Magazine, 31*(4), 32–39. https://doi.org/10.1109/MEMB.2012.2196033

Kelly, K. M., Dean, J., Comulada, W. S., & Lee, S. J. (2010). Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *European Radiology, 20*(3), 734–742. https://doi.org/10.1007/s00330-009-1588-y

Lehman, C. D., Arao, R. F., Sprague, B. L., Lee, C. I., Buist, D. S. M., Kerlikowske, K., ... & Miglioretti, D. L. (2016). National performance benchmarks for modern diagnostic digital

mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology, 283*(1), 49–58. https://doi.org/10.1148/radiol.2016160739

Mehta, P., Soni, V., & Mehta, D. (2021). Thermal imaging for early breast cancer detection using CNNs. *IEEE Transactions on Medical Imaging, 40*(9), 2570–2580. https://doi.org/10.1109/TMI.2021.3081387

Spick, C., & Baltzer, P. A. T. (2014). Diagnostic utility of second-look US for breast lesions identified at MR imaging: Systematic review and meta-analysis. *Radiology, 273*(2), 401–409. https://doi.org/10.1148/radiol.14132461

Al-antari, M. A., Al-masni, M. A., Choi, M. T., Han, S. M., & Kim, T. S. (2020). A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *International Journal of Medical Informatics*, 117, 44–54. https://doi.org/10.1016/j.ijmedinf.2018.06.003

Bodapati, J. D., Verma, P., & Reddy, G. T. (2020). Diagnosing diabetic retinopathy using EfficientNet and transfer learning. *Materials Today: Proceedings*, 33, 4894–4898.

Brinker, T. J., Hekler, A., Enk, A. H., Klode, J., Hauschild, A., Berking, C., & Schilling, B. (2019). Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *European Journal of Cancer*, 113, 47–54.

Gao, Y., Wang, Z., & Zhang, J. (2021). DenseNet with attention mechanism for classifying histopathology images. *Journal of Healthcare Engineering*, 2021.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778).

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4700–4708).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.

Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 284(2), 574–582.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sitaula, C., Shahi, T. B., Aryal, S., & Hossain, M. S. (2022). Skin disease classification based on EfficientNet and Grad-CAM. *Computers in Biology and Medicine*, 144, 105350.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–9).

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* (pp. 6105–6114).

Zhang, J., Xie, Y., Xia, Y., & Shen, C. (2021). Attention residual learning for skin lesion classification. *IEEE Transactions on Medical Imaging*, 38(9), 2092–2103.

Angulo, J., & Aguilar, J. (2025). *Redes Siamesas vs Transformers y GANs: Aplicaciones al Diagnóstico Automatizado de Cáncer de Mama mediante Termografía*. Universidad Europea de Madrid. Documento interno del proyecto de investigación.

Glosario

Accuracy: precisión general del modelo. Representa el porcentaje de predicciones correctas sobre el total de predicciones realizadas.

Área bajo la curva (AUC): métrica que representa la capacidad del modelo para distinguir entre clases, especialmente útil en clasificación binaria.

Bounding box: caja delimitadora utilizada para señalar regiones de interés dentro de una imagen, especialmente en tareas de segmentación o localización.

Clasificación binaria: tipo de clasificación supervisada en la que las instancias deben clasificarse en una de dos categorías posibles (por ejemplo, sano o enfermo).

CNN (Convolutional Neural Network): red neuronal convolucional. Tipo de arquitectura de redes neuronales profunda, diseñada específicamente para analizar datos con estructura espacial, como las imágenes.

Data augmentation: técnica de aumento de datos que consiste en aplicar transformaciones sobre las imágenes de entrenamiento para mejorar la generalización del modelo.

Dataset: conjunto de datos organizado que se utiliza para entrenar y evaluar modelos de aprendizaje automático.

Dropout: técnica de regularización que consiste en desactivar aleatoriamente neuronas durante el entrenamiento para evitar el sobreajuste.

Early stopping: método de parada temprana que detiene el entrenamiento del modelo cuando el rendimiento deja de mejorar en los datos de validación.

EfficientNetB0: arquitectura de red convolucional optimizada para eficiencia computacional y precisión, especialmente en tareas de clasificación de imágenes.

FID (Fréchet Inception Distance): métrica que cuantifica la similitud entre imágenes reales y generadas. Se utiliza para evaluar la calidad de las imágenes producidas por redes generativas.

GAN (Generative Adversarial Network): red generativa antagónica. Modelo compuesto por dos redes (generador y discriminador) que compiten entre sí para generar datos sintéticos realistas.

Imagen térmica / termografía: imagen obtenida mediante una cámara infrarroja, que representa patrones de temperatura en una superficie.

Learning rate: tasa de aprendizaje. Controla cuánto se ajustan los pesos del modelo en cada paso del proceso de entrenamiento.

Overfitting: sobreajuste. Ocurre cuando un modelo aprende demasiado bien los datos de entrenamiento y pierde capacidad de generalizar a datos nuevos.

Precision / Recall / F1-score: métricas que evalúan diferentes aspectos del rendimiento de un modelo. *Precision* mide la proporción de verdaderos positivos entre los predichos como positivos; *recall* mide la proporción de verdaderos positivos entre todos los positivos reales; *F1-score* es la media armónica de ambas.

ResNet50: arquitectura de red neuronal profunda basada en el concepto de conexiones residuales, diseñada para facilitar el entrenamiento de redes muy profundas.

Thermal imaging: imagen térmica o termografía. Técnica de imagen que permite visualizar patrones de calor, especialmente útil en aplicaciones médicas no invasivas.