Universidad Nacional de Educación a Distancia
Escuela Técnica Superior de Ingeniería Informática

Trabajo Fin de Máster
Ingeniería y Ciencia de Datos

Deep Learning y Mecanismos de Atención para la detección del cáncer de mama a través de imágenes térmicas

Autor
Juan Carlos Ortegón Aguilar

Dirigido por
José Manuel Cuadra Troncoso
Mariano Rincón Zamorano

Curso: 2022-2023: Convocatoria extraordinaria de septiembre

Agradecimientos

Quisiera dedicar este trabajo a mi familia y amigos por el apoyo y motivación a lo largo de todo el recorrido. Agradecer también a José Manuel Cuadra Troncoso y Mariano Rincón Zamorano por haber depositado su confianza en mí y haber guiado y supervisado este proyecto.

Resumen

De acuerdo con la Organización Mundial de la Salud (OMS), el carcinoma de mama es el tipo de cáncer más común, con más de 2,2 millones de casos en 2020. Cerca de una de cada 12 mujeres enfermarán de cáncer de mama a lo largo de su vida, siendo la principal causa de mortalidad en las mujeres. Dado el riesgo de muerte asociado a la metástasis durante las últimas fases del cáncer, la detección temprana y precisa del cáncer de mama es un desafío crucial en el campo de la medicina. La Iniciativa de la Comisión Europea para el cribado y diagnóstico del cáncer de mama recomienda el cribado mediante mamografía digital, la ultrasonografía manual y la ultrasonografía mamaria, entre otras técnicas. Sin embargo, estas metodologías tradicionales presentan limitaciones en términos de accesibilidad, costo y en algunos casos, la exposición a la radiación. Por otro lado, diferentes estudios han demostrado que el uso de imágenes térmicas puede ser una herramienta prometedora para la detección del cáncer de mama, siendo una técnica no invasiva, indolora y libre de radiación.

Recientemente, gracias a los avances en las técnicas de aprendizaje automático y deep learning, la detección y el diagnóstico del cáncer de mama mediante estas técnicas son cada vez más eficaces. Asimismo, se ha demostrado que la aplicación de mecanismos de atención produce resultados prometedores en este contexto. Es por ello que en este trabajo se pretende contribuir al campo de la detección del cáncer de mama aplicando técnicas de aprendizaje profundo junto con mecanismos de atención sobre imágenes térmicas usando la base de datos DMR-IR, desarrollada por el Hospital Universitário Antônio Pedro (HUAP) de la Universidad Federal Fluminense (Brasil).

El propósito de este estudio consiste en analizar cómo los mecanismos de atención impactan en el diagnóstico del cáncer de mama a partir de imágenes térmicas. Para este fin, se han desarrollado cuatro modelos diferentes. El primero de estos modelos es una Red Neuronal Convolucional (CNN) que sirve como punto de partida y se considera el modelo base. Los tres modelos restantes son variaciones de las CNNs que incorporan mecanismos de atención: Auto-atención, Atención Suave y Atención Dura. Estas variantes se diseñan con la finalidad de explorar cómo la inclusión de estos mecanismos puede mejorar la capacidad de detección y clasificación del cáncer de mama en imágenes térmicas.

Diferentes experimentos se han llevado a cabo en este estudio. En primer lugar, se ha empleado la validación cruzada para ajustar los parámetros de los modelos, buscando mejorar su rendimiento y prevenir el sobreajuste. Asimismo, se ha aplicado la técnica de Grad-CAM con el propósito de analizar cómo cada uno de los modelos interpreta y utiliza la información visual para realizar sus predicciones. En cuanto a la evaluación de los modelos, se consideraron métricas fundamentales como la exactitud, especificidad, precisión, exhaustividad, puntuación F1, así como el área bajo la curva ROC (AUC). Estas métricas nos proporcionan una comprensión del rendimiento de los modelos.

Los resultados obtenidos en la evaluación de los modelos revelan un incremento en el rendimiento gracias a la aplicación de mecanismos de atención. El modelo base ha alcanzado una tasa de exactitud del 86.21 %. Por otro lado, las variantes de las CNNs que incorporan mecanismos de atención mostraron mejoras significativas en su rendimiento. La CNN con Auto-atención y la CNN con Atención Suave han logrado una exactitud del 91.38 %, mientras que la que implementa Atención Dura ha alcanzado un 93.10 %. Estos resultados sugieren que la implementación de mecanismos de atención en los modelos puede contribuir de manera significativa a la mejora en la precisión en la detección del cáncer de mama mediante imágenes térmicas.

Palabras clave: Aprendizaje Profundo, Red Neuronal Convolucional, Cáncer de mama, Termografía, Mecanismos de Atención, Auto-atención, Atencion Suave, Atención Dura.

Abstract

According to the World Health Organization (WHO), breast carcinoma is the most common type of cancer, with over 2.2 million cases in 2020. Nearly one out of every 12 women will develop breast cancer during their lifetime, making it a leading cause of mortality among women. Given the risk of death associated with metastasis in the later stages of cancer, early and accurate detection of breast cancer is a critical challenge in the field of medicine. The European Commission Initiative for Breast Cancer Screening and Diagnosis recommends screening through digital mammography, manual ultrasound and automated breast ultrasound, among other techniques. However, these traditional methodologies have limitations in terms of accessibility, cost, and, in some cases, radiation exposure. On the other hand, various studies have shown that thermal imaging can be a promising tool for breast cancer detection, being a non-invasive, painless, and radiation-free technique.

Recent advancements in machine learning and deep learning techniques have improved the effectiveness of breast cancer detection and diagnosis. Moreover, the application of attention mechanisms has demonstrated promising results in this context. Hence, this work aims to contribute to the field of breast cancer detection by applying deep learning techniques along with attention mechanisms to thermal images, using the DMR-IR database, developed by the Hospital Universitário Antônio Pedro (HUAP) of the Federal Fluminense University (Brazil).

The objetive of this study is to analyze how attention mechanisms impact the diagnosis of breast cancer from thermal images. For this purpose, four different models have been developed. The first of these models is a Convolutional Neural Network (CNN), which serves as a starting point and is considered the base model. The remaining three models are variations of CNNs that incorporate attention mechanisms: Self-attention, Soft Attention, and Hard Attention. These variants are designed to explore how the inclusion of these mechanisms can enhance the detection and classification capabilities of breast cancer in thermal images.

Various experiments have been conducted in this study. Firstly, cross-validation has been employed to fine-tune model parameters, aiming to improve performance and prevent overfitting. Additionally, the Grad-CAM technique has been applied to analyze how each model interprets and utilizes visual information to make predictions. In terms of model evaluation, fundamental metrics such as accuracy, specificity, precision, recall, F1 score, and the Area Under the ROC Curve (AUC) have been considered. These metrics provide insights into model performance.

The results obtained in the model evaluation reveal an increase in performance due to the application of attention mechanisms. The base model achieved an accuracy rate of 86.21 %. On the other hand, the CNN variants incorporating attention mechanisms showed significant improvements in their performance. The CNN with Self-attention and the CNN with Soft Attention achieved an accuracy of 91.38 %, while the one implementing Hard Attention

reached 93.10 %. These results suggest that the implementation of attention mechanisms in the models can significantly contribute to improving the accuracy in breast cancer detection using thermal images.

Keywords: Deep Learning, Convolutional Neural Network, Breast Cancer, Thermography, Attention Mechanisms, Self-Attention, Soft Attention, Hard Attention.

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Introducción general y objetivos

El propósito de este primer capítulo es proporcionar una introducción a la problemática que abordaremos a lo largo de este trabajo fin de máster. Comenzaremos por establecer el contexto general en el que se enmarca nuestro estudio, destacando la importancia de la temática elegida. A continuación, estableceremos los objetivos de éste, definiendo que deseamos lograr. Finalmente, presentaremos la metodología que se ha seguido y proporcionaremos una vista previa de la estructura y organización del documento.

1.1. Motivación

De acuerdo con la Organización Mundial de la Salud (OMS) [Organización Mundial de la Salud, 2023], el carcinoma de mama es el tipo más común de cáncer, habiéndose registrado más de 2.2 millones de casos en el año 2020. Cerca de una de cada 12 mujeres experimentará el desarrollo de cáncer de mama a lo largo de su vida, convirtiéndolo en la principal causa de mortalidad entre las mujeres. En el año 2020, alrededor de 685 000 mujeres fallecieron debido a esta enfermedad.

El cáncer de mama se origina en las células del revestimiento (epitelio) de los conductos mamarios (en un 85 %) o en los lóbulos (en un 15 %) del tejido glandular en los senos. Inicialmente, el tumor canceroso se encuentra limitado en el conducto o lóbulo (en un estado llamado *in situ*), en donde generalmente no produce síntomas y tiene un potencial mínimo de diseminación (metástasis).

A medida que avanza el tiempo, este cáncer *in situ* (etapa 0) puede evolucionar y empezar a invadir los tejidos mamarios circundantes (convirtiéndose en cáncer de mama invasivo), y posteriormente puede propagarse a los ganglios linfáticos cercanos (metástasis regional) o a otros órganos del cuerpo (metástasis distante). La metástasis generalizada es la causa de fallecimiento en casos en los que una mujer muere a causa de cáncer de mama [Organización Mundial de la Salud, 2023].

Dado el riesgo de muerte asociado a la metástasis durante las últimas fases del cáncer, la detección temprana y precisa del cáncer de mama es un desafío crucial en el campo de la medicina, ya que el diagnóstico precoz puede significar la diferencia entre un tratamiento exitoso y un pronóstico desfavorable. Por esa razón, la Iniciativa de la Comisión Europea para el cribado y diagnóstico del cáncer de mama [Schünemann et al., 2020] sugiere llevar a cabo el cribado mediante mamografía digital, además de considerar la incorporación de ultrasonografía manual, ultrasonografía mamaria automatizada o resonancia magnética, en comparación con la mamografía en solitario, para mujeres de edades comprendidas entre los 40 y 75 años.

Aunque, como se ha mencionado, existen diversas técnicas y tecnologías para la detección de esta enfermedad, estas metodologías tradicionales presentan limitaciones en términos de accesibilidad, costo y en algunos casos, la exposición a la radiación. Sin embargo, diferentes estudios [Rakhunde et al., 2022, Kennedy et al., 2009] han demostrado que el uso de imágenes térmicas puede ser una herramienta prometedora para la detección del cáncer de mama. Estas imágenes capturan la distribución de la temperatura en la superficie del cuerpo, lo que puede reflejar patrones de flujo sanguíneo y metabólico asociados con la presencia de tumores.

Algunas de las ventajas que se mencionan en [Rakhunde et al., 2022] sobre esta técnica son:

- Procedimiento simple y similar a tomar una fotografía.
- No emite radiación ni requiere inyecciones, lo que minimiza el riesgo de daño a las estructuras celulares frágiles.
- No invasiva y sin dolor, ya que no involucra compresión ni contacto directo.
- Costo efectivo en comparación con otras técnicas de diagnóstico por imágenes.
- Permite la detección temprana de malignidades, lo que posibilita la investigación y la intervención oportuna antes de que aparezcan signos evidentes.
- Adecuada para el tamizaje de la salud de la mujer en etapas tempranas, en mujeres de todas las edades, desde la preadolescencia hasta la postmenopausia.
- Puede utilizarse en mujeres con diversas formas y tamaños de senos, incluyendo aquellas con tejido mamario denso, fibroquísticos, embarazadas, en período de lactancia y posmenopáusicas. Permite la detección en mujeres bajo terapia de reemplazo hormonal.
- Sobresale en la detección temprana de cambios como vasodilatación, angiogénesis y cambios en los vasos sanguíneos, características presentes en las primeras etapas del cáncer.

Aunque, desafortunadamente, también presenta ciertas desventajas:

- Dificultad para distinguir la causa del aumento de temperatura.
- Partes cálidas en el seno pueden indicar tanto cáncer como condiciones no cancerosas como mastitis.
- Inflamación en el seno debido a infecciones bacterianas o virales puede aumentar la temperatura de los tejidos, afectando los resultados de la termografía.

Recientemente, gracias a los avances en las técnicas de aprendizaje automático y deep learning, la detección y el diagnóstico del cáncer de mama mediante estas técnicas son cada vez más eficaces. Asimismo, se ha demostrado que la aplicación de mecanismos de atención produce resultados prometedores para la detección del cáncer de mama [Toğaçar et al., 2020]. De esta manera, se logra enfocar la atención en las regiones destacadas de la imagen en lugar de considerar todas las partes de ésta como igualmente importantes, lo que, a su vez, incrementa las posibilidades de potenciar el rendimiento de las técnicas empleadas.

Es por ello que en este trabajo se pretende contribuir al campo de la detección del cáncer de mama aplicando técnicas de aprendizaje profundo junto con mecanismos de atención sobre imágenes térmicas.

1.2. Objetivos

El propósito de este trabajo es lograr la detección del cáncer de mama utilizando imágenes infrarrojas y aplicando mecanismos de atención. Específicamente, busca clasificar a cada paciente como sano o enfermo. Este objetivo principal se puede dividir en una serie de subobjetivos, que incluyen:

- Obtener un conjunto de imágenes infrarrojas de calidad.
- Analizar la influencia de distintos mecanismos de atención en el comportamiento de diferentes modelos de aprendizaje profundo para la tarea de clasificación del cáncer de mama en imágenes térmicas.
- Analizar las áreas de enfoque de los mecanismos de atención durante el proceso de clasificación.
- Construir modelos de deep learning competitivos para el diagnóstico del cáncer de mama a partir de imágenes termográficas.

1.3. Metodología

La obtención de un conjunto de imágenes infrarrojas de calidad es un paso fundamental en el éxito de este estudio. En esta etapa, se llevará a cabo la recopilación de datos utilizando la base de datos para la Investigación Mastológica con Imágenes Infrarrojas (DMR-IR). Este conjunto de datos adquirido será sometido a un proceso de limpieza y preprocesamiento para garantizar su calidad y coherencia. Además, con el fin de enriquecer y ampliar el conjunto de datos, se aplicarán técnicas de aumento de datos que incluirán el volteo horizontal y la alteración del brillo, lo que contribuirá a mejorar la robustez de los modelos y su capacidad para generalizar a partir de una cantidad limitada de datos originales.

Con el propósito de lograr la detección del cáncer de mama utilizando el conjunto de datos previamente mencionado, se llevará a cabo el desarrollo de diversos modelos de Deep Learning. Esto incluirá la implementación de variantes de redes neuronales convolucionales que incorporan mecanismos de atención. Estos mecanismos permitirán que el modelo se centre en áreas específicas de la imagen durante la clasificación, lo que podría mejorar la precisión de la detección.

Para comprender cómo los modelos interpretan y utilizan la información visual para realizar predicciones, se aplicará Grad-CAM. Esta técnica permitirá la visualización de las áreas de enfoque de los mecanismos de atención durante el proceso de clasificación.

Finalmente, se realizará una evaluación de los resultados obtenidos en este estudio. Se analizarán métricas clave, como la exactitud, especificidad, precisión, exhaustividad y la puntuación F1, para evaluar el rendimiento de los modelos. Además, se compararán los resultados con investigaciones previas relacionadas con la detección de cáncer de mama utilizando imágenes.

1.4. Organización del documento

La estructura de este trabajo es la siguiente. En este capítulo introductorio, se ha presentado la motivación detrás del trabajo y se han establecido los objetivos que se buscan alcanzar. Además, se describe la metodología general del estudio y cómo está organizado el documento.

El capítulo 2 se centra en los fundamentos teóricos esenciales, incluyendo el Aprendizaje Profundo, las Redes Neuronales Convolucionales (CNNs), los mecanismos de atención, la validación cruzada, Grad-CAM y las métricas de evaluación. También se exploran trabajos relacionados en el ámbito de la detección del cáncer de mama mediante el uso de imágenes y se destacan sus contribuciones.

En el capítulo 3 se explora el conjunto de datos utilizado en el estudio, incluyendo detalles sobre su adquisición, preprocesamiento y se destaca la relevancia del aumento de datos,

una estrategia clave en el desarrollo de este trabajo. Además, se presenta en detalle los modelos propuestos. Se discuten aspectos sobre las decisiones de diseño tomadas así como el preprocesamiento de las imágenes, la extracción de características, la inclusión de capas de atención, etc.

En el capítulo 4, se describen los experimentos llevados a cabo en este estudio. Estos experimentos abarcan la validación cruzada, utilizada para mejorar y ajustar los modelos, así como el empleo de Grad-CAM para entender las áreas de interés de los mecanismos de atención durante el proceso de clasificación. Además, este capítulo presenta en profundidad los resultados obtenidos y concluye con un análisis y comparación de éstos con trabajos previos de otros autores.

El capítulo 5 se dedica al análisis de las conclusiones obtenidas a partir del trabajo realizado, además de presentar propuestas para futuras investigaciones.

Finalmente, el apéndice A contiene una descripción más detallada del conjunto de datos visto en el capítulo 3.1, incluyendo los problemas relacionados con los procesos de descarga y limpieza.

Capítulo 2

Estado del arte

Este capítulo proporciona una base fundamental para nuestro estudio. Comenzaremos con una introducción a conceptos esenciales como el Aprendizaje Profundo y las Redes Neuronales Convolucionales, junto con los mecanismos de atención, validación cruzada, Grad-CAM y métricas de evaluación.

En la segunda sección, exploraremos investigaciones relacionadas con la detección del cáncer de mama, en particular en imágenes histopatológicas, mamográficas y termográficas. Esto nos brindará una visión general del estado actual de la investigación en esta área clave.

2.1. Fundamentos

A continuación, presentaremos una introducción a conceptos fundamentales, que incluyen Aprendizaje Profundo, Redes Neuronales Convolucionales, así como mecanismos de atención, validación cruzada, Grad-CAM y métricas de evaluación.

2.1.1. Aprendizaje Profundo y Redes Neuronales Convolucionales (CNNs)

El Aprendizaje Profundo o Deep Learning en inglés, representa una rama del campo de la Inteligencia Artificial que busca emular la capacidad del cerebro humano para procesar información y extraer patrones complejos a partir de datos. En su esencia, el Aprendizaje Profundo se basa en la construcción de modelos computacionales llamados redes neuronales, que están compuestas por capas interconectadas de neuronas artificiales.

Uno de los pilares fundamentales del Aprendizaje Profundo es la categoría de redes neuronales conocida como Redes Neuronales Convolucionales (CNN por sus siglas en inglés, que significa *Convolutional Neural Network*) [Alzubaidi et al., 2021]. La principal ventaja de las CNNs es que identifican automáticamente las características relevantes sin supervisión humana [Gu et al., 2018]. Las CNNs se han aplicado ampliamente en diversos campos, como

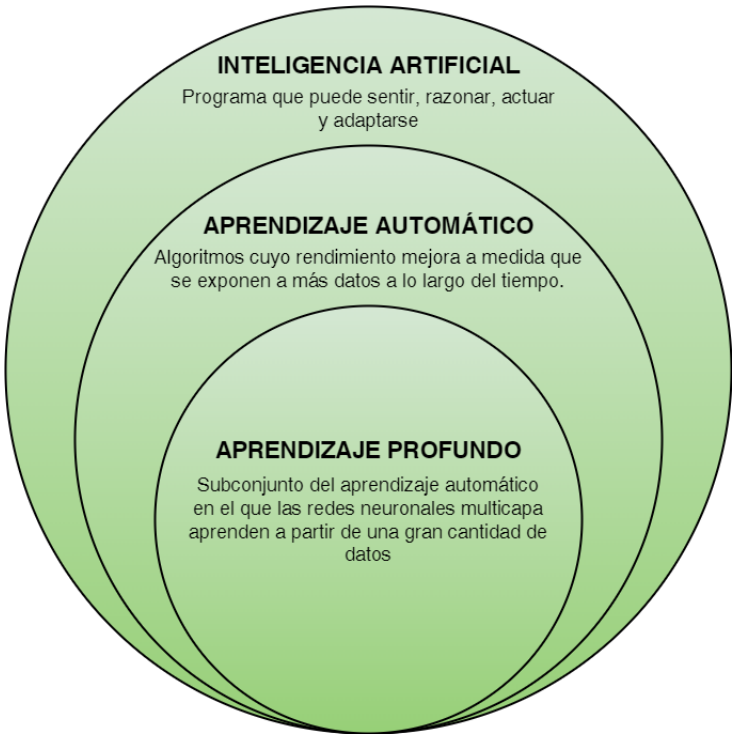


Figura 2.1: Familia del Deep Learning. Imagen extraída y traducida de [Alzubaidi et al., 2021].

la visión por ordenador [Fang et al., 2020], el procesamiento del habla [Palaz et al., 2019], el reconocimiento facial [Li et al., 2020], etc. La estructura de las CNNs se inspira en las neuronas del cerebro humano y animal, de forma similar a una red neuronal convencional. A diferencia de redes totalmente conectadas (FC por sus siglas en inglés) convencionales, los pesos compartidos y las conexiones en la Redes Neuronales Convolucionales se emplean para aprovechar al máximo las estructuras de datos de entrada 2D, como las señales de imagen. Esta operación utiliza un número extremadamente pequeño de parámetros, lo que simplifica el proceso de entrenamiento y acelera el funcionamiento de la CNN.

La arquitectura de las Redes Neuronales Convolucionales se compone de dos bloques principales: bloque de extracción de características y bloque de clasificación. El principal objetivo del primer bloque es aprender y extraer características relevantes de las imágenes de entrada. Algunas de las capas más utilizadas se describen detalladamente a continuación [Alzubaidi et al., 2021].

1. Capa Convolucional: En la arquitectura de las CNN, el componente más significativo es la capa convolucional. Esta capa está compuesta por un conjunto de filtros convolucionales (llamados *kernels*). La imagen de entrada es convolucionada con estos filtros para generar el mapa de características de salida.

2. Capa de Normalización por Lotes (*Batch Normalization*): Esta capa se utiliza para normalizar la salida de una capa anterior, lo que ayuda a acelerar el entrenamiento de la red y mejorar su estabilidad. La normalización por lotes ajusta la media y la varianza de las activaciones de cada neurona en una capa mediante el uso de estadísticas calculadas en lotes de datos.

3. Capa de *Pooling*: La tarea principal de esta capa es reducir el tamaño de los mapas de características. Estos mapas se generan siguiendo las operaciones convolucionales. En otras palabras, este enfoque reduce los mapas de características de gran tamaño para crear mapas de características más pequeños. Al mismo tiempo, mantiene la mayoría de la información dominante (o características) en cada etapa de la capa de *pooling*.

4. Función de Activación: Es un componente fundamental que introduce no linealidad en la red, permitiendo que la red pueda aprender relaciones y patrones más complejos en los datos. Esta función es aplicada a la salida de cada neurona en una capa para determinar si la neurona debe activarse o no, es decir, si debe enviar una señal a las neuronas de la capa siguiente.

   La función de activación toma la suma ponderada de las entradas a la neurona (que incluye las salidas de las neuronas de la capa anterior y los pesos asociados) y, en función de ese valor, decide si la neurona debe activarse. Si el valor resultante supera un cierto umbral, la neurona se activa y envía una señal a las neuronas de la siguiente capa. Algunas de las funciones de activación más usadas son:

   - Sigmoid: La entrada de esta función de activación son números reales, mientras que la salida está restringida entre cero y uno.
   - Tanh (*Tangent Hyperbolic*): Es similar a la función sigmoidea, ya que su entrada son números reales, pero la salida está restringida entre -1 y 1.
   - ReLU (*Rectified Linear Unit*): Es la función más utilizada en el contexto de las Redes Convolucionales. Convierte los valores enteros de la entrada en números positivos. Una carga computacional más baja es el principal beneficio de la función ReLU en comparación con las demás.

5. Capa *Dropout*: Se utiliza como una técnica de regularización para prevenir el sobre-ajuste en la red. Durante el entrenamiento, esta capa aleatoriamente establece a cero un porcentaje de las neuronas en la capa anterior en cada iteración. Esto ayuda a evitar que la red se vuelva demasiado dependiente de ciertas neuronas y, en última instancia, mejora la generalización del modelo.

Después de la extracción de características, el siguiente bloque se compone de capas totalmente conectadas (también conocidas como capas densas). En estas capas, cada neurona

está conectada a todas las neuronas de la capa anterior y toman las características aprendidas en el bloque anterior para realizar la clasificación [Alzubaidi et al., 2021].

En la figura 2.2 se puede observar un ejemplo de la arquitectura de una Red Neuronal Convolucional para la clasificación de dígitos.

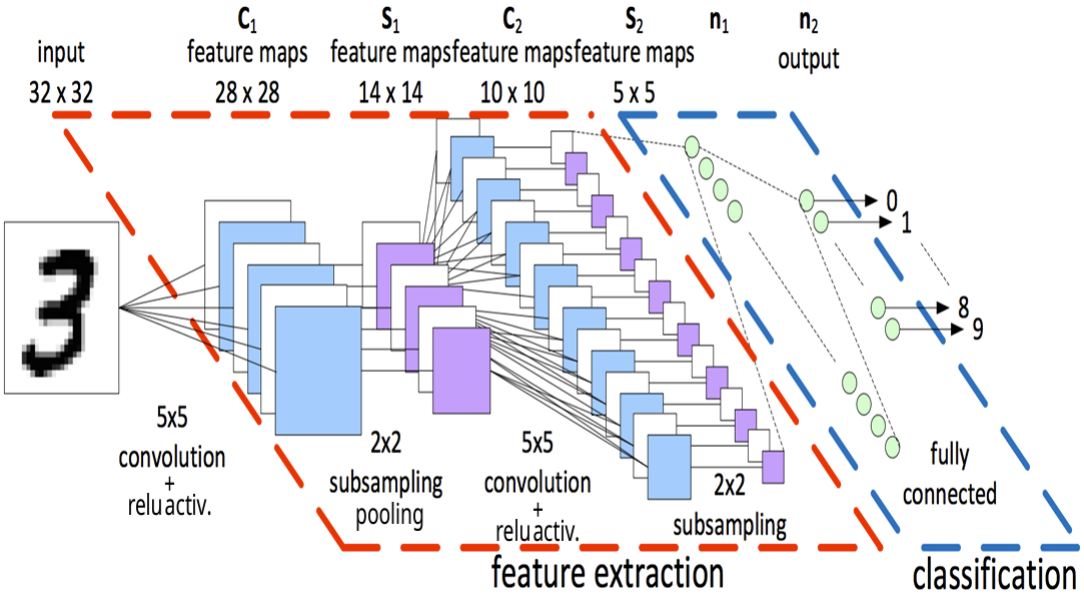


Figura 2.2: Ejemplo de la arquitectura de una CNN.

2.1.2. Mecanismos de Atención

Tal y como se explica en [Alshehri and AlSaeed, 2022], por lo general, cuando se aplican algoritmos de aprendizaje automático en imágenes, se les otorga igual importancia, es decir, misma "atención" a todas las partes de una imagen sin prestar interés especial a las áreas prominentes. Al incorporar Mecanismos de Atención (AMs por sus siglas en inglés, que significa *Attention Mechanisms*) en el aprendizaje automático, se enfatizan las áreas más destacadas de la imagen, aumentando la posibilidad de mejorar el rendimiento de las técnicas aplicadas. Los AMs pueden resaltar las partes más críticas de la información de entrada y, por lo tanto, mejorar su capacidad para extraer la información más relevante para cada parte de la salida, suprimir o incluso ignorar por completo la información irrelevante. Esto resulta en mejoras en la calidad de la salida generada de manera dinámica. Los mecanismos de atención se utilizaron por primera vez en la traducción automática y luego se implementaron en redes neuronales. Su uso se desarrolló rápidamente en el procesamiento de imágenes, la respuesta a preguntas y la traducción automática [Tian et al., 2021]. En [Alshehri and AlSaeed, 2022] se explican de una manera breve lo tres tipos principales de mecanismos de atención.

- Auto-atención (Self-Attention): En el contexto de auto-atención, se evalúa cómo los diferentes elementos de entrada interactúan entre sí. Esto permite que la entrada interaccione con las otras partes y tome decisiones sobre qué aspectos merecen una mayor atención. Una de las ventajas clave de este mecanismo es su habilidad para realizar cálculos en paralelo, lo que resulta especialmente beneficioso cuando se trabaja con entradas grandes. A diferencia de los enfoques suaves y duros, la auto-atención aprovecha esta capacidad de procesamiento paralelo. Este mecanismo utiliza cálculos de matrices simples y fácilmente paralelizables para verificar la atención de todos los elementos de entrada idénticos [de Santana Correia and Colombini, 2022].

- Atención suave (Soft Attention): Se utiliza una serie de elementos para calcular una distribución categórica. Las posibilidades resultantes reflejan la importancia de cada elemento y se utilizan como pesos para generar un codificador consciente del contexto que representa la suma ponderada de todos los elementos [Shen et al., 2018]. Identifica cuánta atención se debe prestar a cada elemento, considerando la interdependencia entre el mecanismo de la red neuronal profunda y el objetivo, mediante la asignación de un peso de 0 a 1 a cada elemento de entrada. Las capas de atención calculan los pesos utilizando funciones *softmax*, lo que hace que el modelo de atención general sea determinista y diferenciable. La atención suave tiene la capacidad de actuar tanto espacial como temporalmente. La función principal del contexto espacial es extraer las características o pesos de las características más esenciales. Ajusta los pesos de todas las muestras en ventanas de tiempo deslizantes para el contexto temporal, ya que las muestras en diferentes períodos contribuyen de manera diferente. Los mecanismos suaves tienen un alto costo de procesamiento a pesar de ser deterministas y diferenciables [de Santana Correia and Colombini, 2022].

- Atención dura (Hard Attention): A partir de la secuencia de entrada, se elige un subconjunto de elementos. Este mecanismo obliga al modelo a centrarse únicamente en los elementos importantes, ignorando todos los demás, donde el peso asignado a una parte de entrada es 0 o 1. Como resultado, el objetivo no es diferenciable ya que los elementos de entrada se observan o no. El procedimiento implica tomar una serie de decisiones sobre qué partes resaltar. Por ejemplo, en el contexto temporal, el modelo presta atención a una parte de la entrada para adquirir información, decidiendo en qué centrarse en el siguiente paso en función de la información conocida. Esto permite que una red neuronal tome decisiones respaldadas por información previamente procesada. Los mecanismos de atención dura se representan mediante procesos estocásticos ya que no hay una verdad absoluta que sugiera la política de selección óptima. En comparación con los mecanismos suaves, el tiempo de inferencia y los costos computacionales se reducen cuando no se almacena ni procesa la entrada completa [de Santana Correia

and Colombini, 2022].

2.1.3. Validación cruzada

[Refaeilzadeh et al., 2009] define la validación cruzada como un método estadístico para evaluar y comparar algoritmos de aprendizaje dividiendo los datos en dos segmentos: uno utilizado para entrenar al modelo y otro utilizado para validarlo. En la validación cruzada convencional, los conjuntos de entrenamiento y validación deben cruzarse en rondas sucesivas para que cada subconjunto de datos tenga una oportunidad de ser validado. La forma más básica de validación cruzada es la validación cruzada con k pliegues.

En esta validación cruzada, los datos se dividen primero en k subconjuntos de igual (o casi igual) tamaño. A continuación, se realizan k iteraciones de entrenamiento y validación, de forma que en cada iteración se reserva un subconjunto diferente de los datos para la validación, mientras que los $k-1$ pliegues restantes se utilizan para el entrenamiento.

Esta técnica permite estimar el error verdadero de predicción de los modelos y el ajuste de los parámetros de los mismos [Berrar, 2019]. Gracias a ello podemos evitar el posible sobreajuste que pueda existir.

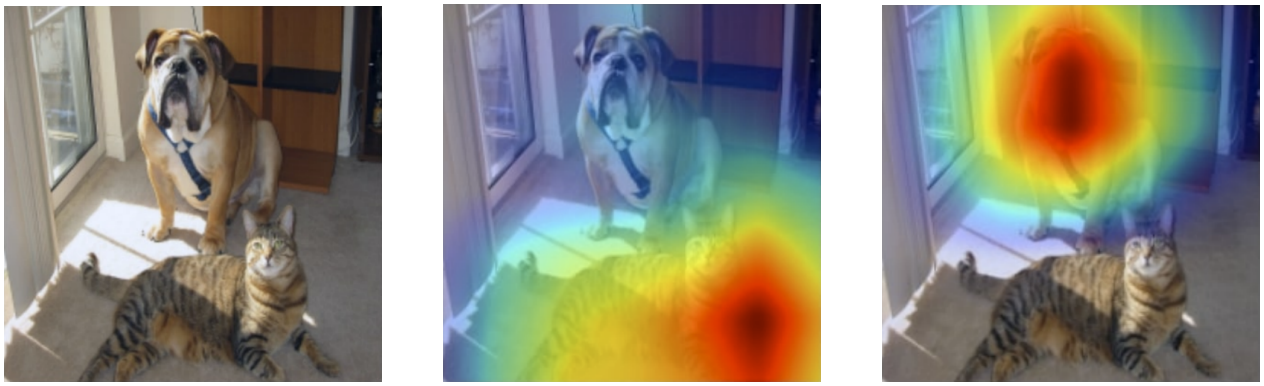
2.1.4. Grad-CAM

Grad-CAM (*Gradient-weighted Class Activation Mapping*) se trata de una técnica, presentada por [Selvaraju et al., 2017], que consiste en una metodología de localización discriminatoria por clases que genera "explicaciones visuales" para redes basadas en CNN, sin requerir modificaciones en la arquitectura ni reentrenamiento. Grad-CAM utiliza los gradientes de una clase objetivo en una red de clasificación o una secuencia de palabras, los cuales se propagan hacia atrás hasta la capa convolucional. Esto produce un mapa de localización en términos generales que resalta las regiones significativas de la imagen que influyen en la predicción de la clase en cuestión. La figura 2.3 muestra un ejemplo incluido en [Selvaraju et al., 2017] donde se aplica Grad-CAM usando la red ResNet sobre una imagen en la que aparece un perro y un gato.

Como hemos podido observar, dependiendo del objetivo de la clasificación (gato o perro), Grad-CAM es capaz de identificar las regiones específicas en las cuales la red neuronal ha concentrado su atención para realizar la predicción.

2.1.5. Métricas de evaluación

En este apartado se detallan diversas métricas comúnmente empleadas para evaluar el desempeño de los modelos en tareas de clasificación. Estas métricas ofrecen una evaluación



(a) Imagen original (b) ResNet Grad-CAM "gato" (c) ResNet Grad-CAM "perro"

Figura 2.3: Grad-CAM: Explicaciones visuales de redes profundas mediante localización basada en gradientes.

cuantitativa de la precisión, cobertura y eficacia en la detección de los modelos. A continuación, se presentan estas métricas, incluyendo su definición y la ecuación correspondiente para su cálculo:

- Matriz de Confusión: Es una herramienta fundamental para evaluar el desempeño de un modelo de clasificación. Divide las predicciones en cuatro categorías:

  - Verdaderos Positivos (TP, *True Positives*): Son los casos en los que el modelo predijo correctamente una instancia como positiva y, de hecho, la instancia es positiva según la etiqueta real. En el contexto de detección de cáncer de mama, sería un caso en el que el modelo identifica correctamente a un paciente con cáncer.
  - Verdaderos Negativos (TN, *True Negatives*): Son los casos en los que el modelo predijo correctamente una instancia como negativa y, de hecho, la instancia es negativa según la etiqueta real. En el contexto de detección de cáncer de mama, sería un caso en el que el modelo identifica correctamente a un paciente sano.
  - Falsos Positivos (FP, *False Positives*): Son los casos en los que el modelo predijo incorrectamente una instancia como positiva, pero en realidad la instancia es negativa según la etiqueta real. En el contexto de detección de cáncer de mama, sería un caso en el que el modelo predice que un paciente tiene cáncer cuando en realidad no lo tiene.
  - Falsos Negativos (FN, *False Negatives*): Son los casos en los que el modelo predijo incorrectamente una instancia como negativa, pero en realidad la instancia es positiva según la etiqueta real. En el contexto de detección de cáncer de mama, sería un caso en el que el modelo no detecta el cáncer en un paciente que realmente lo tiene.

- Exactitud (*Accuracy*): Mide la proporción de predicciones correctas realizadas por el modelo en relación con el total de predicciones. Se calcula dividiendo la suma de los verdaderos positivos y verdaderos negativos entre el total de predicciones.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.1}$$

- Especificidad (*Specificity*): Mide la capacidad del modelo para identificar correctamente ejemplos negativos. Se calcula dividiendo los verdaderos negativos entre la suma de los verdaderos negativos y los falsos positivos.

$$specificity = \frac{TN}{TN + FP} \tag{2.2}$$

- Precisión (*Precision*): Representa la proporción de predicciones positivas realizadas correctamente por el modelo en relación con todas las predicciones positivas. Se calcula dividiendo los verdaderos positivos entre la suma de los verdaderos positivos y los falsos positivos.

$$precision = \frac{TP}{TP + FP} \tag{2.3}$$

- Exhaustividad (*Recall*): Mide la capacidad del modelo para identificar correctamente ejemplos positivos. Se calcula dividiendo los verdaderos positivos entre la suma de los verdaderos positivos y los falsos negativos.

$$recall = \frac{TP}{TP + FN} \tag{2.4}$$

- Puntuación F1 (*F1-score*): Combina la precisión y la exhaustividad en un solo valor que refleja el equilibrio entre ambas métricas.

$$F1\text{-}score = 2 \cdot \frac{Precision \cdot Recall}{(Precision + Recall)} \tag{2.5}$$

- Curva ROC y Área bajo la Curva (AUC): La curva ROC es una representación gráfica que muestra la tasa de verdaderos positivos frente a la tasa de falsos positivos para diferentes umbrales de decisión. El área bajo la curva (AUC) mide la capacidad discriminatoria del modelo; un AUC más alto indica un mejor rendimiento de clasificación.

2.2. Trabajos relacionados

Son muchos los estudios que existen actualmente acerca del diagnóstico y clasificación del cáncer de mama mediante el uso de imágenes. En esta sección haremos un recorrido sobre los más importantes haciendo especial hincapié en aquellos que se centran en la clasificación basada en imágenes térmicas.

Para una mejor lectura se va a dividir la sección en diferentes subsecciones dependiendo de la modalidad de imagen usada en los estudios. Son 3 los tipos de imágenes en los que se basan: histopatológicas, mamográficas y termográficas.

2.2.1. Imágenes histopatológicas

Las imágenes histopatológicas son representaciones visuales de tejidos biológicos tomadas mediante microscopía. Estas imágenes permiten examinar las estructuras celulares y tisulares con alta resolución [Gurcan et al., 2009]. En el contexto del cáncer de mama, las imágenes histopatológicas muestran secciones de tejido mamario obtenidas mediante biopsias o resecciones quirúrgicas. Los patólogos estudian estas imágenes para identificar características morfológicas y detectar anomalías, como células cancerosas, cambios precancerosos y otros signos de enfermedad. A continuación se presentan estudios relacionados con este tipo de imágenes.

En [Nahid et al., 2018], los autores presentan tres modelos diferentes. El primero emplea técnicas de redes neuronales convolucionales (CNN), el segundo utiliza la estructura de memoria a largo plazo (LSTM, por sus siglas en inglés), y el tercer modelo combina las estructuras CNN y LSTM para el análisis de los datos. Para llevar a cabo los experimentos, utilizaron el conjunto de datos de imágenes mamarias BreakHis. Dicho conjunto de datos consta de cuatro grupos de imágenes según el factor de ampliación: 40x, 100x, 200x y 400x. Cada una de las imágenes en este conjunto de datos es en formato RGB y tiene un tamaño de 760 $\times$ 460 píxeles. El modelo que combinó CNN y LSTM obtuvo el mejor resultado al alcanzar un 91 % de precisión en el conjunto de datos con un factor de ampliación de 200x.

En [Fan et al., 2020], los autores propusieron un modelo de imágenes de granularidad fina basado en CNN (Resnet-18) para extraer características, al que aplican un mecanismo de atención para localizar objetos en el conjunto datos *BreAst Cancer Histology* (BACH). Este conjunto de datos consta de 400 imágenes de alta resolución, 2018 $\times$ 1356 píxeles. Obtuvieron una tasa de clasificación del 97 %.

Finalmente, en el artículo [Zhang et al., 2019], los autores presentaron un modelo basado en ResNet utilizando el conjunto de datos BreakHis, mencionado previamente. Al modelo propuesto, le incorporaron un módulo de atención convolucional en bloque (CBAM). Los resultados exhibieron una mejora significativa en comparación con los modelos de referencia,

logrando una precisión del 92.6 %, una sensibilidad del 94.7 %, una especificidad del 88.9 %, un puntaje F1 del 94.1 % y un área bajo la curva (AUC) del 91.8 % para el conjunto de datos con un factor de ampliación de 200x.

2.2.2. Imágenes mamográficas

Las imágenes mamográficas son radiografías de las mamas. Estas imágenes pueden revelar la presencia de masas, microcalcificaciones u otras anomalías que podrían indicar la presencia de tumores malignos o benignos [Centers for Disease Control and Prevention, 2023]. A continuación se describen trabajos relacionados con este tipo de imágenes.

En [Rashed and El Seoud, 2019], los autores propusieron una novedosa arquitectura de CNN con el objetivo de tratar el problema de la variación de las anomalías en la mamografía digital. La red está inspirada en la estructura U-net y el conjunto de datos usado fue *Curated Breast Imaging Subset of Digital Database of Screening Mammography* (CBIS-DDSM). La tasa de clasificación para microcalcificaciones y masas fue del 94,31 % y 95,01 %, respectivamente.

Los autores de [Patil and Biradar, 2021] realizaron una combinación optimizada de redes neuronales convolucionales (CNN) y recurrentes (RNN), para la detección automatizada del cáncer de mama en mamografías. El resultado de este estudio muestra que la combinación de los dos clasificadores tiende a proporcionar una precisión diagnóstica global superior a la de los modelos convencionales, con una precisión del 90.59 %.

Finalmente, en [Deng et al., 2020] los autores mejoraron las CNNs mediante la integración de un innovador mecanismo *SE-Attention* para aprender características discriminatorias, con el objetivo de clasificar automáticamente la densidad mamaria en mamografías. El modelo fue entrenado usando 18 157 imágenes de mamografías, segmentadas manualmente en 4 niveles basados en el *Breast Imaging and Reporting Data System* (BI-RADS): A (graso), B (fibroglandular), C (heterogéneamente denso) y D (extremadamente denso). El modelo superó a otros estudios con una precisión del 92.17 %.

2.2.3. Imágenes termográficas

Las imágenes termográficas capturan las emisiones de calor de una región corporal específica. En el contexto del cáncer de mama, éstas se utilizan para detectar diferencias en la temperatura superficial de las mamas. Las células cancerosas y los tejidos anormales pueden generar más calor que los tejidos normales debido al aumento del flujo sanguíneo y el metabolismo [Rakhunde et al., 2022]. A continuación se describen trabajos relacionados con este tipo de imágenes. A continuación se exponen investigaciones que se centran en imágenes de este tipo.

En [Santana et al., 2018] se implementaron clasificadores basados en redes neuronales artificiales, árboles de decisión, clasificadores bayesianos y atributos de Haralick y Zernike. El conjunto de datos usado está compuesto por imágenes termográficas adquiridas en el Hospital Universitario de la Universidad Federal de Pernambuco. Utilizaron varios modelos como una red Bayesiana, *Naive Bayes*, *Support Vector Machine*, árboles de decisión, Perceptrón Multicapa (MLP), *Random Forest* (RF), *Extreme Learning Machine* (ELM), etc. Usando el 75 % de los datos para entrenar los modelos, los resultados mostraron que MLP y ELM dieron resultados prometedores en comparación con los otros clasificadores, con una tasa de clasificación del 73.38 %. Estos resultados aumentaron al 76,01 % de exactitud al utilizar el método de validación cruzada de 10 *folds* para realizar las pruebas.

Por otro lado, en [Ekici and Jawzal, 2020], los autores propusieron un nuevo algoritmo para la extracción de los rasgos característicos de la mamas basado en bio-datos, análisis de imagen y estadísticas de la imagen. Estas características se han extraído del conjunto de datos DMI, que contiene imágenes termográficas capturadas por una cámara térmica. Estos datos se utilizaron para clasificar las imágenes de las mamas como normales o sospechosas mediante el uso de redes neuronales convolucionales optimizadas por el algoritmo de Bayes. Se obtuvo una tasa de clasificación del 98.95 % para las imágenes térmicas del conjunto de datos pertenecientes a 140 individuos.

En [Tello-Mijares et al., 2019], los autores presentan un método eficaz y eficiente para segmentar imágenes termográficas de mamas y diagnosticar el cáncer, clasificándolo como sano o enfermo. Las contribuciones principales incluyen el innovador uso de la función de curvatura combinada k (cvt k) y el método de flujo vectorial gradiente (GVF) para la segmentación de la mama. Además, para el análisis y la clasificación de las imágenes termográficas segmentadas, proponen el empleo de una red neuronal convolucional. Utilizaron 63 imágenes del conjunto de datos DMR-IR y, mediante un enfoque de validación cruzada de 2 pliegues, lograron alcanzar, como mejor resultado, un 100 % de exactitud, sensibilidad y especificidad. Es importante destacar que en este caso, la evaluación se llevó a cabo en un conjunto de datos bastante reducido, y lo que es aún más relevante, ellos mismos generaron el *ground truth* en lugar de utilizar el proporcionado en el conjunto de datos original. Para generar este *ground truth*, se apoyaron en la opinión de dos oncólogos que delimitaron la región de interés en las imágenes mamarias y realizaron medidas en función de la cual se estableció el etiquetado correcto. Por lo tanto, si la segmentación automática obtenida mediante el método GVF *snakes* se asemejaba a la realizada por los oncólogos, la clasificación resultante solía ser coincidente.

En el estudio realizado por [Sánchez-Cauce et al., 2021], los autores presentaron una novedosa red neuronal convolucional de múltiples entradas para la detección del cáncer de mama, la cual combina imágenes térmicas capturadas desde diferentes ángulos de visión junto con información personal y clínica de los pacientes. Aplicaron este enfoque a la base de datos

DMR-IR, en la cual el mejor modelo alcanzó una precisión del 97 %, un área bajo la curva ROC de 0.99, una especificidad del 100 % y una sensibilidad del 83 %.

En el trabajo [de Freitas Oliveira Baffa and Grassano Lattari, 2018], los investigadores entrenaron redes neuronales convolucionales para clasificar imágenes térmicas a través de protocolos estáticos y dinámicos usando la base de datos DMR-IR. El enfoque propuesto demostró obtener resultados competitivos en ambos protocolos. En el protocolo estático, se logró una precisión del 98 % en imágenes a color y del 95 % en escala de grises, mientras que en el protocolo dinámico se obtuvo un 95 % en imágenes a color y un 92 % en escala de grises. Estos resultados superaron a otros métodos aplicados al mismo conjunto de datos.

Finalmente, en [Alshehri and AlSaeed, 2022], los autores se propusieron investigar el potencial de las redes neuronales convolucionales con mecanismos de atención para lograr resultados satisfactorios en la detección del cáncer de mama utilizando imágenes térmicas. Presentaron diversos modelos que fueron entrenados y evaluados empleando la base de datos DMR-IR. Los mecanismos de atención aplicados a los modelos CNN alcanzaron resultados alentadores en las pruebas, con una tasa de clasificación del 99.46 %, 99.37 % y 99.30 %. En contraste, las CNN sin mecanismos de atención obtuvieron una exactitud del 92.32 %, lo que implica que la incorporación de mecanismos de atención mejoró la precisión en un 7 %. Además, los modelos propuestos superaron a los modelos previamente revisados en la literatura.

Capítulo 3

Métodos

En este capítulo, presentaremos en detalle los métodos y enfoques empleados en nuestro estudio para la detección del cáncer de mama a partir de imágenes térmicas. Abordaremos tanto el conjunto de datos utilizado, que servirá como base para entrenar y validar nuestros modelos, como las arquitecturas de redes neuronales convolucionales (CNN) implementadas junto a los mecanismos de atención empleados.

3.1. Conjunto de datos

El conjunto de datos que se ha usado en este trabajo es la Base de Datos para la Investigación Mastológica con Imágenes Infrarrojas. Contiene información de exámenes mamarios y datos clínicos de 280 pacientes voluntarias del Hospital Universitário Antônio Pedro (HUAP) de la Universidad Federal Fluminense (Brasil).

Las imágenes infrarrojas que componen esta base de datos han sido obtenidas usando dos protocolos, uno estático y otro dinámico. El protocolo estático consiste en 5 imágenes tomadas desde diferentes ángulos: 1 frontal, 2 laterales a 45º (lado derecho e izquierdo) y 2 laterales a 90º (lado derecho e izquierdo). El protocolo dinámico consiste en una serie de imágenes tomadas cada 15 segundos durante 5 minutos o hasta que la temperatura original del cuerpo es alcanzada, en las que previamente se ha enfriado la zona de las mamas. Finalmente se realizan 2 capturas más, una de la mama izquierda y otra de la mama derecha, ambas a 90º.

Cada imagen de cada paciente tiene un tamaño de 640x480 píxeles con un solo canal de color, es decir, las imágenes son en blanco y negro. Para obtener más información acerca de la descripción, descarga y limpieza de este conjunto de datos es conveniente la lectura del Apéndice A.

Después de la limpieza, el número de pacientes del que disponemos finalmente es de 260, donde 166 son pacientes sanas y 94 enfermas. Cabe destacar que para los experimentos realizados solo se ha tenido en cuenta una imagen por paciente, de tipo frontal obtenida del

protocolo estático. Además, ya que la región de interés para entrenar a nuestros modelos es la zona de las mamas, las imágenes han sido recortadas de forma manual usando la herramienta *Label Studio* [HumanSignal, 2023]. De manera resumida, *Label Studio* es una plataforma de código abierto desarrollada por HumanSignal que se utiliza para etiquetar y anotar datos de manera colaborativa. Esta herramienta es especialmente valiosa en el ámbito del aprendizaje automático y la inteligencia artificial, donde la disponibilidad de datos etiquetados con precisión es crucial para entrenar y mejorar los modelos. Aunque en nuestro caso, simplemente se ha usado para delimitar la región de interés y su posterior recorte.

El criterio empleado para delimitar las imágenes ha consistido principalmente en enfocarse en la región de interés que abarca la zona de las mamas, extendiéndose aproximadamente hasta la altura de las axilas. En la figura 3.1 se muestra un ejemplo de la herramienta *Label Studio* donde en la parte izquierda se muestra el listado de imágenes a recortar y en la parte derecha se ilustra un ejemplo de la región de interés marcada en color verde sobre una de las imágenes del listado. Por tanto, un total de 260 imágenes han sido recortadas. Algunos ejemplos de éstas pueden ser consultadas en la figura 3.2.

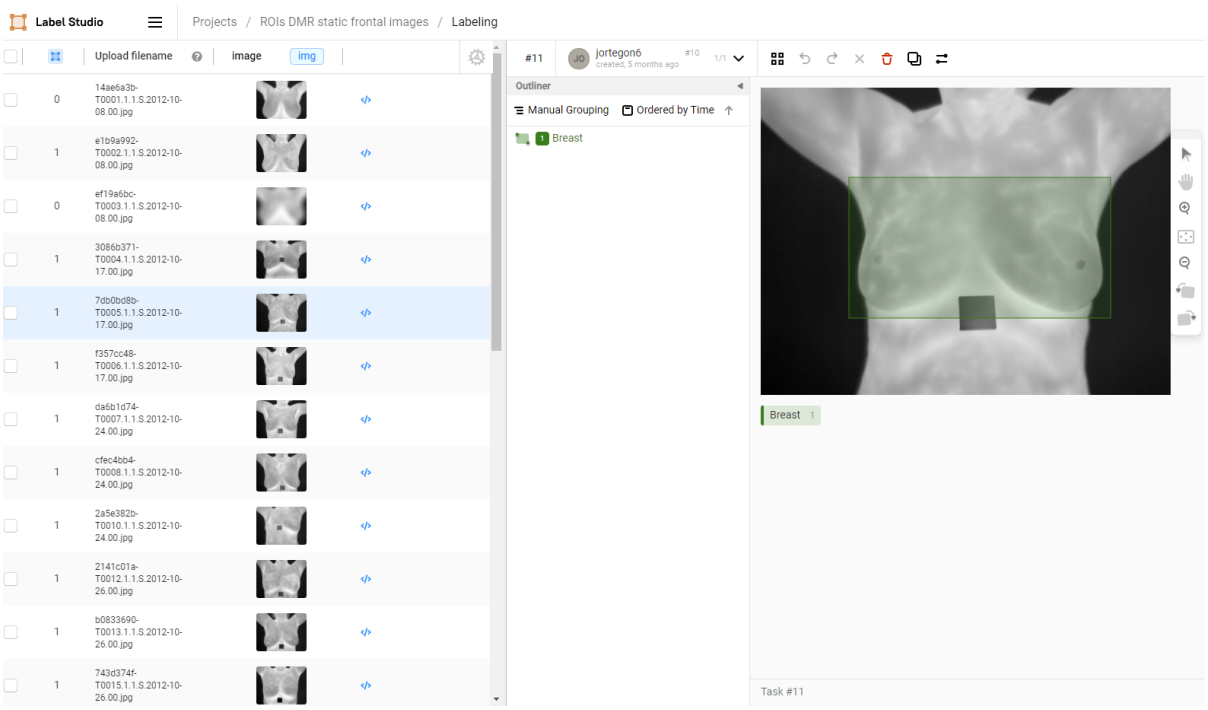
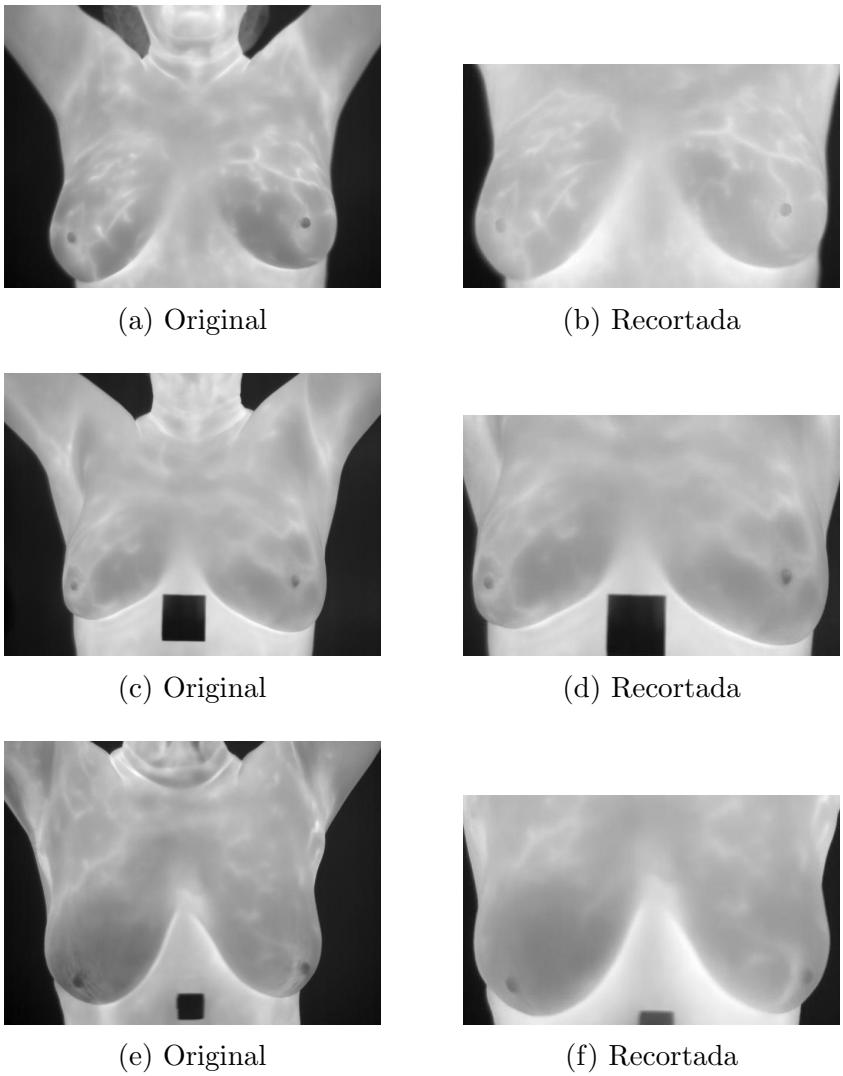


Figura 3.1: Ejemplo de *Label Studio* para marcar la región de interés de las imágenes.

Hemos optado por realizar una partición en relación 80/20 para la creación de los conjuntos de entrenamiento y prueba. Esto significa que el 80 % de las imágenes se asignarán al conjunto que se usará para entrenar a los modelos, mientras que el 20 % restante se destinará al conjunto de prueba. En el proceso de creación de estos conjuntos, hemos tenido especial cuidado en asegurarnos de que la distribución de clases en ambos conjuntos refleje fielmente



(a) Original (b) Recortada

(c) Original (d) Recortada

(e) Original (f) Recortada

Figura 3.2: Ejemplo de algunas imágenes originales junto al recorte.

la distribución del conjunto de datos original. Esta consideración es especialmente importante debido al desequilibrio de clases presente en el *dataset* de origen.

3.1.1. Aumento de datos

Uno de los inconvenientes que encontramos en el conjunto de datos de entrenamiento es su reducida cantidad de imágenes para el entrenamiento de una red neuronal profunda, lo que es muy posible que limite su capacidad para generalizar y obtener resultados sólidos en la detección del cáncer de mama. Sin embargo, es en este punto donde entra en juego una técnica crucial en el ámbito del Deep Learning: el aumento de datos.

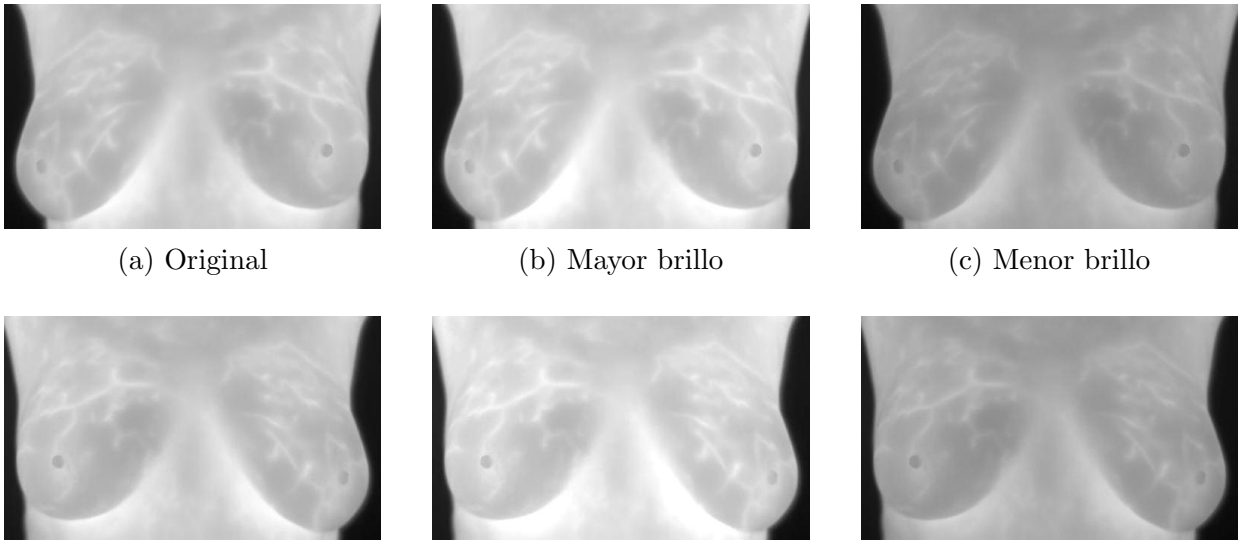
El aumento de datos, también conocido como "data augmentation" en inglés, es una estrategia que busca abordar la escasez de datos al generar variaciones artificiales de las muestras existentes. Esta técnica consiste en aplicar transformaciones aleatorias y controladas

a las imágenes originales de manera que las que son generadas sean realistas y preserven las características esenciales.

En el contexto de imágenes térmicas para la detección del cáncer de mama, el aumento de datos implica la aplicación de transformaciones como volteo horizontal y alteraciones en el brillo a las imágenes originales. Estas modificaciones no solo aumentan el tamaño del conjunto de datos sino que aumentan la robustez del sistema a pequeños giros y cambios de iluminación.

Por otra parte, se han considerado otras técnicas de aumento de datos, como la rotación y la distorsión óptica. Sin embargo, al aplicar estas técnicas a nuestro conjunto de datos, los resultados obtenidos fueron peores, lo que nos ha llevado a descartar su uso. Es posible que las imágenes generadas con estas dos técnicas hubieran perdido información relevante para el entrenamiento del modelo.

El procedimiento que se ha seguido ha sido generar, por cada imagen del conjunto de entrenamiento, 10 imágenes aumentadas aplicando con un 50 % de probabilidad el volteo horizontal y de forma aleatoria una pequeña alteración del brillo. En la figura 3.3 se puede apreciar algunas de las imágenes generadas aplicando el aumento de datos descrito sobre una de la imágenes originales del conjunto de datos de entrenamiento.



(a) Original (b) Mayor brillo (c) Menor brillo

(d) Volteo horizontal (e) Volteo horizontal y mayor brillo (f) Volteo horizontal y menor brillo

Figura 3.3: Ejemplo de aumento de datos sobre una de las imágenes del *dataset*.

3.2. Modelos propuestos

En esta sección, se detallará la implementación de los diversos modelos propuestos para la detección del cáncer de mama mediante imágenes termográficas. En total, se presentarán

cuatro modelos. Comenzaremos con un modelo simple, basado en una Red Convolucional, que servirá como modelo base. Luego, a partir de este modelo base, implementaremos tres enfoques adicionales que incorporarán los mecanismos de atención descritos en la sección 2.1.2: auto-atención, atención suave y atención dura. La figura 3.4 muestra la arquitectura que van a seguir los últimos 3 modelos mencionados. El modelo base no incluye la capa LSTM bidireccional ni la capa que aplica el mecanismo de atención, como es lógico.

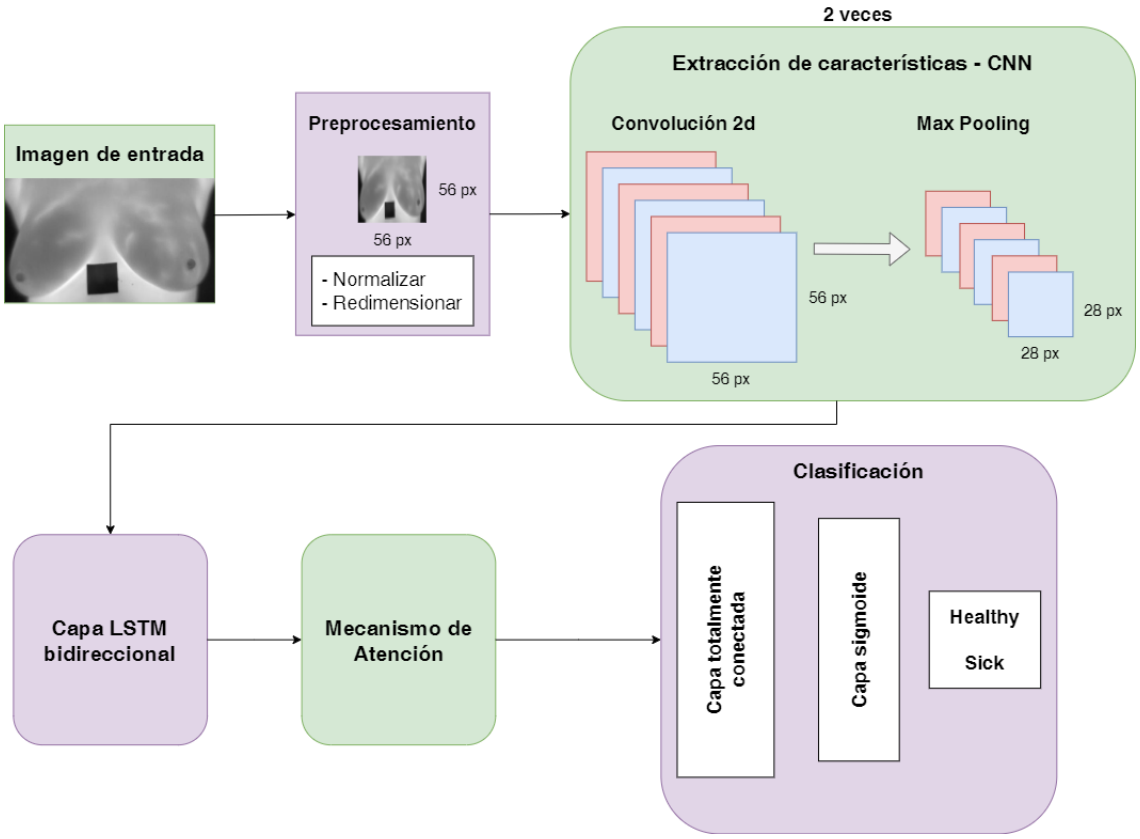


Figura 3.4: Arquitectura de los modelos implementados.

En las próximas subsecciones se va a describir cada una de las etapas que constituyen la arquitectura vista en la figura 3.4. Al final del capítulo se muestra la arquitectura final junto a sus parámetros de cada uno de los modelos implementados.

3.2.1. Preprocesamiento

En la primera etapa utilizamos las imágenes que constituyen el conjuntos de datos que fueron previamente recortadas, tal y como se ha explicado en el capítulo 3.1. Antes de que la CNN reciba las imágenes, éstas también han sido normalizadas y redimensionadas a un tamaño de 56x56 píxeles, que es una resolución que a menudo se utiliza en las redes neuronales convolucionales [de Freitas Oliveira Baffa and Grassano Lattari, 2018].

3.2.2. Extracción de características

La extracción de patrones la logramos gracias a las redes neuronales convolucionales. Como mencionamos en la sección 2.1.1, estas redes son extremadamente útiles en el Aprendizaje Profundo y desempeñarán un papel fundamental en la extracción de características en nuestros modelos. La arquitectura de esta CNN se compone de dos bloques de capas apiladas en serie, que se explican a continuación:

- Capa de Convolución: Esta es la primera capa convolucional de la red. Toma una imagen en escala de grises (1 canal) y aplica, en el primer bloque, 64 filtros de convolución y en el segundo 128 filtros, ambos de tamaño 3x3. Esto ayuda a extraer características específicas de la imagen.

- Capa de Normalización por Lotes (*Batch Normalization*): Después de cada capa de convolución, se aplica normalización por lotes para estandarizar las activaciones de las neuronas. Esto ayuda a estabilizar y acelerar el proceso de entrenamiento.

- Función de Activación ReLU: Se aplica la función ReLU (Rectified Linear Unit) a las activaciones convolucionales normalizadas, permitiendo la detección de patrones más complejos en los datos.

- Capa de *Pooling*: Esta capa reduce el tamaño espacial de las características al realizar un muestreo máximo en regiones de 2x2. Esto reduce la cantidad de parámetros y hace que la red sea más eficiente.

- Capa de *Dropout*: Desactiva aleatoriamente una fracción de las unidades en la capa anterior durante el entrenamiento. Esto ayuda a prevenir el sobreajuste (*overfitting*) y a mejorar la generalización del modelo.

3.2.3. Capa *Long Short-Term Memory* bidireccional

Cabe destacar que las imágenes presentan características espaciales propias. Con el objetivo de explotar al máximo la información a largo plazo contenida en estas características, se va a emplear una capa LSTM (*Long Short-Term Memory*) para aprender de ellas. Con la finalidad de utilizar tanto el contexto previo como el futuro de una secuencia en el proceso de clasificación, hemos optado por implementar una capa LSTM bidireccional (BLSTM), lo que permite extraer características espaciales tanto en dirección ascendente como descendente, enriqueciendo así la comprensión de los patrones en el análisis [Liu et al., 2017].

3.2.4. Capa de atención

Basándonos en la salida de la capa LSTM, aplicamos de forma independiente los diferentes mecanismos de atención descritos en la sección 2.1.2: auto-atención, atención suave y atención dura.

- Auto-atención [Vaswani et al., 2017]: Se ha implementado usando MultiheadAttention [Pytorch, 2023] que permite que el modelo atienda conjuntamente a la información de diferentes subespacios de representación.

- Atención suave [Alshehri and AlSaeed, 2022]: Ignora áreas irrelevantes multiplicando el mapa de características correspondiente por un peso bajo. En consecuencia, una zona de alta atención mantiene su valor original, mientras que las áreas de baja atención se acercan a 0. Utilizamos el estado oculto $C = h_t - 1$ de la salida de la capa LSTM para calcular un peso α_i para cada subparte de una imagen[1]. Calculamos una puntuación s_i para medir cuánta atención se otorga, como se muestra en la siguiente ecuación:

$$s_i = tanh(W_c C + W_i X_i) = tanh(W_c h_{c-1} + W_x x_i) \tag{3.1}$$

  Pasamos s_i a una función *softmax* para normalizar y calcular el peso α_i.

$$\alpha_i = softmax(s_1, s_2, ...s_i) \tag{3.2}$$

  Con *softmax*, α_i suma 1, y lo utilizamos para calcular una media ponderada para x_i.

$$Z = \sum_i \alpha_i x_i \tag{3.3}$$

- Atención dura [Alshehri and AlSaeed, 2022]: Obliga al modelo a enfocarse únicamente en los elementos importantes, ignorando todos los demás, donde el peso asignado a una parte de entrada es o 0 o 1. Como resultado, el objetivo no es diferenciable ya que los elementos de entrada están presentes o no lo están. Para calcular la atención dura, utilizamos el valor α_i como una tasa de muestreo para elegir uno de los x_i como entrada a la capa siguiente, en lugar de un promedio ponderado como en la atención suave.

$$Z \sim \alpha_i, x_i \tag{3.4}$$

[1]Cuando hablamos de "subparte de una imagen" nos referimos a regiones específicas o partes de la imagen que se consideran y evalúan individualmente en función de su relevancia para la tarea de análisis o clasificación. Como se ha indicado, estas áreas específicas se determinan multiplicando el mapa de características por un peso de bajo valor. Este peso es un parámetro que se ajusta durante el proceso de entrenamiento de la red neuronal. A medida que el modelo se entrena, el valor de este peso se adapta de manera iterativa para minimizar la función de pérdida del modelo, mejorando así su capacidad de atención y capacidad de clasificación. Después de esta multiplicación, los valores más altos se consideran áreas de mayor relevancia.

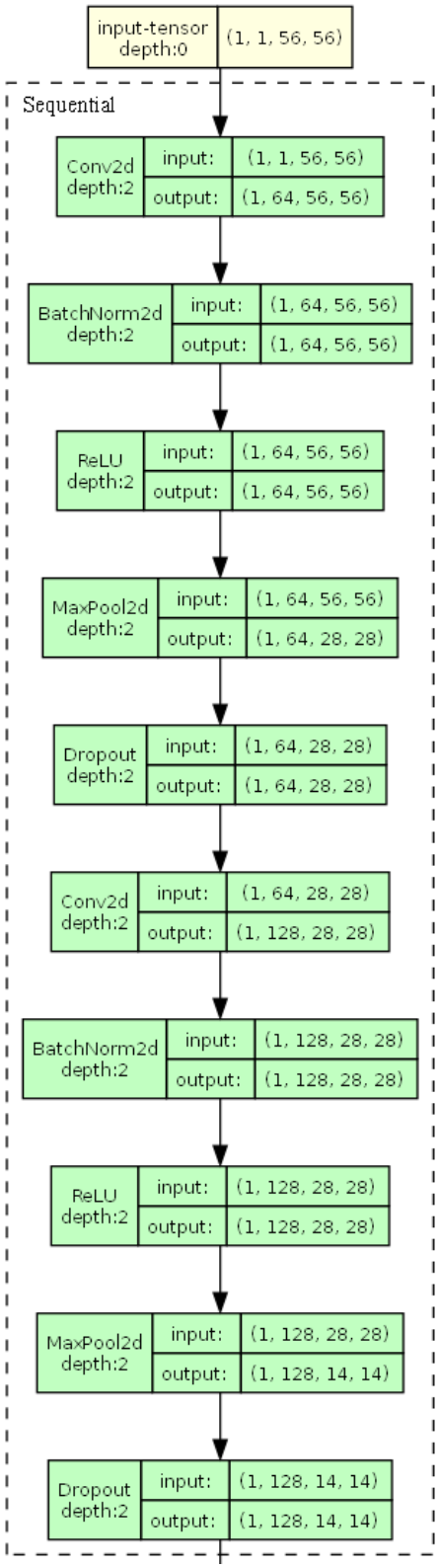
3.2.5. Capa totalmente conectada

Se encarga de combinar y transformar las características extraídas de las capas anteriores para realizar tareas específicas, en este caso, clasificación.

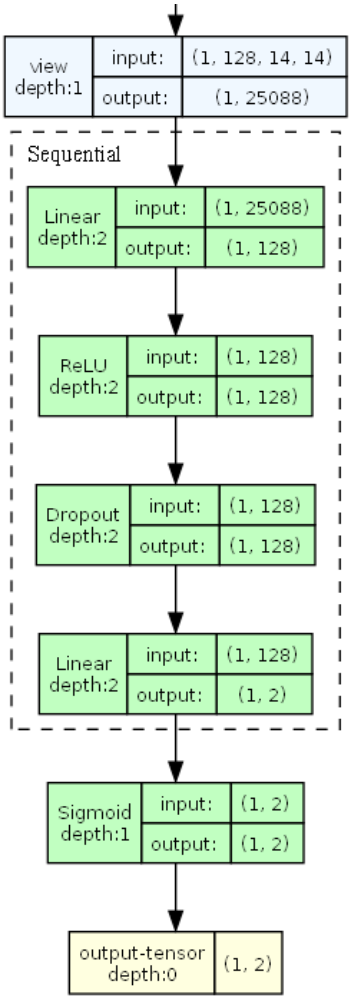
3.2.6. Capa sigmoide

Última capa de la red. Toma la salida de la capa totalmente conectada y aplica la función sigmoide, que comprime los valores en el rango $[0, 1]$.

Las figuras 3.5, 3.6, 3.7 y 3.8 presentan las arquitecturas finales y sus parámetros correspondientes de los modelos CNN, CNN + Auto-atención, CNN + Atención suave y CNN + Atención dura. Estos modelos son los que han sido entrenados finalmente utilizando el conjunto de entrenamiento y evaluados utilizando el conjunto de prueba.

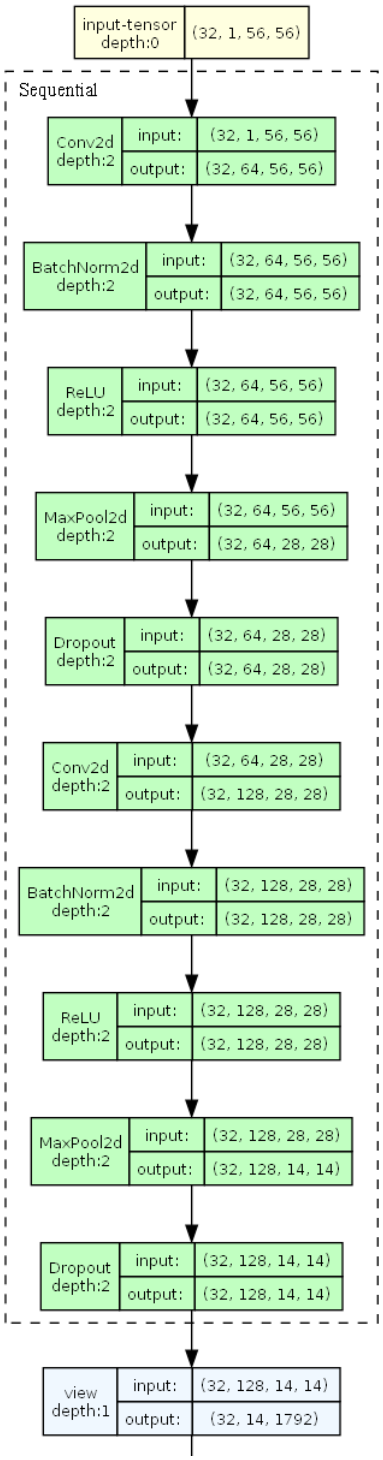


(a) CNN - Parte 1

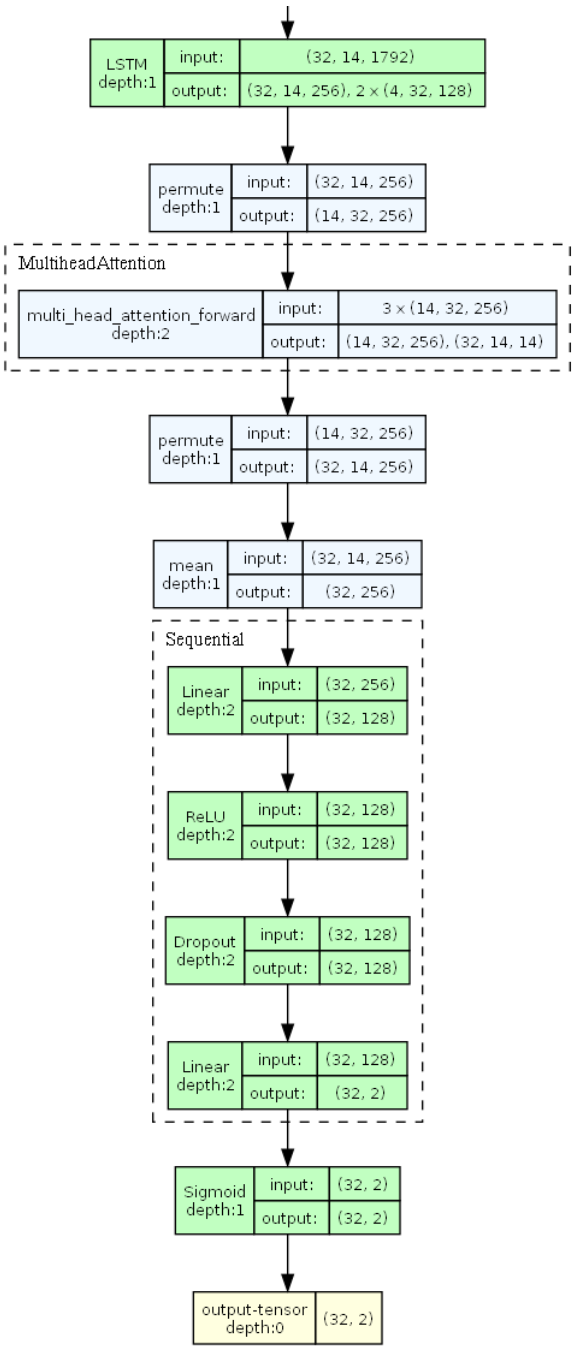


(b) CNN - Parte 2

Figura 3.5: Arquitectura y parámetros de la red CNN.

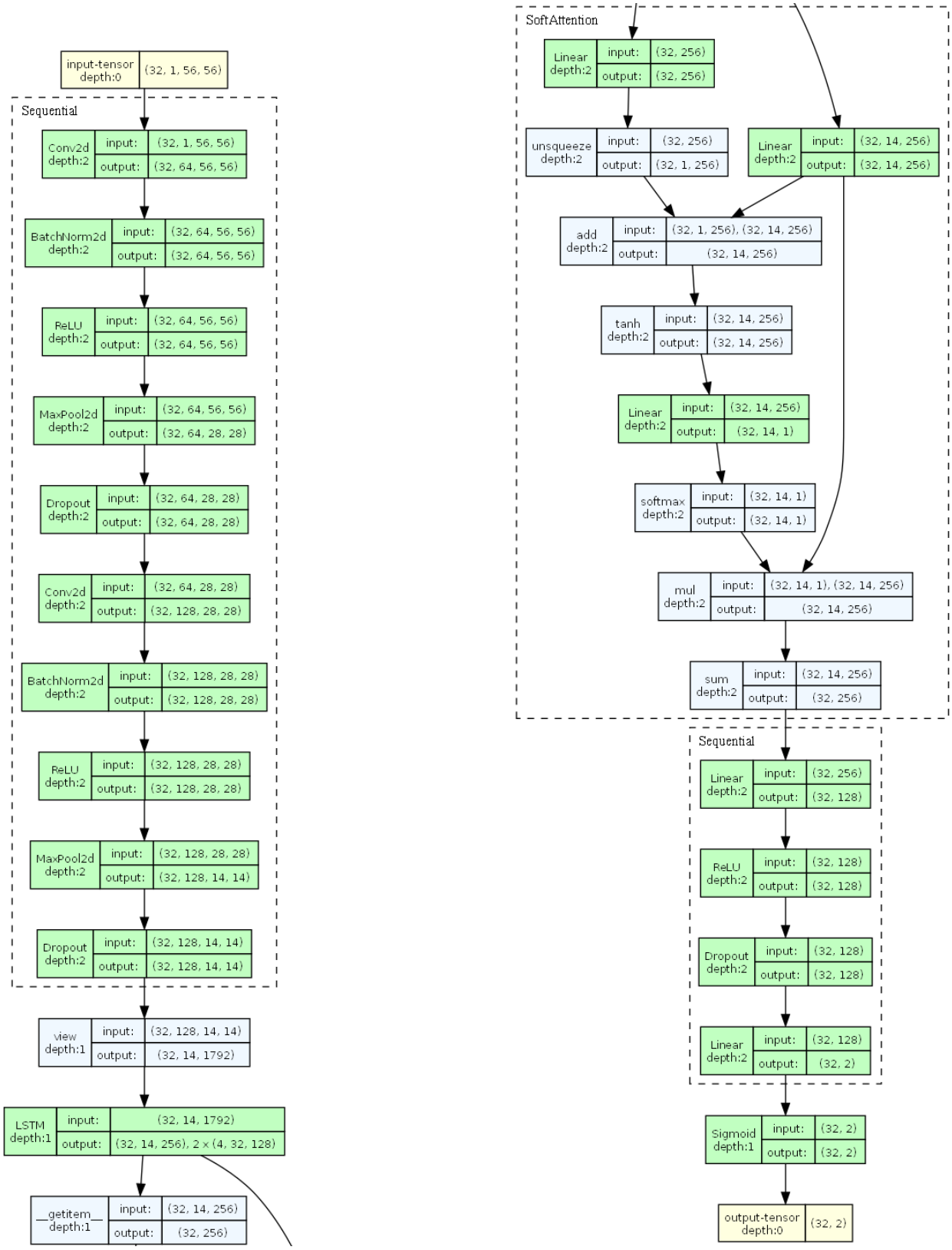


(a) CNN + Auto-atención - Parte 1



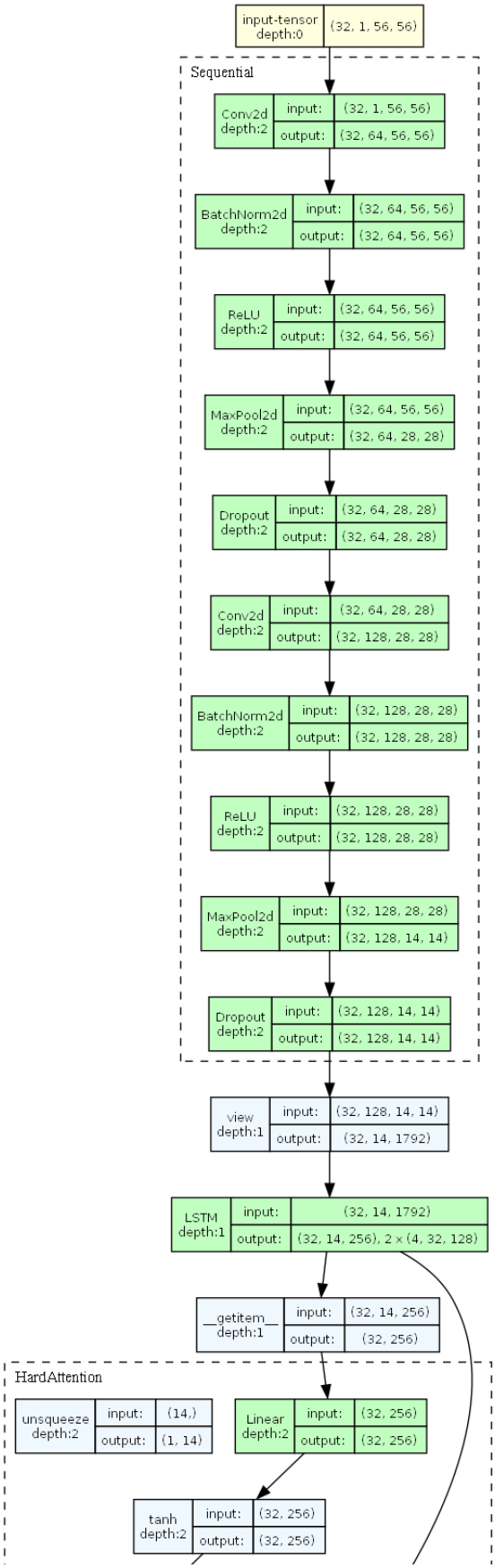
(b) CNN + Auto-atención - Parte 2

Figura 3.6: Arquitectura y parámetros de la red CNN + Auto-atención.



(a) CNN + Atención suave - Parte 1 (b) CNN + Atención suave - Parte 2

Figura 3.7: Arquitectura y parámetros de la red CNN + Atención suave.



(a) CNN + Atención dura - Parte 1 (b) CNN + Atención dura - Parte 2

Figura 3.8: Arquitectura y parámetros de la red CNN + Atención dura.

Capítulo 4

Experimentos y resultados

A continuación se presentan los experimentos que se han realizado junto con los resultados obtenidos. Finalmente, se dedica una sección de discusión de nuestros resultados en comparación con los obtenidos por otros autores.

4.1. Experimentos

A lo largo de esta sección, exploraremos los diversos experimentos llevados a cabo en el contexto de este trabajo. En primer lugar, abordaremos la validación cruzada, una metodología que nos ha permitido evaluar el rendimiento de los modelos y, al mismo tiempo, llevar a cabo la selección de hiperparámetros. Además, exploraremos la aplicación de la técnica Grad-CAM, una herramienta que permite visualizar las áreas en las que los modelos con mecanismos de atención han centrado su enfoque y atención durante la detección.

4.1.1. Validación cruzada

Con el objetivo de evaluar el desempeño de los modelos propuestos y ajustar algunos de los parámetros que tienen, se ha utilizado la técnica de validación cruzada (ver subsección 2.1.3).

En nuestro caso, hemos utilizado 5 subconjuntos, denotados por $k = 5$, con un total de 50 épocas por iteración. En otras palabras, dividimos nuestro conjunto de entrenamiento en 5 partes. Durante cada iteración, entrenamos el modelo con 4 de estas partes durante 50 épocas, y luego lo validamos con la parte restante.

Bajo esta configuración, se han llevado a cabo múltiples ejecuciones, realizando ajustes en la arquitectura de los modelos y explorando distintos hiperparámetros de las capas. Entre los parámetros más significativos se encuentran las variaciones en el número de bloques de capas en las CNNs así como el número de filtros en la convolución. Los resultados de algunas de estas ejecuciones se presentarán en la sección 4.2.

Cabe destacar que el conjunto que se ha usado para la validación cruzada contiene las imágenes aumentadas que fueron generadas a partir de las originales. Por este motivo, tenemos que asegurarnos de que las imágenes que pertenecen a un mismo paciente queden agrupadas en la misma partición ya que no queremos que el conjunto de validación se contamine con datos que ya han sido "vistos" durante el entrenamiento. Cuando realizamos aumento de datos en imágenes, estamos generando versiones ligeramente modificadas de las mismas imágenes. Sin embargo, estas versiones todavía provienen del mismo paciente y, por lo tanto, comparten similitudes y características. Este inconveniente se ha solventado mediante el uso de StratifiedGroupKFold [Scikit-learn, 2023], el cual genera subconjuntos estratificados con grupos no solapados. Los pliegues son creados manteniendo el porcentaje de muestras de cada clase. En cada subconjunto, cada grupo (conjunto de imágenes correspondientes a un mismo paciente) aparecerá exactamente una vez en el conjunto de validación.

Además, dado que el conjunto de validación se genera a partir del conjunto de entrenamiento durante la validación cruzada, es importante garantizar que esta partición contenga solo imágenes originales y no aumentadas. Por lo tanto, antes de evaluar el modelo, se realiza un paso adicional para eliminar todas las imágenes que hayan sido aumentadas del conjunto de validación. Esto se lleva a cabo con el propósito de evaluar el rendimiento del modelo en condiciones realistas y evitar cualquier sesgo en la evaluación.

4.1.2. Grad-CAM

El objetivo principal de este experimento es comprender en qué áreas específicas se enfocan los diferentes modelos entrenados durante el proceso de predicción. Para ello se ha aplicado una técnica conocida como Grad-CAM (ver subsección 2.1.4), con la cual obtendremos una perspectiva visual sobre las regiones que cada modelo considera más relevantes para tomar sus decisiones de clasificación. En la sección 4.2.2 se muestran los resultados obtenidos con esta técnica sobre imágenes de diferentes pacientes.

4.2. Resultados

Esta sección presenta los resultados de los experimentos realizados durante la validación cruzada para ajustar los modelos. Además, mostraremos visualizaciones de los resultados de Grad-Cam en diferentes pacientes. Finalmente, evaluaremos el rendimiento de los modelos en el conjunto de datos de prueba utilizando las métricas detalladas en la sección 2.1.5.

4.2.1. Validación cruzada

En uno de los primeros intentos de ajuste de parámetros se utilizó un único bloque de capas en la red convolucional, donde el número de filtros estaba fijado en 64. En la tabla 4.1 se muestran los resultados de este primer ajuste.

Modelos	Exactitud (%)	Especificidad (%)	Precisión (%)	Exhaustividad (%)	Puntuación F1 (%)
CNN	80.89 %	84.50 %	78.49 %	76.04 %	77.25 %
CNN + Auto-atención	83.56 %	86.82 %	81.72 %	79.17 %	80.42 %
CNN + Atención suave	82.46 %	82.95 %	78.64 %	81.82 %	80.20 %
CNN + Atención dura	**86.67** %	**87.60** %	**83.67** %	**85.42** %	**84.54** %

Tabla 4.1: Resultados obtenidos por los distintos modelos en la validación cruzada con un solo bloque en las CNNs y 64 filtros en la convolución.

Los resultados distan mucho de ser óptimos, pero ya se pueden apreciar indicios de que la aplicación de los mecanismos de atención están mejorando el rendimiento, especialmente en el caso del modelo que incorpora atención dura. Así que, con el objetivo de intentar mejorar los resultados anteriores, se reemplazó el número de filtros por 128, manteniendo aún un único bloque de capas en la CNN. En la tabla 4.2 se muestran los resultados del segundo ajuste de parámetros en los modelos.

Modelos	Exactitud (%)	Especificidad (%)	Precisión (%)	Exhaustividad (%)	Puntuación F1 (%)
CNN	79.11 %	82.95 %	76.34 %	73.96 %	75.13 %
CNN + Auto-atención	81.78 %	85.27 %	79.57 %	77.08 %	78.31 %
CNN + Atención suave	80.89 %	81.40 %	76.24 %	80.21 %	78.17 %
CNN + Atención dura	**84.00** %	**86.05** %	**81.25** %	**81.25** %	**81.25** %

Tabla 4.2: Resultados obtenidos por los distintos modelos en la validación cruzada con un solo bloque en las CNNs y 128 filtros en la convolución.

En este caso, el aumento en la cantidad de filtros en la convolución no conduce a una mejora en los resultados, como se evidencia en la tabla anterior. De hecho, los resultados son considerablemente inferiores, a pesar de que los modelos con mecanismos de atención siguen superando al modelo base.

Finalmente, se ha optado por aplicar dos bloques de capas en la red convolucional, utilizando 64 filtros en la primera convolución y 128 en la segunda. Estos parámetros finales son los que se han utilizado en la arquitectura de los modelos propuestos, como se detalla en la sección 3.2. Los resultados de esta validación cruzada se presentan en la tabla 4.3.

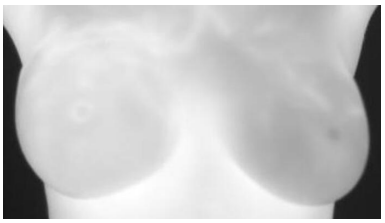
Modelos	Exactitud (%)	Especificidad (%)	Precisión (%)	Exhaustividad (%)	Puntuación F1 (%)
CNN	87.11 %	89.92 %	86.02 %	83.33 %	84.66 %
CNN + Auto-atención	89.78 %	92.25 %	89.25 %	86.46 %	87.83 %
CNN + Atención suave	90.22 %	89.92 %	87.00 %	90.62 %	88.78 %
CNN + Atención dura	**94.22** %	**96.12** %	**94.62** %	**91.67** %	**93.12** %

Tabla 4.3: Resultados obtenidos por los distintos modelos en la validación cruzada con dos bloques de CNNs.

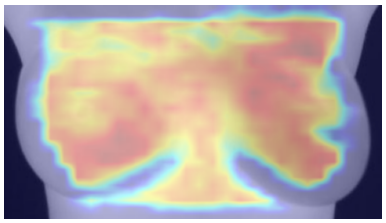
La inclusión de dos bloques de capas en la CNN ha resultado en una mejora notable en las métricas, destacando especialmente la exactitud del modelo CNN + Atención dura, que alcanzó un 94.22 %. Se realizaron pruebas adicionales para intentar superar estos resultados, pero en ningún caso se logró una mejora.

4.2.2. Grad-CAM

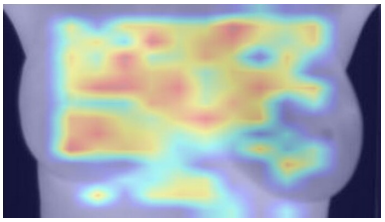
Esta técnica se ha aplicado sobre diferentes pacientes usando los distintos modelos implementados. En la figura 4.1 se pueden observar las áreas más relevantes que han considerado los modelos para realizar la predicción sobre la paciente 373. En el caso del modelo más simple (ver figura 4.1 b), observamos que prácticamente se enfoca en toda la imagen, asignando importancia a todas sus partes.



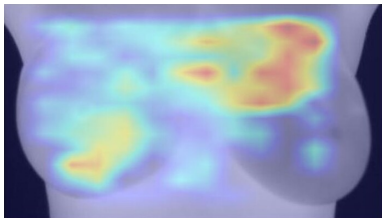
(a) Original



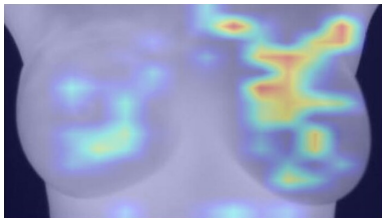
(b) CNN



(c) CNN + Auto-atención



(d) CNN + Atención suave



(e) CNN + Atención dura

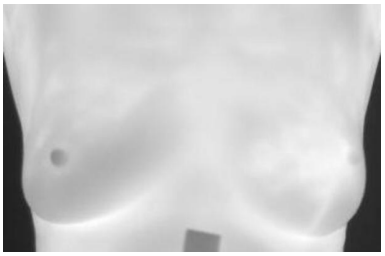
Figura 4.1: Método Grad-CAM usando los modelos entrenados sobre la paciente 373.

Sin embargo, la dinámica cambia en las demás figuras. Por ejemplo, en el caso del modelo que emplea auto-atención (figura 4.1 c), se pueden notar resaltadas solo ciertas zonas específicas como relevantes. Lo mismo ocurre para la atención suave y dura, que también presentan enfoques selectivos en las regiones clave de las imágenes. En el caso de la atención suave (Figura 4.1 d), se puede apreciar una distribución más ponderada de la atención, donde ciertas áreas reciben un mayor relevancia, mientras que otras zonas también contribuyen, pero con menos énfasis.

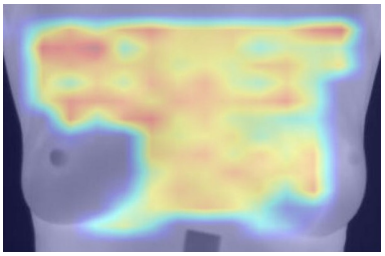
Por otro lado, la atención dura (Figura 4.1 e) se destaca por su enfoque específico en áreas altamente distintivas y decisivas. Esta estrategia se alinea con la naturaleza discreta de la atención dura, donde las regiones consideradas críticas para la predicción son seleccionadas con mayor precisión.

En el contexto de la detección de cáncer de mama en imágenes térmicas, las áreas más cálidas, que pueden indicar la presencia de anomalías o cambios en la temperatura de la mama, son típicamente las que los modelos con atención seleccionan como las más relevantes para la tarea de diagnóstico.

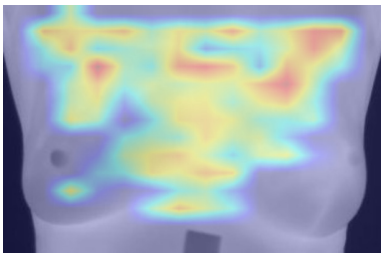
Las figuras 4.2, 4.3 y 4.4 presentan ejemplos adicionales de los resultados obtenidos mediante Grad-CAM en otras pacientes. Estos ejemplos exhiben patrones muy similares a los mostrados en la figura 4.1, con la atención dura destacándose como la más selectiva a la hora de la predicción.



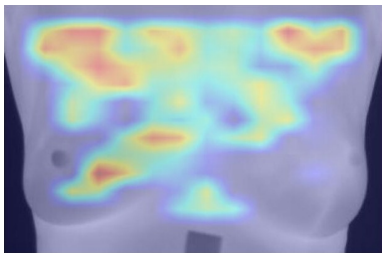
(a) Original



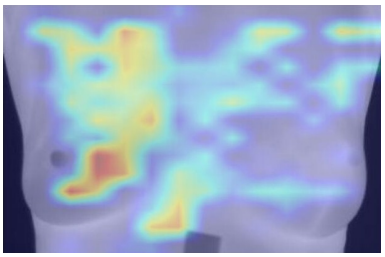
(b) CNN



(c) CNN + Auto-atención



(d) CNN + Atención suave



(e) CNN + Atención dura

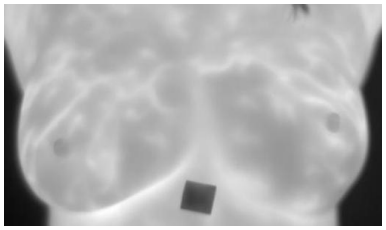
Figura 4.2: Método Grad-CAM usando los modelos entrenados sobre la paciente 344.

En conjunto, el análisis de los mapas de atención generados por Grad-CAM proporciona una comprensión más profunda de cómo cada modelo interpreta y utiliza la información visual para realizar sus predicciones. Estos *insights* visuales pueden ser de gran utilidad para interpretar y validar el comportamiento de los modelos en tareas de detección de cáncer de mama utilizando imágenes térmicas.

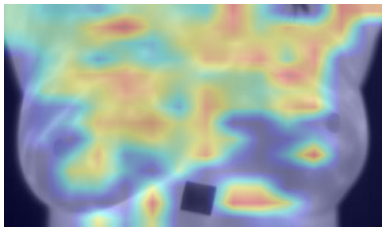
4.2.3. Evaluación sobre el conjunto de prueba

Por último, procederemos a evaluar el desempeño de los modelos implementados sobre el conjunto de datos de prueba, es decir, sobre imágenes que aún no han visto. A modo de resumen, cada modelo fue entrenado durante 50 épocas utilizando la función de pérdida de entropía cruzada y el optimizador Adam con una tasa de aprendizaje de 0.001.

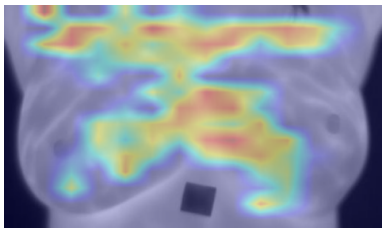
Comenzaremos con la visualización de las matrices de confusión obtenidas por cada uno de los modelos (ver figura 4.5). Se observa que el modelo base ha logrado clasificar correctamente



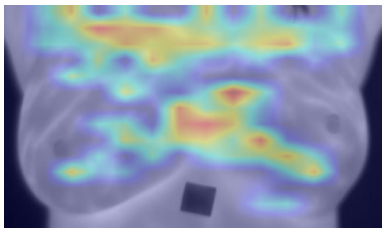
(a) Original



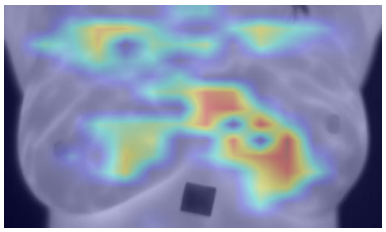
(b) CNN



(c) CNN + Auto-atención

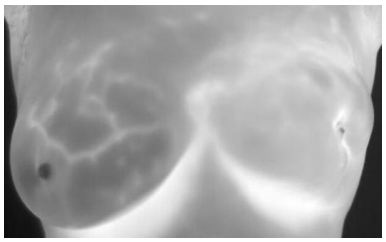


(d) CNN + Atención suave

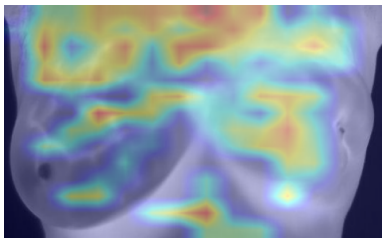


(e) CNN + Atención dura

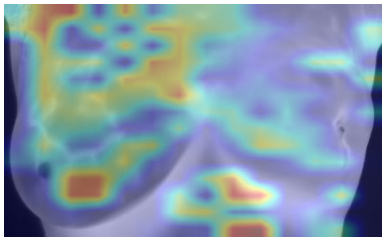
Figura 4.3: Método Grad-CAM usando los modelos entrenados sobre la paciente 283.



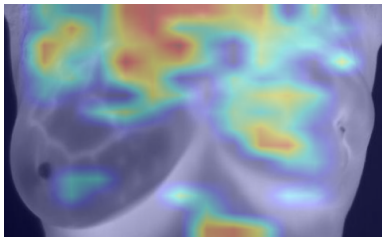
(a) Original



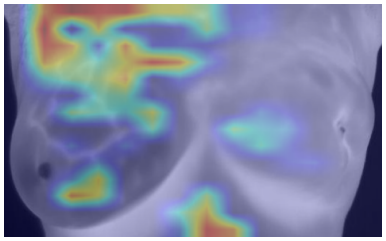
(b) CNN



(c) CNN + Auto-atención



(d) CNN + Atención suave

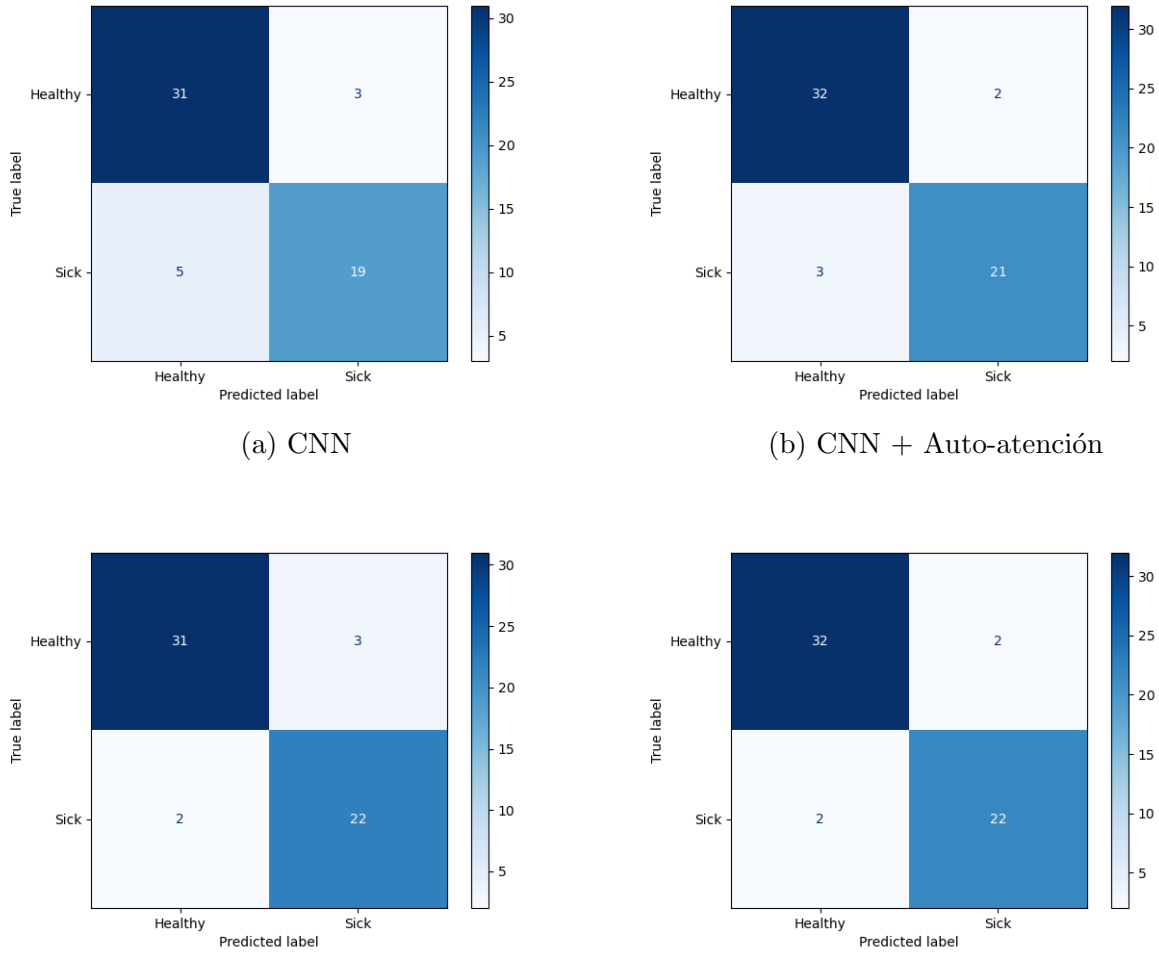


(e) CNN + Atención dura

Figura 4.4: Método Grad-CAM usando los modelos entrenados sobre la paciente 192.

en la mayoría de los casos, aunque con 8 equivocaciones en total. En contraste, el modelo que incorpora atención dura solo ha tenido 4 predicciones incorrectas. Los modelos con auto-atención y atención suave han mostrado resultados cercanos al modelo con atención dura, con un total de 5 fallos en cada caso. En todos los escenarios, se nota un equilibrio en las predicciones, sin una tendencia clara a clasificar una clase mejor que la otra.

La tabla 4.4 recoge los resultados de las diferentes métricas propuestas para la evaluación de los modelos. En términos de exactitud, el modelo base logró un 86.21 %, lo que indica un rendimiento sólido en la identificación general de casos. Sin embargo, como ya veníamos



(a) CNN

(b) CNN + Auto-atención

(c) CNN + Atención suave

(d) CNN + Atención dura

Figura 4.5: Matrices de confusión obtenidas por los diferentes modelos.

viendo en la validación cruzada, al introducir mecanismos de atención se observa una mejora notable. Tanto el modelo CNN + Auto-atención como el modelo CNN + Atención suave lograron un 91.38 % de exactitud, superando al modelo base. Y por delante del resto, el modelo que implementa la atención dura ha obtenido un 93.10 % de aciertos.

Modelos	Exactitud (%)	Especificidad (%)	Precisión (%)	Exhaustividad (%)	Puntuación F1 (%)	AUC
CNN	86.21 %	91.18 %	86.36 %	79.17 %	82.61 %	0.8517
CNN + Auto-atención	91.38 %	**94.12 %**	91.30 %	87.50 %	89.36 %	0.9081
CNN + Atención suave	91.38 %	91.18 %	88.00 %	**91.67 %**	89.80 %	0.9142
CNN + Atención dura	**93.10 %**	**94.12 %**	**91.67 %**	**91.67 %**	**91.67 %**	**0.9289**

Tabla 4.4: Resultados obtenidos sobre el conjunto de prueba por los distintos modelos.

En lo que respecta a la especificidad, el modelo CNN + Auto-atención y el modelo CNN + Atención dura alcanzaron un 94.12 %, lo que sugiere una gran habilidad para identificar con precisión los casos negativos verdaderos.

La precisión, que mide la proporción de predicciones positivas verdaderas, demostró un

aumento en los modelos con atención. El modelo CNN + Atención dura lidera con una precisión del 91.67 %, seguido por el modelo CNN + Auto-atención y el modelo CNN + Atención suave.

La exhaustividad muestra que los modelos con atención presentan una capacidad equilibrada para identificar casos positivos verdaderos. El modelo CNN + Atención suave junto al modelo CNN + Atención dura lideran en este aspecto con un 91.67 %, seguido de cerca por el modelo CNN + Auto-atención. En este aspecto el modelo base cae bastante con un 79.17 %.

La puntuación F1, que combina precisión y exhaustividad, sigue la tendencia de mejora en los modelos con atención. El modelo CNN + Atención dura destaca con una puntuación F1 del 91.67 %, demostrando su capacidad para encontrar un equilibrio entre la identificación precisa de casos positivos y la reducción de falsos positivos.

Finalmente, el área bajo la curva ROC (AUC) muestra que el modelo CNN + Atención dura también tiene el mejor rendimiento general con un valor de 0.9289, indicando su alta capacidad pero no excelente para distinguir entre clases positivas y negativas (ver figura 4.6).

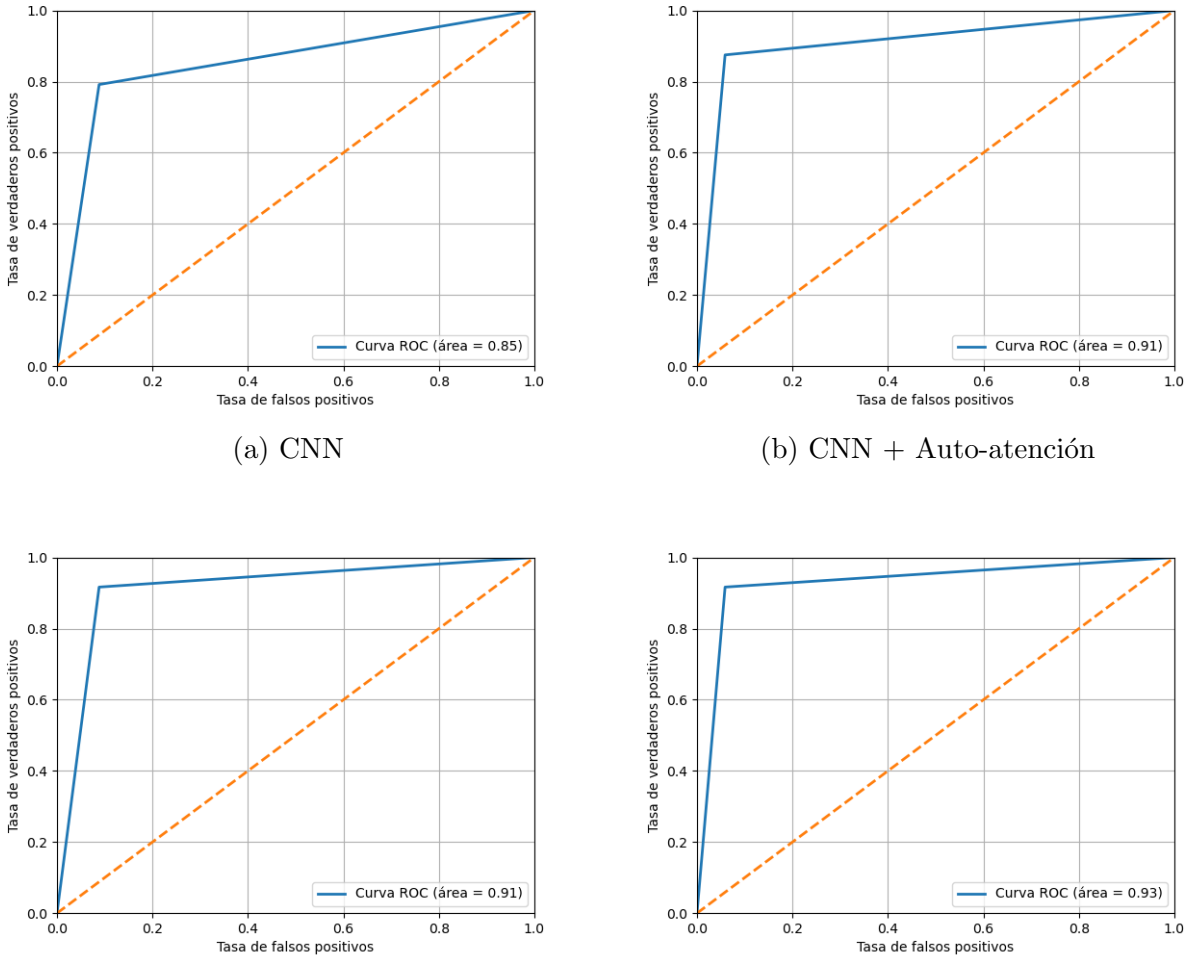
En resumen y aunque los resultados no sean asombrosos, la incorporación de mecanismos de atención en los modelos de detección de cáncer de mama ha demostrado ser altamente beneficiosa. Los modelos que aplican mecanismos de atención han superado consistentemente al modelo CNN base en todas las métricas evaluadas. Estos resultados sugieren que la atención juega un papel esencial en la mejora de la precisión y el rendimiento general de los modelos en esta tarea crítica de clasificación médica.

4.3. Discusión

Esta sección se centra en la comparación entre los resultados obtenidos por los modelos propuestos y los resultados reportados por otros autores, tal como se presentaron en la sección 2.2.

Numerosos estudios han contribuido significativamente a la detección del cáncer de mama, empleando diversas arquitecturas de modelos y utilizando distintos tipos de imágenes como conjuntos de datos, ya sean histopatológicas, mamográficas o térmicas. En la mayoría de estos estudios, la métrica de exactitud es ampliamente utilizada como medida de evaluación. Algunos de ellos, pero en menor medida, también aportan métricas como la especificidad o la exhaustividad. Sin embargo, se observa una tendencia en la comunidad científica a no explorar en detalle otras métricas esenciales, como la precisión, la puntuación F1 o el área bajo la curva ROC. Esta omisión de métricas adicionales puede limitar la comprensión completa del rendimiento de los modelos y la capacidad de comparación entre diferentes enfoques de detección de cáncer de mama.

Destacamos que el estudio realizado por [Alshehri and AlSaeed, 2022] logró el mejor



(a) CNN

(b) CNN + Auto-atención

(c) CNN + Atención suave

(d) CNN + Atención dura

Figura 4.6: Curvas ROC y valores AUC obtenidos por los diferentes modelos.

resultado de los que se han descrito, obteniendo una impresionante tasa de clasificación del 99.46 % al emplear redes neuronales convolucionales aplicando mecanismos de atención, utilizando el conjunto de datos DMI-IR. En contraste, el mejor resultado obtenido en nuestro estudio ha sido una exactitud del 93.10 %, utilizando una CNN con atención dura.

Es relevante destacar que, a pesar de que tanto su estudio como el nuestro ha utilizado el mismo conjunto de datos para realizar sus investigaciones, han habido diferencias significativas en la metodología. En particular, en [Alshehri and AlSaeed, 2022], los autores llevaron a cabo una segmentación previa de las imágenes, centrándose exclusivamente en las regiones mamarias de interés. Este enfoque podría haber contribuido a la obtención de tasas de clasificación más altas al reducir el ruido y las distracciones en los datos. Por lo tanto, es importante considerar estas diferencias metodológicas al comparar los resultados.

Otro trabajo con muy buenos resultados es [Ekici and Jawzal, 2020], donde los autores, utilizando redes neuronales convolucionales optimizadas por el algoritmo de Bayes, alcanzaron

una tasa de clasificación del 98.95 %. Cabe destacar que, en este estudio, además de utilizar imágenes térmicas para el entrenamiento del modelo, se aprovechó información adicional de los pacientes, como los bio-datos, lo que posiblemente contribuyó a su notable desempeño. Esta combinación de datos ofrece una perspectiva interesante para futuras investigaciones en la detección del cáncer de mama basada en imágenes térmicas y datos clínicos.

En un enfoque similar, pero empleando una arquitectura diferente, encontramos un ejemplo en [Sánchez-Cauce et al., 2021], donde los investigadores introdujeron una red neuronal convolucional de múltiples entradas innovadora. Esta red combina imágenes térmicas capturadas desde diferentes ángulos de visión con datos personales y clínicos de los pacientes. Sus resultados son igualmente notables, con una exactitud excepcional del 97 %, una especificidad perfecta del 100 %, y una exhaustividad del 83 %. En comparación, nuestros modelos CNN + Atención suave y CNN + Atención dura lograron una exhaustividad del 91.67 %, lo que indica que nuestros enfoques tienen una capacidad equilibrada aún mayor para detectar casos positivos verdaderos.

Por otro lado, es importante destacar que nuestros resultados superan a los obtenidos en [Patil and Biradar, 2021], un estudio en el cual los autores llevaron a cabo una combinación optimizada de redes neuronales convolucionales y recurrentes para la detección automatizada del cáncer de mama en mamografías. A pesar de que su investigación logró una precisión del 90.59 %, nuestro mejor modelo supera este desempeño al alcanzar una precisión del 91.67 %. Cabe destacar que el tipo de imágenes utilizado es diferente; mientras que [Patil and Biradar, 2021] se centró en mamografías, nuestro enfoque se basó en imágenes térmicas de las mamas. Por lo tanto, nuestros resultados destacan la eficacia de la detección del cáncer de mama basada en imágenes térmicas como una alternativa prometedora y complementaria a los métodos tradicionales de detección basados en mamografías.

En resumen, este apartado ha proporcionado una visión general de la investigación relacionada con la detección del cáncer de mama utilizando imágenes médicas, destacando algunos de los avances logrados por otros investigadores. Hemos observado que existen enfoques muy prometedores y exitosos en esta área, con tasas de clasificación impresionantes. Nuestro estudio se ha sumado a este esfuerzo al presentar resultados sólidos y, en algunos casos, competitivos utilizando imágenes térmicas de las mamas y modelos de aprendizaje profundo. En particular, destacamos el rendimiento de nuestra CNN con atención dura. Aún así, reconocemos que hay margen para futuras mejoras y optimizaciones en nuestro estudio para la detección del cáncer de mama basada en imágenes termográficas.

Capítulo 5

Conclusiones y trabajos futuros

En este último capítulo, concluiremos nuestro estudio sobre la detección del cáncer de mama a través de imágenes infrarrojas y presentaremos una visión general de los resultados y logros obtenidos en este trabajo. Además, identificaremos una serie de trabajos futuros que podrían ampliar y mejorar más los resultados obtenidos.

5.1. Conclusiones

En este trabajo, se ha abordado el desafío de la detección del cáncer de mama a través de imágenes térmicas utilizando técnicas de aprendizaje profundo y mecanismos de atención. Se han implementado y evaluado diferentes modelos, incluyendo una red neuronal convolucional como base y variantes de ésta que incorporaban mecanismos de atención: Auto-atención, Atención Suave y Atención Dura.

Aunque los resultados obtenidos en la evaluación de los modelos están lejos de ser óptimos, sí que han arrojado valiosas conclusiones. Se ha observado que la inclusión de mecanismos de atención en las CNNs produjo mejoras significativas en la capacidad de detección y clasificación de imágenes termográficas para el cáncer de mama. El modelo base logró una precisión del 86.21 %, mientras que las variantes con atención alcanzaron un rango de precisión entre 91.38 % y 93.10 %. Esta diferencia evidencia el impacto positivo de los mecanismos de atención en la mejora del rendimiento de los modelos.

El análisis mediante Grad-CAM ha permitido comprender cómo cada modelo interpreta y utiliza la información visual para realizar predicciones. Se ha observado que los modelos con atención tenían un enfoque más selectivo en regiones específicas de las imágenes, lo que respalda la idea de que los mecanismos de atención ayudan a destacar las áreas relevantes para la detección del cáncer. En particular, el modelo con atención dura demostró una alta concentración en áreas altamente relevantes, lo que sugiere que este mecanismo de atención toma decisiones muy específicas al asignar peso a las subpartes de la imagen. Esta atención

selectiva puede ser especialmente beneficiosa en aplicaciones médicas donde la precisión y la identificación de regiones críticas son fundamentales para el diagnóstico.

El tamaño del conjunto de datos utilizado supone un reto para cualquier método de aprendizaje automático, ya que actualmente se dispone de un número limitado de muestras. Esto afecta a la generalización de los modelos y limita su capacidad para aprender. A pesar de haber empleado técnicas de aumento de datos, es importante recordar que éstas no tienen siempre un efecto milagroso y pueden no abordar por completo el inconveniente del tamaño. La diversidad y cantidad de datos son fundamentales para que los modelos de aprendizaje automático puedan generalizar de manera efectiva y capturar la variabilidad presente en los casos reales.

5.2. Trabajos futuros

Para mejorar los resultados obtenidos en este trabajo, se plantean diversas líneas de investigación que podrían ser abordadas en futuros estudios. Estas incluyen:

- Uso del *transfer learning*: Consiste en una categoría dentro del aprendizaje automático donde se reutilizan modelos preexistentes para resolver retos actuales [Hosna et al., 2022]. Aplicar esto utilizando modelos preentrenados en grandes conjuntos de datos puede mejorar la capacidad de generalización de los modelos, especialmente en casos de conjuntos de datos pequeños como en este estudio. A estos modelos se les añadiría los mecanismos de atención vistos en el presente trabajo.

- Consideración de otros tipos de datos médicos: Combinar imágenes térmicas con datos clínicos de los pacientes, tal y como se hizo en parte en [Sánchez-Cauce et al., 2021]. Esto podría aumentar la precisión del diagnóstico y brindar una visión más completa de la salud de los pacientes.

- Optimización de hiperparámetros y arquitecturas: Realizar una búsqueda más exhaustiva de hiperparámetros y arquitecturas de modelos para encontrar combinaciones óptimas que puedan maximizar el rendimiento y la eficiencia de la detección.

Bibliografía

[Alshehri and AlSaeed, 2022] Alshehri, A. and AlSaeed, D. (2022). Breast cancer detection in thermography using convolutional neural networks (cnns) with deep attention mechanisms. *Applied Sciences*, 12(24).

[Alzubaidi et al., 2021] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L. (2021). Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1):53.

[Berrar, 2019] Berrar, D. (2019). Cross-validation. In Ranganathan, S., Gribskov, M., Nakai, K., and Schönbach, C., editors, *Encyclopedia of Bioinformatics and Computational Biology*, pages 542–545. Academic Press, Oxford.

[Centers for Disease Control and Prevention, 2023] Centers for Disease Control and Prevention (2023). What is a mammogram? `https://www.cdc.gov/cancer/breast/basic_info/mammograms.htm` [Accedido: 26 de agosto de 2023].

[de Freitas Oliveira Baffa and Grassano Lattari, 2018] de Freitas Oliveira Baffa, M. and Grassano Lattari, L. (2018). Convolutional neural networks for static and dynamic breast infrared imaging classification. In *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 174–181.

[de Santana Correia and Colombini, 2022] de Santana Correia, A. and Colombini, E. L. (2022). Attention, please! a survey of neural attention models in deep learning. *Artificial Intelligence Review*, 55(8):6037–6124.

[Deng et al., 2020] Deng, J., Ma, Y., Deng-ao, L., Zhao, J., Liu, Y., and Zhang, H. (2020). Classification of breast density categories based on se-attention neural networks. *Computer Methods and Programs in Biomedicine*, 193:105489.

[Ekici and Jawzal, 2020] Ekici, S. and Jawzal, H. (2020). Breast cancer diagnosis using thermography and convolutional neural networks. *Medical Hypotheses*, 137:109542.

[Fan et al., 2020] Fan, M., Chakraborti, T., Chang, E. I., Xu, Y., and Rittscher, J. (2020). Microscopic fine-grained instance classification through deep attention. *CoRR*, abs/2010.02818.

[Fang et al., 2020] Fang, W., Love, P. E. D., Luo, H., and Ding, L. (2020). Computer vision for behaviour-based safety in construction: A review and future directions. *Adv. Eng. Informatics*, 43:100980.

[Gu et al., 2018] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., and Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, 77:354–377.

[Gurcan et al., 2009] Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., and Yener, B. (2009). Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, 2:147–171.

[Hosna et al., 2022] Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., and Azim, M. A. (2022). Transfer learning: a friendly introduction. *Journal of Big Data*, 9(1):102.

[HumanSignal, 2023] HumanSignal (2023). Open source data labeling platform. `https://labelstud.io/` [Accedido: 27 de mayo de 2023].

[Kennedy et al., 2009] Kennedy, D. A., Lee, T., and Seely, D. (2009). A comparative review of thermography as a breast cancer screening technique. *Integrative Cancer Therapies*, 8(1):9–16. PMID: 19223370.

[Li et al., 2020] Li, Hsiao-Chi, Deng, Zong-Yue, Chiang, and Hsin-Han (2020). Lightweight and resource-constrained learning network for face recognition with performance optimization. *Sensors*, 20(21).

[Liu et al., 2017] Liu, Q., Zhou, F., Hang, R., and Yuan, X. (2017). Bidirectional-convolutional lstm based spectral-spatial feature learning for hyperspectral image classification. *Remote Sensing*, 9(12).

[Nahid et al., 2018] Nahid, A., Mehrabi, A., and Kong, Y. (2018). Histopathological breast cancer image classification by deep neural network techniques guided by local clustering. *BioMed Research International*, 2018:1–20.

[Organización Mundial de la Salud, 2023] Organización Mundial de la Salud (2023). Cáncer de mama. `https://www.who.int/es/news-room/fact-sheets/detail/breast-cancer` [Accedido: 5 de agosto de 2023].

[Palaz et al., 2019] Palaz, D., Magimai-Doss, M., and Collobert, R. (2019). End-to-end acoustic modeling using convolutional neural networks for hmm-based automatic speech recognition. *Speech Communication*, 108:15–32.

[Patil and Biradar, 2021] Patil, R. S. and Biradar, N. (2021). Automated mammogram breast cancer detection using the optimized combination of convolutional and recurrent neural network. *Evolutionary Intelligence*, 14(4):1459–1474.

[Pytorch, 2023] Pytorch (2023). Multiheadattention. `https://pytorch.org/docs/stable/generated/torch.nn.MultiheadAttention.html` [Accedido: 10 de agosto de 2023].

[Rakhunde et al., 2022] Rakhunde, B., M., Gotarkar, S., Choudhari, and G., S. (2022). Thermography as a breast cancer screening technique: A review article. *Cureus*, 14(11):e31251.

[Rashed and El Seoud, 2019] Rashed, E. and El Seoud, M. S. A. (2019). Deep learning approach for breast cancer diagnosis. *Proceedings of the 2019 8th International Conference on Software and Information Engineering*, 13(1):161.

[Refaeilzadeh et al., 2009] Refaeilzadeh, P., Tang, L., and Liu, H. (2009). *Cross-Validation*, pages 532–538. Springer US, Boston, MA.

[Santana et al., 2018] Santana, M., Pereira, J. M., Da Silva, F., Lima, N., Sousa, F., Arruda, G., Lima, R., Azevedo, W., and Dos Santos, W. (2018). Breast cancer diagnosis based on mammary thermography and extreme learning machines. 34:45–53.

[Schünemann et al., 2020] Schünemann, H., Lerda, D., Quinn, C., Follmann, M., Alonso, P., Giorgi Rossi, P., Lebeau, A., Nyström, L., Broeders, M., Ioannidou-Mouzaka, L., Duffy, S., Borisch, B., Fitzpatrick, P., Hofvind, S., Castells, X., Giordano, L., Canelo-Aybar, C., Warman, S., Mansel, R., and Saz-Parkinson, Z. (2020). Breast cancer screening and diagnosis: A synopsis of the european breast guidelines. *ANNALS OF INTERNAL MEDICINE*, 172(1):45–56.

[Scikit-learn, 2023] Scikit-learn (2023). Stratifiedgroupkfold. `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedGroupKFold.html` [Accedido: 11 de agosto de 2023].

[Selvaraju et al., 2017] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626.

[Shen et al., 2018] Shen, T., Zhou, T., Long, G., Jiang, J., Wang, S., and Zhang, C. (2018). Reinforced self-attention network: a hybrid of hard and soft attention for sequence modeling. pages 4345–4352.

[Sánchez-Cauce et al., 2021] Sánchez-Cauce, R., Pérez-Martín, J., and Luque, M. (2021). Multi-input convolutional neural network for breast cancer detection using thermal images and clinical data. *Computer Methods and Programs in Biomedicine*, 204:106045.

[Tello-Mijares et al., 2019] Tello-Mijares, S., Woo, F., and Flores, F. (2019). Breast cancer identification via thermography image segmentation with a gradient vector flow and a convolutional neural network. *Journal of Healthcare Engineering*, 2019:9807619.

[Tian et al., 2021] Tian, H., Wang, P., Tansey, K., Han, D., Zhang, J., Zhang, S., and Li, H. (2021). A deep learning framework under attention mechanism for wheat yield estimation using remotely sensed indices in the guanzhong plain, pr china. *International Journal of Applied Earth Observation and Geoinformation*, 102:102375.

[Toğaçar et al., 2020] Toğaçar, M., Özkurt, K. B., Ergen, B., and Cömert, Z. (2020). Breastnet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer. *Physica A: Statistical Mechanics and its Applications*, 545:123592.

[Vaswani et al., 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *CoRR*, abs/1706.03762.

[Zhang et al., 2019] Zhang, X., Zhang, Y., Qian, B., Liu, X., Li, X., Wang, X., Yin, C., Lv, X., Song, L., and Wang, L. (2019). Classifying breast cancer histopathological images using a robust artificial neural network architecture. In Rojas, I., Valenzuela, O., Rojas, F., and Ortuño, F., editors, *Bioinformatics and Biomedical Engineering*, pages 204–215, Cham. Springer International Publishing.

Anexo A

Conjunto de datos

A.1. Descripción

La Base de Datos para la Investigación Mastológica con Imágenes Infrarrojas (DMR-IR, por sus siglas en inglés, que significa *Database for Mastology Research with Infrared Image*) es un conjunto de datos que contiene información de exámenes mamarios y datos clínicos obtenidos de pacientes voluntarias del Hospital Universitário Antônio Pedro (HUAP) de la Universidad Federal Fluminense (Brasil).

Las imágenes infrarrojas que componen la base de datos han sido obtenidas usando dos protocolos diferentes: estático y dinámico.

Estático:

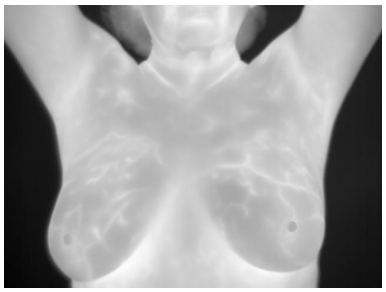
Consiste en 5 imágenes tomadas desde diferentes ángulos: 1 frontal, 2 laterales a 45^{o} (lado derecho e izquierdo) y 2 laterales a 90^{o} (lado derecho e izquierdo). En la figura A.1 se puede ver un ejemplo de este protocolo para una de las pacientes.

Dinámico:

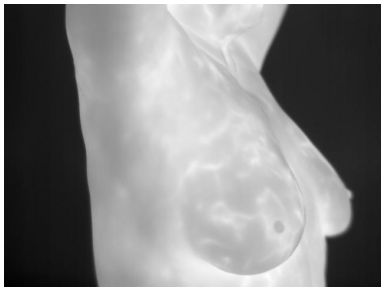
Consiste en una serie de imágenes (normalmente unas 20), en la que previamente se ha enfriado la zona de las mamas, tomadas cada 15 segundos durante 5 minutos o hasta que la temperatura original del cuerpo es alcanzada. Finalmente se realizan 2 capturas más, una de la mama izquierda y otra de la mama derecha, ambas a 90^{o}. En la figura A.2 se puede ver un ejemplo de este protocolo para una de las pacientes.

Estos protocolos pueden ser consultados con más detalle en la página web oficial de esta base de datos (`http://visual.ic.uff.br/dmi`).

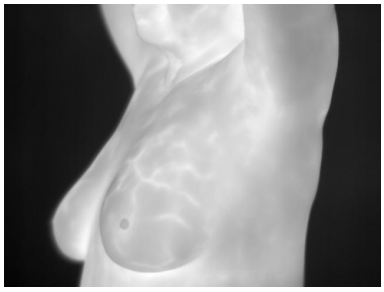
Actualmente, DMR-IR cuenta con información sobre 280 pacientes, de los cuales 176



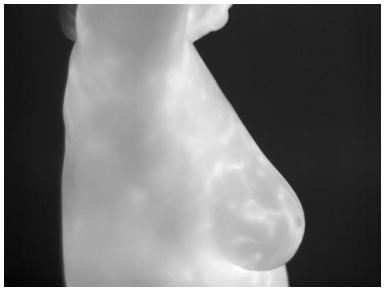
(a) Frontal



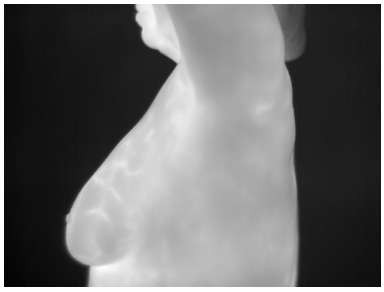
(b) Derecha 45º



(c) Izquierda 45º



(d) Derecha 90º



(e) Izquierda 90º

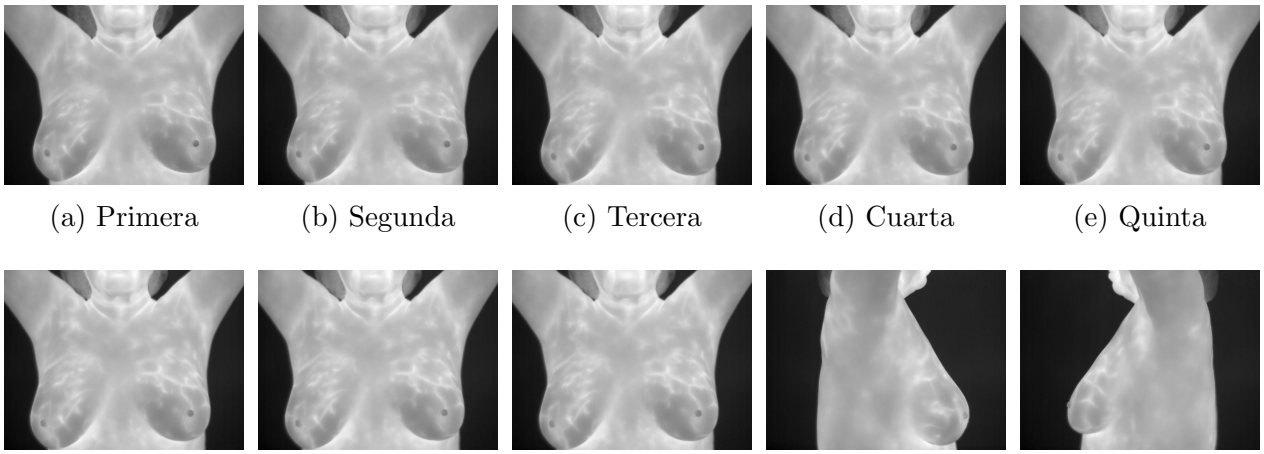
Figura A.1: Ejemplo de imágenes tomadas usando el protocolo estático.

están clasificados como *Healthy*, 100 como *Sick* y 4 como *Unknown*. Las imágenes tienen un tamaño de 680 píxeles de ancho y 480 de alto. En la figura A.3 se puede observar un ejemplo de cada imagen con su respectiva clase a la que pertenece.

A.2. Descarga

Aunque acceder a los datos es una tarea sencilla, simplemente debemos crearnos una cuenta en `http://visual.ic.uff.br`, descargar el conjunto de datos completo no es nada fácil ya que no existe forma ninguna para hacerlo.

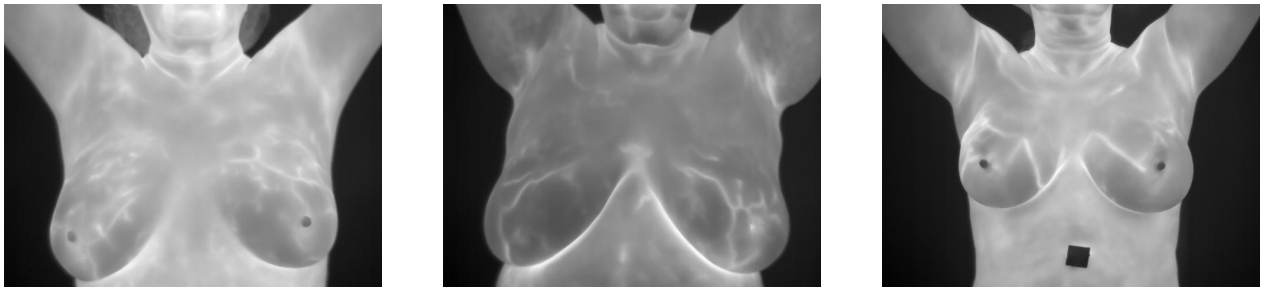
En mi caso, he tenido que desarrollar un script usando Python para descargarlo. Básicamente lo que hace este script es *web scrapping*, es decir, extraer contenidos y datos del sitio



(a) Primera (b) Segunda (c) Tercera (d) Cuarta (e) Quinta

(f) Sexta (g) Séptima (h) Octava (i) Derecha 90º (j) Izquierda 90º

Figura A.2: Ejemplo de imágenes tomadas usando el protocolo dinámico.



(a) *Healthy* (b) *Unknown* (c) *Sick*

Figura A.3: Ejemplo de imágenes con su respectiva etiqueta.

web donde se encuentra la base de datos.

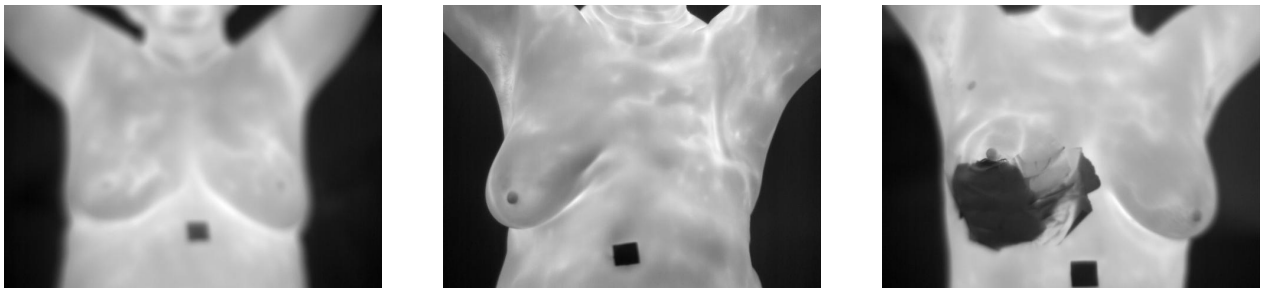
Las imágenes térmicas las podemos encontrar en dos formatos distintos. Por un lado en JPG y por otro lado en un TXT, que almacena matrices de números con el valor de la temperatura en cada píxel. Se ha optado por la segunda opción por lo que el script ha ido visitando cada una de las pacientes disponibles y ha descargado los archivos TXT correspondientes de las imágenes térmicas. Finalmente otro script, también desarrollado en Python, se ha encargado de transformar los ficheros TXT en imágenes.

A.3. Limpieza

La limpieza de conjuntos de datos es un proceso fundamental en la preparación de datos para el entrenamiento de modelos. En esta base de datos encontramos ciertos casos que debemos de tratar para mejorar la calidad y fiabilidad de los modelos entrenados. Por un lado, y como ya se ha mencionado en la sección A.1, existen 4 pacientes etiquetados con la clase *Unknown*, es decir, se desconoce si esa paciente está enferma o no. Son las pacientes 1, 275, 280 y 284, que no se tendrán en cuenta en el dataset final.

Por otra parte, también existe la presencia de pacientes duplicados como son los casos de los IDs 90/91, 153/154 y 189/193, por lo que serán borrados también del conjunto de datos final.

Por último, hay ciertos pacientes en las que todas o algunas de las fotografías tomadas pueden no ser válidas. Las razones encontradas son que hay imágenes borrosas, pacientes con mastectomías y pacientes con algún tipo de ropa en la parte del pecho que deforma la imagen. El uso de estas imágenes puede alterar el rendimiento del modelo ya que se trata de casos anómalos. En la figura A.4 se puede ver ejemplos de imágenes no válidas debido a los casos comentados.



(a) Imagen borrosa (b) Mastectomía (c) Ropa en el seno

Figura A.4: Ejemplo de imágenes no válidas.

Los casos de imágenes borrosas e imágenes en los que hay ropa deben de ser borradas ya que impiden la visión de la temperatura en ciertas zonas importantes que necesitan los modelos a entrenar. El caso de las imágenes con mastectomías se van a dejar en el conjunto de datos debido a que aparecen en ambas clases con una proporción muy parecida y no provocarán sesgo en el conjunto. En la tabla A.1 se pueden ver los IDs de las pacientes cuyas imágenes sufren de alguna las causas comentadas.

Anomalía	**IDs pacientes sanas**	**IDs pacientes enfermas**
Borrosa	1, 3, 18, 141, 159, 182, 183, 184	255, 268, 285
Mastectomía	10, 46, 47, 92, 94, 107, 114, 156, 183, 197, 206	192, 203, 256, 258, 345, 361, 363, 381
Ropa	109, 185	242

Tabla A.1: Anomalías encontradas en el conjunto de datos junto con el ID del paciente.

Con esta limpieza, el número de pacientes final del conjunto de datos es 260, de los cuales 166 son pacientes sanas y 94 enfermas.