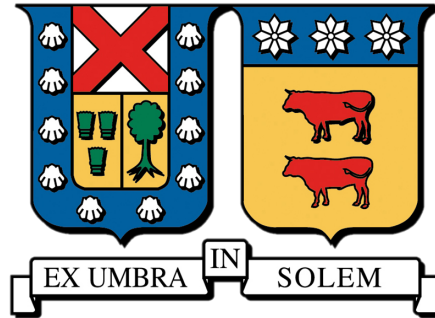


**UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA**  
**DEPARTAMENTO DE ELECTRÓNICA**  
**VALPARAÍSO - CHILE**



**“CLASIFICACIÓN DE IMÁGENES DE  
MAMOGRAFÍAS MEDIANTE MODELOS DE  
INTELIGENCIA ARTIFICIAL BASADOS EN  
VISION TRANSFORMER PARA LA DETECCIÓN  
DE TEJIDO CANCEROSO”**

**JULIO ANÍBAL MATURANA SOLÍS**

**MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE INGENIERO  
CIVIL TELEMÁTICO**

**PROFESOR GUÍA:**

**MAURICIO ARAYA**

**PROFESOR CORREFERENTE:**

**MOHAMED ABDELHAMID**

**Octubre 2025**



## CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

### 1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

**Tipo de monografía (marcar una opción):**  Memoria o trabajo de título;  Tesis de Postgrado;

**Título del trabajo:** Clasificación de imágenes de mamografías mediante modelos de inteligencia artificial basados en visión transformer para la detección de tejido canceroso.

**Nombre del candidato(a):** Julio Aníbal Maturana Solís

**Carrera / Grado:** Ingeniería Civil Telemática

**Campus:** Casa Central Valparaíso ; **Departamento:** Electrónica

### 2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Mauricio Araya López, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución

### 3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL

El trabajo **NO contiene información que amerite confidencialidad** y puede ser publicado de inmediato en repositorio con acceso abierto.

El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (embargo) por:

6 meses;  12 meses;  2 años;  3 años;  5 años;  10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

### 4.- FIRMAS

**Profesor(a) guía o director(a) de memoria o tesis:**

**Fecha:** 20-10-2025

**Firma:** 

**Estudiante o Candidato(a):**

**Fecha:** 20-10-2025

**Firma:** 

*Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.*

BIB-FO2385 01

## Agradecimientos

Quisiera expresar mi más sincero agradecimiento a mi mamá, Paola, por todo su cariño, apoyo y aliento durante todos estos años; sin ella nada de esto sería posible. A mi abuelita Sara, a quien extraño y recuerdo todos los días, este logro se lo dedico en especial a ella. A mis tíos y primos, cada uno de ellos contribuyó de una u otra forma a que llegara hasta esta instancia.

Agradezco a la universidad por permitirme crecer tanto como persona como profesional, y a los profesores, por su dedicación en nuestro aprendizaje. A mis compañeros y con quienes compartí durante este tiempo.

Gracias a Dios, por permitirme llegar hasta aquí y nunca estar solo.

Y gracias a *twenty one pilots*, por acompañarme con su música y letras durante las largas horas de estudio y trabajo.

*No, I don't know which way I'm going*

*But I can hear my way around*

— *twenty one pilots*, The Hype



## Resumen

El cáncer de mama es la principal causa de muerte por cáncer en la población femenina mundial y lograr una detección temprana es crucial para un tratamiento efectivo. El desarrollo de sistemas CAD (Diagnóstico asistido por computador) basados en CNN (Redes neuronales convolucionales) han demostrado ser de gran ayuda para los médicos, sin embargo, este tipo de modelos trabajan sobre una porción de la imagen, dejando fuera características relevantes y además, presentan una gran complejidad computacional debido a las múltiples convoluciones que se realizan.

En este trabajo se presenta un modelo de deep learning basado en el paradigma de vision transformer, el cual emplea transfer-learning al tomar como base modelos pre-entrenados y que son adaptados mediante fine-tuning. Como resultado, se obtiene una arquitectura compuesta por tres clasificadores entrenados de forma independiente, que colectivamente se encargan de entregar como resultado si hay presencia de lesiones, junto con su tipo y nivel BI-RADS correspondiente.

***Keywords: Vision transformer, Inteligencia artificial, Cáncer de mama, Mamografía, Clasificación***



## **Abstract**

Breast cancer is the leading cause of cancer-related death among the female population worldwide, and achieving early detection is crucial for effective treatment. The development of CAD (computer-aided diagnosis) systems based on Convolutional Neural Networks (CNNs) has proven to be highly beneficial for physicians. However, these models typically operate on a portion of the image, potentially excluding relevant features, and they also involve high computational complexity due to the numerous convolutions performed.

This work presents a deep learning model based on the Vision Transformer paradigm, which employs transfer learning by using pre-trained models that are adapted through fine-tuning. As a result, an architecture composed of three independently trained classifiers is obtained; collectively, they are responsible for determining the presence of lesions, their type, and the corresponding BI-RADS level.

***Keywords: Vision transformer, Artificial intelligence, Breast cancer, Mammography, Classification***



## Glosario

***Bounding Box*** (Cuadro delimitador): Es un rectángulo dibujado alrededor de una región de interés en una imagen. Esta región debe corresponder a un objeto o área específica que se desea identificar.

***Data augmentation*** (Aumento de datos): Es el proceso de crear variaciones de imágenes existentes en un conjunto de datos para crear nuevas imágenes.

***Dataset*** (Conjunto de datos): Es una colección estructurada de datos organizados y almacenados juntos para su análisis o procesamiento.

***DICOM***: Es un estándar internacional para la gestión, el almacenamiento, la transmisión y el intercambio de datos de imágenes médicas, incluyendo imágenes, información relacionada y documentación clínica.

***Epoch***: En el contexto del aprendizaje automático, se refiere a un paso completo del conjunto de datos de entrenamiento a través del algoritmo.

***Fine-Tuning***: El ajuste fino en aprendizaje automático consiste en adaptar un modelo ya entrenado a tareas o casos de uso específicos.

***Full-Field Digital Mammography***: Es un sistema de mamografía digital en el que la película se reemplaza por detectores de estado sólido que convierten los rayos X en señales eléctricas.

***Incidencia***: En epidemiología, proporción de casos nuevos de una enfermedad en un determinado período de tiempo, respecto a la población expuesta a padecerla.

***Metadatos***: Los metadatos son información, como el autor, la fecha de creación, el tamaño del archivo, entre otros, que describe un registro o un conjunto de datos, independiente del contenido de los mismos.

***Prevalencia***: Es una medida del número total de personas en un grupo específico que tienen (o tuvieron) cierta enfermedad, afección o factor de riesgo en un momento específico o durante un período determinado.

***Screen-Film Mammography***: Es un sistema de mamografía de pantalla-película en la que se emiten rayos X a través de la mama hacia un casete que contiene una pantalla y

una película que debe revelarse.

***Sobrevida:*** Se refiere al tiempo que una persona vive después de haber sido diagnosticada o tratado por una enfermedad. Se utiliza para evaluar la eficacia de los tratamientos.

***Transfer Learning*** (Aprendizaje por transferencia): es una técnica de aprendizaje automático en la que el conocimiento adquirido a través de una tarea o un conjunto de datos se utiliza para mejorar el rendimiento del modelo en otra tarea relacionada o en un conjunto de datos diferente.

# Índice General

<b>1. Introducción</b>	<b>1</b>
<b>2. Estado del Arte</b>	<b>5</b>
<b>3. Diseño de la Solución e Implementación del Módulo de Clasificación de Imágenes</b>	<b>13</b>
3.1. Contexto General del Prototipo . . . . .	13
3.2. Arquitectura Propuesta . . . . .	14
3.3. Adquisición de Datos . . . . .	15
3.3.1. Selección del Dataset . . . . .	15
3.3.2. Descripción del Dataset a Utilizar . . . . .	16
3.4. Preprocesamiento de los datos . . . . .	17
3.4.1. Exploración de los datos . . . . .	17
3.4.2. Subsets de datos a utilizar . . . . .	19
3.5. Arquitectura de los Clasificadores . . . . .	22
3.5.1. Clasificador Principal . . . . .	23
3.5.2. Clasificador de Malignidad de Masas . . . . .	24
3.5.3. Clasificador de Malignidad de Calcificaciones . . . . .	25
3.6. Ambiente de Desarrollo . . . . .	26
<b>4. Resultados</b>	<b>27</b>
4.1. Clasificador Principal . . . . .	27

4.2. Clasificador de Malignidad de Masas . . . . .	31
4.3. Clasificador de Malignidad de Calcificaciones . . . . .	35
<b>5. Factibilidad de la Solución</b>	<b>41</b>
<b>6. Conclusiones</b>	<b>45</b>
6.1. Trabajo Futuro . . . . .	46
<b>Anexos</b>	<b>47</b>
<b>Anexo A: Descripción de las Métricas de Evaluación</b>	<b>47</b>

# Índice de Figuras

2.1. Metodología propuesta en [18] para la segmentación de masas. . . . .	7
2.2. Diagrama esquemático de la arquitectura propuesta en [19] . . . . .	8
2.3. Descripción general del modelo Vision Transformer introducida en [21].	9
3.1. Arquitectura del sistema “Smart Mammo” . . . . .	14
3.2. Ejemplo de Windowing Operation [28] . . . . .	19
3.3. Arquitectura de clasificación de una mamografía. . . . .	22
4.1. Matriz de confusión - Clasificador principal . . . . .	28
4.2. Matriz de confusión normalizada - Clasificador principal . . . . .	29
4.3. Ejemplo de predicción del modelo sobre el conjunto de test. <b>Predicción:</b> <i>Masas</i> (99.99 %) <b>Etiqueta real:</b> <i>Masas</i> . Se observa que el modelo concentra su atención en el tejido mamario, pero no particularmente en donde se encuentra la masa, sino que al rededor de esta. . . . .	30
4.4. Ejemplo de predicción del modelo sobre el conjunto de test. <b>Predicción:</b> <i>Calcificaciones</i> (72.39 %) <b>Etiqueta real:</b> <i>Sin Hallaz-</i> <i>gos</i> . En este caso el modelo centra la atención en el interior del seno, sin enfocarse en un área en particular. . . . .	30
4.5. Matriz de confusión - Clasificador Masas . . . . .	32
4.6. Matriz de confusión normalizada - Clasificador Masas . . . . .	33

4.7.	Ejemplo de predicción del modelo sobre el conjunto de test. <b>Predicción:</b> <i>Benigno</i> (99.99 %) <b>Etiqueta real:</b> <i>Benigno</i> . El modelo centra la atención en la zona interior del seno, sin enfocarse particularmente en la lesión presente. . . . .	34
4.8.	Ejemplo de predicción del modelo sobre el conjunto de test. <b>Predicción:</b> <i>Benigno</i> (99.88 %) <b>Etiqueta real:</b> <i>Sospechoso</i> . Se observa que las regiones que presentan un nivel de atención mayor se encuentran al rededor de la masa y no precisamente sobre esta. . . . .	34
4.9.	Matriz de confusión - Clasificador calcificaciones . . . . .	36
4.10.	Matriz de confusión normalizada - Clasificador calcificaciones . . . . .	37
4.11.	Ejemplo de predicción del modelo sobre el conjunto de test. <b>Predicción:</b> <i>Maligno</i> (100.00 %) <b>Etiqueta real:</b> <i>Maligno</i> . La anotación manual es extensa, abarcando una gran región del seno, esta se encuentra dentro de la zona de mayor atención del modelo. . . . .	38
4.12.	Ejemplo de predicción del modelo sobre el conjunto de test. <b>Predicción:</b> <i>Benigno</i> (99.99 %) <b>Etiqueta real:</b> <i>Benigno</i> . En este caso, la anotación es pequeña y las zonas con alta atención igual son de menor tamaño y se encuentran alrededor de la anotación real. . . . .	38
5.1.	Vista de revisión de múltiples imágenes usando Smart Mammo . . . . .	43
5.2.	Vista de un reporte generado por Smart Mammo . . . . .	43

# Índice de Tablas

2.1. Resumen de las arquitecturas y resultados reportados en los trabajos relacionados . . . . .	11
3.1. Principales Hallazgos presentes en el dataser VinDr-Mammo . . . . .	18
3.2. Distribución de BI-RADS para la categoría Masas . . . . .	18
3.3. Distribución de BI-RADS para la categoría Calcificaciones . . . . .	19
3.4. Subsets clasificador de hallazgos . . . . .	20
3.5. Subsets clasificador de masas . . . . .	20
3.6. Subsets clasificador de calcificaciones . . . . .	21
3.7. Conjunto de entrenamiento para el clasificador de hallazgos . . . . .	24
3.8. Conjunto de entrenamiento para el clasificador de malignidad correspondiente a la clase de masas . . . . .	25
3.9. Conjunto de entrenamiento para el clasificador de malignidad correspondiente a la clase de calcificaciones . . . . .	26
4.1. Métricas de rendimiento para el clasificador principal sobre los datos de test . . . . .	28
4.2. Métricas de rendimiento para el clasificador de malignidad de masas sobre los datos de test . . . . .	31
4.3. Métricas de rendimiento para el clasificador de malignidad de calcificaciones sobre los datos de test . . . . .	35



# Capítulo 1

## Introducción

El cáncer de mama es una enfermedad que consiste en que células alteradas presentes en la mama se multiplican sin control y forman tumores, que al no ser tratados, pueden propagarse por todo el cuerpo y causar la muerte. El desarrollo de las células cancerosas comienza dentro de los conductos galactóforos o de los lobulillos que producen leche del seno, las cuales pueden propagarse, invadiendo al tejido mamario cercano, produciendo nódulos o engrosamiento. El cáncer invasivo puede propagarse a los ganglios linfáticos cercanos o a otros órganos; esto se conoce como metástasis y pueden ser letales, poniendo en peligro la vida de la paciente [1].

Durante el año 2022, esta enfermedad fue la segunda causa principal de incidencia mundial de cáncer, con 2,3 millones de nuevos casos (11,6 % de todos los casos de cáncer). Si se considera solo la población femenina, el cáncer de seno es el diagnóstico de cáncer más común y la principal causa de muerte por cáncer a nivel mundial. El cáncer de mama representa casi uno de cada cuatro casos de cáncer y una de cada seis muertes por cáncer en mujeres en todo el mundo [2].

En Chile, el panorama del cáncer de mama no dista mucho de la escena global, siendo desde el 2009 el cáncer con mayor prevalencia, nivel de incidencia (20,8 %) y

mortalidad (12,5 %), transformándose en la principal causa de muerte en mujeres de 35 años y más en el país [3, 4, 5]. Cuando se compara al país con el resto de Latinoamérica, la tasa de mortalidad es menor, pero sigue siendo superior a la observada en regiones desarrolladas como Europa y Norteamérica [5].

El enfoque del control del cáncer de mama es aumentar el acceso al diagnóstico y lograr una detección temprana, junto con un manejo oportuno y completo del cáncer [2]. En este escenario, existe una sobrevida cercana al 99 % y en la mayoría de los casos permite cirugías menores y más conservadoras, con tratamientos menos agresivos [6]. Generalmente las personas no experimentan ningún síntoma cuando el cáncer está aún en fase temprana [1], por lo tanto, el mejor examen para una detección en etapa precoz es la mamografía, la cual permite encontrar cánceres pequeños, incluso antes de que sean palpables (menores a 5 mm.). De forma preventiva, se sugiere comenzar con mamografías anuales a partir de los 40 años en el caso de mujeres con riesgo promedio [6].

Los médicos utilizan un sistema estándar denominado *Breast Imaging Reporting and Data System* (BI-RADS) para describir los resultados y hallazgos en la revisión de mamografías. Este sistema clasifica cada examen en una categoría numérica de cero a siete: 0-se necesita evaluación adicional, 1-negativo, 2-benigno, 3-probablemente benigno, 4-anomalía sospechosa (esta categoría tiene un rango amplio de niveles de sospecha y suele ser dividida aún más: 4A-bajo, 4B-moderado, 4C-sospecha alta de malignidad), 5-hallazgo altamente probable de ser maligno y 6-malignidad confirmada mediante biopsia [7].

En el informe de mamografía también se incluye una evaluación de la densidad mamaria, que es una medida de la cantidad de tejido fibroso y glandular que hay en el seno, en comparación con el tejido graso. Esta se clasifica en 4 categorías: A-tejido adiposo casi en su totalidad, B-áreas dispersas de tejido fibroso y glandular, C-mayor parte del seno está compuesto de tejido glandular y fibroso denso (se describe como

heterogéneamente denso) y D-seno extremadamente denso. En general, se considera que las pacientes tienen tejido mamario denso cuando se encuentran en las categorías C y D. Una alta densidad mamaria se asocia a un riesgo más alto de padecer cáncer de mama y también dificulta la detección de tejido canceroso en las mamografías, ya que tanto este último como el tejido fibroso y glandular se ven blancos, mientras que el tejido adiposo se ve casi negro. Por lo tanto, la identificación de un tumor, que luce blanco, es más sencilla si la mayor parte del seno está compuesta por tejido adiposo [8].

Los protocolos actuales requieren evaluación del mamograma por dos radiólogos (y un tercero, si no se llega a consenso), lo cual evidencia el desafío que presenta, incluso para expertos, la labor de identificar anormalidades. La clasificación de masas es una de las tareas más difíciles, no solo por las amplias variaciones en tamaño y forma, sino que también por el bajo contraste que presentan las imágenes. Adicionalmente, las masas suelen estar rodeadas por otros elementos, como músculo, vasos sanguíneos y tejido normal [9]. De acuerdo a lo mencionado anteriormente, es posible que se produzcan diagnósticos erróneos, debido no solo al hecho de que leer una mamografía es desafiante, sino que también a causa del alto número de exámenes que deben ser revisados por los profesionales [10]. Esto se acentúa aún más en pacientes con tejido mamario denso, debido a la dificultad de identificar el tejido canceroso, lo cual es un factor relevante, ya que se estima que el 43 % (27,6 millones) de la población femenina entre 40 y 74 años en EE.UU. se encuentra en esta categoría [11] y se espera que la prevalencia de pacientes con senos densos se mantenga estable en el tiempo [12].

Para apoyar en el desafío de la evaluación radiológica de las mamografías, se han desarrollado sistemas de diagnóstico asistido por computador (CAD, por sus siglas en inglés) con herramientas para ayudar a los radiólogos en la detección y clasificación de lesiones mamarias [9]. Sin embargo, la clasificación de anormalidades como benignas o malignas es compleja y a menudo conduce a un alto número de Falsos Positivos, especialmente en sistemas CAD tradicionales (no utilizan técnicas de inteligencia arti-

ficial), lo cual conduce al paciente a escenarios de ansiedad, exposición innecesaria a radiación, biopsias sin sentido y mayor número de exámenes junto con mayores gastos médicos [9, 13]. Nuevos y más precisos sistemas han sido desarrollados aprovechando el avance de las técnicas de *machine learning* y especialmente de *deep learning*, algunas de estas propuestas son revisadas con mayor detalle en la sección 2.

En esta memoria, se desarrolla un modelo de deep learning, basado en *vision transformer* (ViT) y usando *transfer-learning*, para la clasificación de imágenes médicas de mamografías, siguiendo las categorías BI-RADS mencionadas anteriormente. En concreto, se hace uso de un dataset público para el entrenamiento independientes de tres clasificadores. El primero se encarga de clasificar si hay presencia de tejido canceroso en la imagen, obteniendo como resultado: Masas, Calcificaciones o Sin Hallazgos. Si el resultado corresponde a las dos primeras categorías, se utiliza un segundo clasificador, según corresponda a masa o calcificación, el que encasilla la imagen en los niveles de BI-RADS 3, 4 o 5, de acuerdo a la malignidad del hallazgo.

Este documento se estructura de la siguiente forma: En el Capítulo 2 se presenta la revisión de soluciones recientes basadas en CNN y ViT. El contexto general de la solución, es abordado en el Capítulo 3, junto con la implementación del módulo de clasificación de imágenes, incluyendo la adquisición y preprocesamiento de los datos y además, se presenta la configuración utilizada durante el entrenamiento de cada clasificador. En el Capítulo 4 se exponen los resultados conseguidos por cada clasificador, a través de matriz de confusión y métricas de desempeño. Discusiones sobre la factibilidad de la solución propuesta y el paradigma vision transformer se proveen en el Capítulo 5. Finalmente, en el Capítulo 6 se entregan las conclusiones del trabajo de memoria realizado y se identifican áreas de mejoras y trabajos futuros.

# Capítulo 2

## Estado del Arte

El desarrollo de sistemas CAD para la evaluación de mamografías es un amplio campo de investigación, donde han surgido distintos enfoques sobre como asistir a los radiólogos en el diagnóstico de lesiones mamarias. En [9] se realiza una revisión de la literatura, exponiendo las principales líneas de investigación. Los desarrollos pueden ser clasificados de acuerdo a su objetivo en 4 categorías, ya que se centran, o bien en la detección, o bien en la clasificación de 2 anomalías mamarias: masas o calcificaciones sospechosas. También es posible dividirlos de acuerdo a las técnicas de machine learning en que se basan, partiendo por sistemas que no utilizan machine learning, seguidos por los que se basan en características para entrenar modelos y finalmente arquitecturas más avanzadas de aprendizaje automático basados en deep learning, donde destacan las Redes Neuronales Convolucionales (CNN, por sus siglas en inglés). La introducción de esta familia de modelos más sofisticados movió el interés desde la extracción de características hacia la identificación, selección y clasificación automática de lesiones mamarias. A raíz de lo anterior, esta clase de modelos de deep learning, logran eliminar la intervención humana, implicando un *trade-off*, ya que logra un desempeño más robusto, permitiendo mejores resultados frente a datos más complejos, pero a su vez, introduce una mayor incertidumbre en el diagnóstico, reduciendo la explicabilidad de los resultados.

Dentro de las propuestas de modelos basados en CNN, Gengian et al. [14] plantea un enfoque de deep learning basado en EfficientNet para la clasificación de imágenes de mamografías, usando el dataset CBIS-DDSM [15] para el entrenamiento y validación de 5 modelos de EfficientNet con diferentes complejidades, en donde el modelo más complejo (b4) obtuvo el mejor desempeño (accuracy de 0.75 y AUC de 0.83). De esta forma se sugiere que modelos más complejos podrían obtener mejores resultados, pero también se produciría un trade-off entre desempeño y costo computacional. Un sistema CAD basado en *fully connected depthwise-separable CNN* fue introducido en [16] para la detección de micro calcificaciones sospechosas y su posterior clasificación en benignas o malignas. Para entrenar el modelo se usa un dataset privado y posteriormente es validado usando el dataset DDSM [17], obteniendo un accuracy de 0.87 y precisión de 0.85.

Otros desarrollos exploran técnicas alternativas para la detección de anomalías, como en Zeiser et al. [18], donde usaron 6 modelos basados en U-Net para la segmentación de masas. La metodología que se propone en este trabajo se muestra en la figura 2.1. El modelo más simple y que usaba menos datos presentó una alta sensibilidad, pero junto con una alta tasa de falsos positivos, lo que no es deseado en un entorno médico. El mejor resultado, con un accuracy de 0,86, lo obtuvo el modelo más complejo y también fue el que usó más datos. Se concluye que para mejorar los resultados, se requiere una mayor complejidad y un conjunto de datos más grande, además, se menciona que el uso de imágenes digitales por sobre imágenes digitalizadas, también puede contribuir a un mejor desempeño de los modelos.

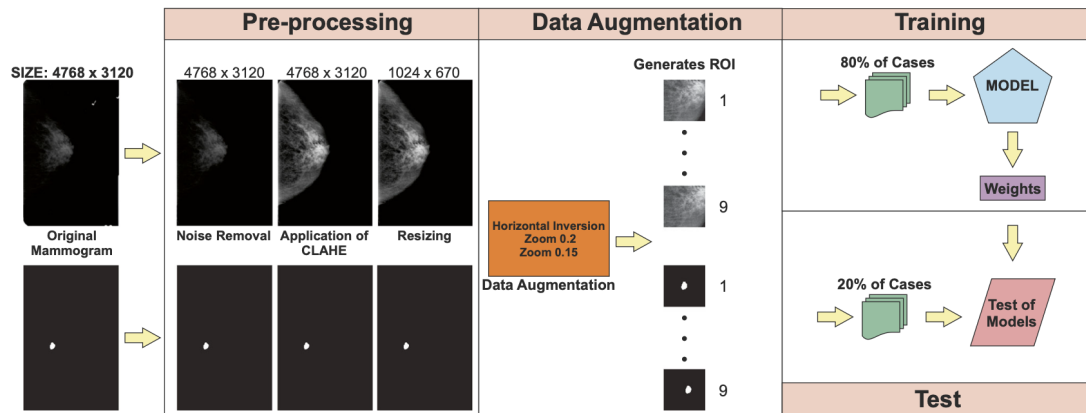


Figura 2.1: Metodología propuesta en [18] para la segmentación de masas.

En [19], se presenta un sistema integrado combinando la detección y clasificación. Para la detección se hace uso de YOLO 9000 y para la clasificación se prueban 3 modelos de deep learning basados en CNN: *regular feedforward CNN*, *ResNet-50* y *InceptionResNet-V2*, usando los datasets INbreast [20] y DDSM. La arquitectura completa se presenta en la figura 2.2. Los autores hacen uso de *data augmentation* y transfer learning para solventar el problema de la falta de datos médicos anotados. La clasificación de las lesiones se enfoca en dos clases, benignas y malignas. Para la evaluación de los modelos se siguió la estrategia de 5-fold (5 subconjuntos de test, cada uno con 60 imágenes), en donde el resultado final corresponde al promedio de las cinco evaluaciones. La arquitectura basada en InceptionResNet-v2 consiguió los mejores resultados con un accuracy de 0.97 y 0.95 sobre DDSM e INbreast, respectivamente.

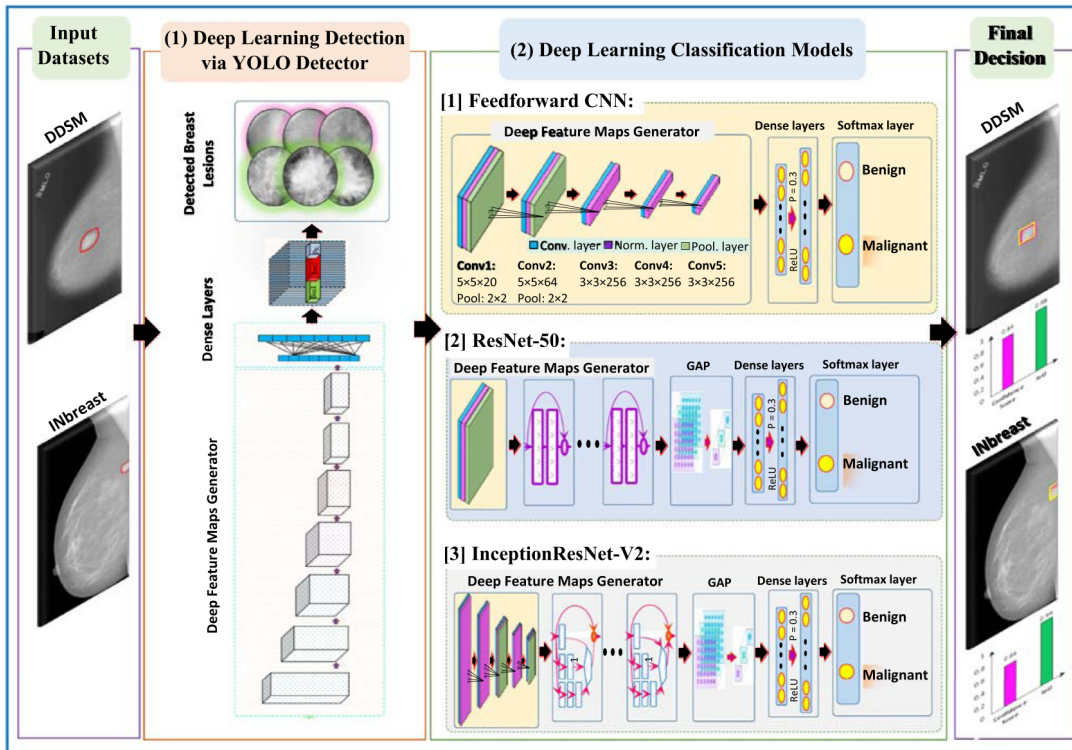


Figura 2.2: Diagrama esquemático de la arquitectura propuesta en [19]

Otras alternativas a las CNN han surgido, en [21] se propone el uso de arquitectura transformer, usada principalmente en el campo del procesamiento de lenguaje natural, pero ahora aplicada en el área de visión por computador. El modelo presentado se denomina Vision Transformer (ViT), donde se entrena un encoder Transformer en ImageNet, obteniendo resultados comparables a los obtenidos por los modelos más actuales de CNN, pero con un costo computacional significativamente menor. El método propuesto (figura 2.3) consiste en transformar la imagen 2D en una secuencia de parches 2D aplanados, es decir vectores y luego a cada uno de estos se le añade su posición y se continúa con el flujo tradicional de un transformer. El modelo fue pre-entrenado en grandes conjuntos de datos y luego ajustado a tareas posteriores más pequeñas. El ajuste fino (fine tuning) se hizo eliminando el cabezal de predicción preentrenado y reemplazándolo con una capa de avance inicializada cero.

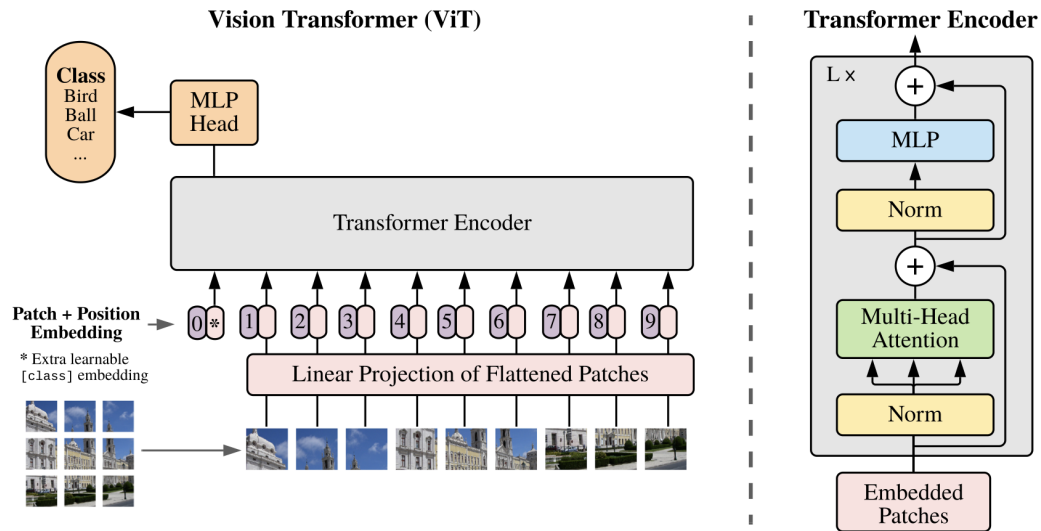


Figura 2.3: Descripción general del modelo Vision Transformer introducida en [21].

Desde la introducción del modelo original de vision transformer en 2021, el uso de esta arquitectura se ha popularizado en diferentes áreas y la tarea de detección temprana de cáncer de mama no está exenta. En 2023, Ayana et al. [13] propuso un enfoque de transfer learning basado en ViT para la clasificación de mamografías, evaluando el desempeño y los requerimientos computacionales de diferentes modelos y sus variantes. Se utilizaron los modelos previamente entrenados en ImageNet como punto de partida y luego fueron entrenados para la clasificación de 2 clases: Benignas y Malignas. Se usó una versión aumentada del dataset DDSM [22]. Se evaluó el método propuesto utilizando 5 configuraciones experimentales: 1. comparación de los diferentes modelos de vision transformer utilizando transfer learning, 2. comparación de los modelos entrenándolos desde cero, comparados con los que utilizaron transfer learning, 3. comparación con modelos que utilizaron transfer learning con arquitecturas de CNN, 4. se evaluó el costo computacional de cada modelo y 5. se comparó el rendimiento del método propuesto con métodos anteriores que usaron el mismo dataset. Se utilizaron diferentes métricas como Accuracy, AUC, F1 Score, entre otras, con un intervalo de confianza del 95 %, obteniendo un resultado  $1 \pm 0$  en cada métrica, con cada modelo de ViT que utilizó transfer learning; en el caso de los modelos entrenados desde cero tu-

vieron un desempeño menor con métricas por debajo del 80 %. Finalmente se concluye que los modelos basados en vision transformer superaron a los basados en CNN y también los modelos con un costo computacional significativamente menor, obtuvieron un rendimiento igual al de los modelos más complejos. De esta forma se demuestra que el enfoque de transfer learning basado en vision transformer es efectivo para la clasificación de imágenes de mamografías, proporcionando un desempeño superior con menor costo computacional.

Para facilitar el trabajo con modelos basados en transformers han surgido soluciones como *Hugging Face*, una plataforma de acceso abierto e integral compuesta por repositorios de datasets y modelos pre-entrenados. Además, ofrece un framework de alto nivel que brinda acceso a diversas bibliotecas y herramientas que cubren todo el ciclo de desarrollo de modelos, desde la gestión de datos, entrenamiento y la evaluación.

Para el desarrollo de esta propuesta, algunos de los modelos presentes en Hugging Face fueron seleccionados como punto de partida para el entrenamiento de los clasificadores. En particular se revisan dos modelos, el primero es GOOGLE/VIT-BASE-PATCH16-224 [23], que cuenta con 12 capas o bloques transformer. El segundo modelo es GOOGLE/VIT-LARGE-PATCH16-224 [24] el cual es una versión más compleja que el anterior, con 24 bloques transformer. Ambos modelos fueron entrenados en ImageNet-21k a una resolución de 224x224 pixels y luego fine-tuned en ImageNet (ILSVRC2012) a la misma resolución.

En la tabla 2.1 se presenta un resumen de los trabajos revisados en este capítulo que abordan la aplicación de modelos de Machine Learning para la evaluación y análisis de imágenes de mamografías.

Tabla 2.1: Resumen de las arquitecturas y resultados reportados en los trabajos relacionados

<b>Paper</b>	<b>Año</b>	<b>Arquitectura de ML</b>	<b>Objetivo</b>	<b>Resultado</b>	<b>Dataset</b>
Gengtian et al.[14]	2023	EfficientNet (CNN)	Clasificación, sin especificar	Accuracy 0.75; AUC 0.83	CBIS-DDSM
Rehman et al.[16]	2021	Fully connected depthwise-separable CNN	Detección y clasificación de microcalcificaciones	Accuracy 0.87; precisión 0.85	DDSM
Zeiser et al.[18]	2020	U-Net (CNN)	Segmentación de masas	Accuracy 0.86; AUC 86.40	DDSM
Mugahad et al.[19]	2020	Regular feedforward CNN ResNet-50 InceptionResNet-V2	Detección y clasificación binaria de lesiones mamarias	InceptionResNet-V2: accuracy 0.97 sobre DDSM	INbreast; DDSM
Ayana et al.[13]	2023	Vision Transformer (ViT)	Clasificación binaria de masas	Accuracy, AUC y precisión $1 \pm 0$ (IC 95 %)	Versión aumentada de DDSM



# Capítulo 3

## Diseño de la Solución e Implementación del Módulo de Clasificación de Imágenes

### 3.1. Contexto General del Prototipo

El prototipo funcional del sistema de clasificación de imágenes de mamografías se desarrolla con el objetivo de proporcionar una solución eficaz y automática, apoyando a los profesionales de la salud en el diagnóstico temprano de tejido canceroso. Esto se logra al combinar tecnologías de aprendizaje profundo, procesamiento de imágenes y visualización interactiva, logrando así, una solución integral.

El diseño del prototipo se centra en la implementación de una herramienta que permite la carga, análisis y clasificación de mamografías haciendo uso de modelos basados en vision transformer ya entrenados. Adicionalmente, se proporciona una interfaz interactiva la cual mejora la experiencia del usuario a través de una navegación clara e intuitiva. La aplicación está orientada a ser utilizada tanto por radiólogos como por otros especialistas de salud para complementar su diagnóstico, aumentando la fiabili-

dad y rapidez en el proceso de análisis.

## 3.2. Arquitectura Propuesta

El prototipo funcional denominado como “**Smart Mammo**” está compuesto por diferentes módulos que interactúan entre sí para completar el flujo de trabajo: carga, análisis y reporte de resultados, tal como se visualiza en la figura 3.1. Cada uno se ha diseñado para realizar una función específica y garantizar la modularidad y eficiencia del sistema. A continuación, se presentan los componentes clave del prototipo, los cuales se integran para dar lugar a un sistema cohesivo y robusto.

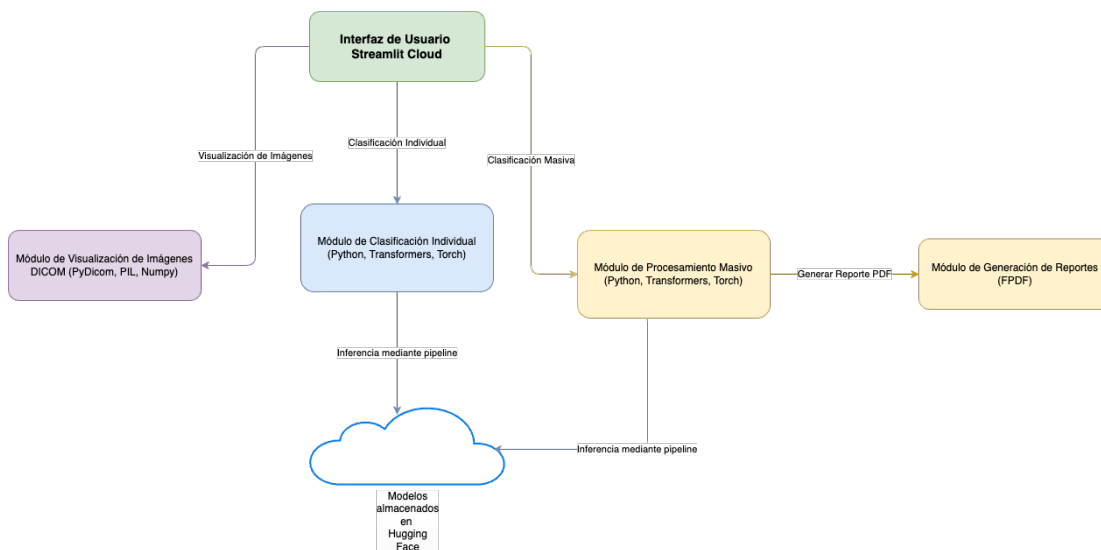


Figura 3.1: Arquitectura del sistema “Smart Mammo”

El esquema general se compone de los siguientes módulos:

- **Frontend - Interfaz de Usuario:** Construido en **Streamlit**, proporciona la capacidad de interactuar con el sistema a través de carga de imágenes, visualización, opciones de clasificación y generación de reportes.
- **Modelos de Clasificación:** Incluye los modelos alojados en **Hugging Face**, que permiten la inferencia y clasificación de las mamografías en masas, calcificaciones y demás categorías.

- **Módulo de Procesamiento Masivo:** Gestiona la clasificación de lotes grandes de imágenes, utilizando técnicas optimizadas para minimizar el tiempo de inferencia.
- **Módulo de Generación de Reportes PDF:** Encargado de consolidar los resultados de las clasificaciones y generar un reporte comprensible para los usuarios.
- **Backend - Servidor en la Nube:** Implementado mediante **Streamlit Cloud** para alojar la interfaz y gestionar la lógica del negocio.

Esta memoria se centra en el módulo de los modelos de clasificación, analizando su diseño, implementación y resultados obtenidos.

### 3.3. Adquisición de Datos

Para el desarrollo de soluciones basadas en modelos de machine learning, el contar con grandes cantidades de datos es indispensable para obtener buenos resultados. Sin embargo, en el área médica la adquisición de datos es compleja debido a la naturaleza de las imágenes y el resguardo a la privacidad de los pacientes. Actualmente existen pocos datasets de mamografías de acceso público, entre los que destacan: *Curated Breast Imaging Subset of Digital Database for Screening Mammography* (CBIS-DDSM) [15], *Mammographic Image Analysis Society* (MIAS) [25], INbreast [20] y VinDr-Mammo [26].

#### 3.3.1. Selección del Dataset

De los anteriores, CBIS-DDSM y MIAS son datasets conformados por *screen-film mammography* (SFM) digitalizadas, las cuales actualmente han sido remplazadas en entornos médicos por imágenes del tipo *full-field digital mammography* (FFDM), por lo cual, han sido descartados para ser utilizados. Pese a que INbreast está compuesto

por FFDM, tiene un tamaño de datos limitados, lo que puede comprometer el rendimiento de un sistema basado en deep learning.

Por tanto se selecciona el dataset **VinDr-Mammo: A large-scale benchmark dataset for computer-aided detection and diagnosis in full-field digital mammography** como dataset principal para el desarrollo y entrenamiento de un modelo de machine learning para clasificar imágenes de mamografías.

### 3.3.2. Descripción del Dataset a Utilizar

VinDr-Mammo introduce un dataset FFDM de gran tamaño conformado por 5.000 exámenes. Cada examen consta de 4 imágenes, ya que por cada seno se obtienen 2 vistas, craneocaudal (CC) y mediolateral oblique (MLO), de esta forma se obtiene un total de 20.000 imágenes de mamografías en formato DICOM (Digital Imaging and Communication in Medicine), el cual es un estándar de transmisión tanto de imágenes médicas como de los metadatos asociados a esa imagen.

El resultado de lectura de cada imagen incluye tanto la evaluación general de la mama, como información sobre regiones anormales. El resultado es informado siguiendo el esquema y léxico BI-RADS [7]. En términos de la evaluación general, se proporcionan las categorías de evaluación, incluyendo los niveles del 1 al 5 y los niveles de densidad mamaria (A, B, C o D). En el caso de los hallazgos con categoría BI-RADS 3, 4 o 5, se incluyen anotaciones del tipo bounding box.

En el proceso de lectura participaron tres radiólogos con más de diez años de experiencia en evaluación de mamografías. El procedimiento de lectura siguió la directriz europea [27] de que cada examen fue leído dos veces por dos radiólogos de forma independiente. Si existe alguna discrepancia entre los dos radiólogos, se resolverá mediante arbitraje con la participación del tercer radiólogo.

## Descripción de los Datos:

El dataset está conformado por las imágenes y 3 archivos CSV:

- Carpeta de imágenes: Contiene 5.000 subcarpetas correspondientes a los exámenes que conforman el dataset. Cada una de estas contiene 4 archivos DICOM, correspondientes a las 2 vistas estándar de cada seno.
- breast-level\_annotations.csv: Cada fila corresponde a una imagen y proporciona la evaluación BI-RADS del seno representado por la imagen junto con algunos metadatos de la imagen.
- finding\_annotations.csv: Cada fila representa una anotación de un hallazgo anormal en el seno. También se proporciona metadatos asociados a la imagen, la categoría del hallazgo junto con su evaluación BI-RADS, las coordenadas de la bounding box, entre otros.
- metadata.csv: Este archivo contiene información proporcionada por las etiquetas de los archivos DICOM, como la edad del paciente, modelo y fabricante del mamógrafo, etc.

## 3.4. Preprocesamiento de los datos

### 3.4.1. Exploración de los datos

Se comienza revisando los diferentes tipos de hallazgos presentes en el dataset VinDr-Mammo, encontrando un total de 35 diferentes categorías, sin embargo, la mayoría no supera los 50 ejemplares. Las principales categorías de hallazgos se presentan en la tabla 3.1.

Tabla 3.1: Principales Hallazgos presentes en el dataser VinDr-Mammo

<b>Hallazgo</b>	<b>Ejemplares</b>
No Finding	18232
Mass	1123
Suspicious Calcification	402
Focal Asymmetry	232
Architectural Distortion	95
Asymmetry	90
Suspicious Calcification, Mass	82
Suspicious Lymph Node	57
Skin Thickening	38
Suspicious Calcification, Focal Asymmetry	31

Debido a lo mostrado anteriormente, solo se seleccionan las tres principales categorías para trabajar, las que corresponden a sin hallazgos, masas y calcificaciones. A continuación, en las tablas 3.2 y 3.3 se revisa la distribución de BI-RADS para las categorías de Masas y Calcificaciones respectivamente.

Tabla 3.2: Distribución de BI-RADS para la categoría Masas

<b>Nivel BI-RADS</b>	<b>Ejemplares</b>
BI-RADS 3	566
BI-RADS 4	453
BI-RADS 5	104

Tabla 3.3: Distribución de BI-RADS para la categoría Calcificaciones

Nivel BI-RADS	Ejemplares
BI-RADS 3	284
BI-RADS 4	65
BI-RADS 5	53

### 3.4.2. Subsets de datos a utilizar

Dado que las imágenes se presentan en archivos DICOM, es necesario convertirlas a un formato que pueda ser utilizado como entrada de una red neuronal como lo son los formatos PNG o JPG. Este proceso no se realiza de forma directa ya que en los archivos DICOM el rango de valores de los pixels va desde los 10 a los 16 bits, mientras que las pantallas suelen tener valores de 8 bits, es decir, solo se dispone de 256 tonos de grises. Esto provoca que se deba hacer un mapeo *many-to-one* de los valores de los pixels. Este proceso se conoce como *Windowing Operation* y se utilizan dos valores denominados *Window Center (WC)* y *Window Width (WW)*, los cuales indican como se realiza el mapeo, tal como se ejemplifica en la figura 3.2

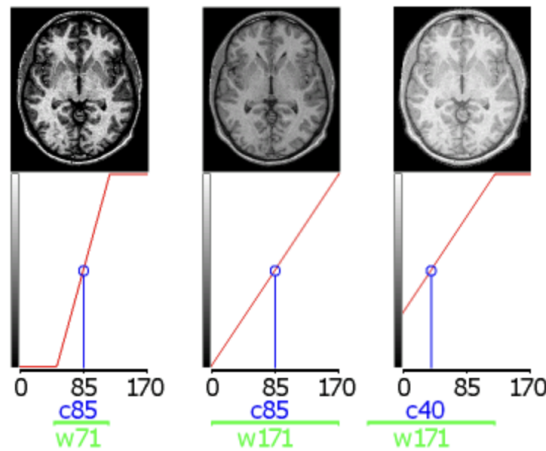


Figura 3.2: Ejemplo de Windowing Operation [28]

Una vez que las imágenes se encuentran en un formato compatible con el entrena-

miento de redes neuronales, se procede a crear 3 subsets de datos que son utilizados para entrenar cada uno de los tres clasificadores que se presentan en la sección 3.5. Debido a que la cantidad de imágenes de cada clase es dispar y también para aumentar la diversidad del conjunto de datos, se realizó un proceso de balanceo de clases utilizando data augmentation. Este procedimiento permite generar nuevas muestras sintéticas a partir de las imágenes originales presentes en el dataset, reduciendo el riesgo de sobreajuste y mejorando la robustez del modelo frente a diferentes escenarios reales. Entre las transformaciones implementadas se incluyen rotaciones aleatorias, volteos horizontales y verticales, así como también, ajustes de brillo, contraste, saturación, junto con inyección de ruido y desenfoque.

A continuación en las tablas 3.4, 3.5 y 3.3 se describe cada uno de los subsets resultantes:

Tabla 3.4: Subsets clasificador de hallazgos

<b>Clase</b>	<b>Imágenes</b>
Masas	2.502
Calcificaciones	2.502
Sin hallazgos	2.500
<b>Total</b>	<b>7.504</b>

Tabla 3.5: Subsets clasificador de masas

<b>Clase</b>	<b>Imágenes</b>
Beningno	834
Maligno	834
Sospechoso	834
<b>Total</b>	<b>2.502</b>

Tabla 3.6: Subsets clasificador de calcificaciones

<b>Clase</b>	<b>Imágenes</b>
Beningno	834
Maligno	834
Sospechoso	834
<b>Total</b>	<b>2.502</b>

### 3.5. Arquitectura de los Clasificadores

De acuerdo a lo revisado en la sección 2, se elige utilizar una arquitectura basada en Vision Transformer para desarrollar un modelo de deep learning que se encargue de clasificar las imágenes de mamografía. Sin embargo, al considerar la complejidad de las imágenes médicas, se hace inviable usar un único modelo que sea capaz de clasificar según el hallazgo y su malignidad, por eso se decide implementar tres modelos de clasificadores como se presenta en la figura 3.3.

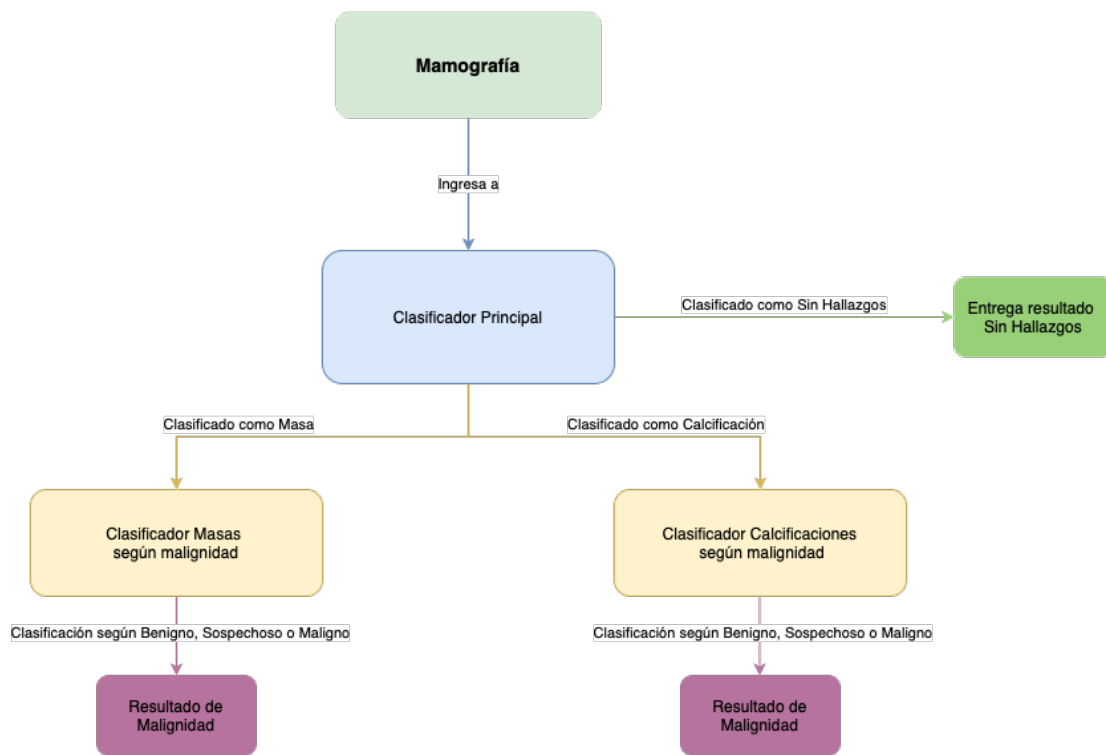


Figura 3.3: Arquitectura de clasificación de una mamografía.

El flujo es el siguiente:

1. La mamografía ingresa al clasificador principal, el que se encarga de clasificar la mamografía en 3 clases: hallazgo de masa, hallazgo de calcificación, sin hallazgo.
2. Si la mamografía se encuentra en alguna de las 2 primeras clases pasa a uno de los clasificadores secundarios dependiendo de la clase. Estos se encargan de

clasificar el hallazgo según su malignidad en alguna de las siguientes categorías:

- **Benigno:** probabilidad no mayor al 2 % de ser maligno. Corresponde al nivel BI-RADS 3
- **Sospechoso:** probabilidad mayor al 2 %, pero menor al 95 % de ser maligno. Corresponde al nivel BI-RADS 4
- **Maligno:** probabilidad mayor al 95 % de ser maligno, se recomienda realización de una biopsia. Corresponde al nivel BI-RADS 5

En las siguientes subsecciones se detalla la arquitectura del modelo junto con la configuración utilizada para cada clasificador.

### 3.5.1. Clasificador Principal

Para este clasificador, mediante transfer learning se utiliza el modelo GOOGLE/VIT-BASE-PATCH16-224 disponible en *Hugging Face*, este modelo cuenta con 12 capas o bloques transformer, de los cuales se descongelan los últimos 7 para el proceso de entrenamiento, de esta forma se obtienen 49.617.411 parámetros entrenables de un total de 85.800.963.

Para entrenar el modelo se usa como base el subset presentado en la tabla 3.4, pero se vuelve a usar data augmentation para mejorar el rendimiento y la generalización del modelo. El conjunto final de datos a utilizar se describe en la tabla 3.7, el cual consiste en un total de 15.008 imágenes, dejando el 80 % para training, 10 % para validation y 10 % para testing.

Tabla 3.7: Conjunto de entrenamiento para el clasificador de hallazgos

<b>División</b>	<b>Calcificaciones</b>	<b>Masas</b>	<b>Sin Hallazgos</b>	<b>Total</b>
Train	3.930	4.034	4.042	12.006
Validation	518	490	492	1.500
Test(Evaluation)	556	480	466	1.502
<b>Total</b>	<b>5.004</b>	<b>5.004</b>	<b>5.000</b>	<b>15.008</b>

La configuración de hiperparámetros utilizada es la siguiente:

- Learning rate:  $1 \cdot 10^{-4}$ .
- Batch size: 32.
- Epochs: 30.
- Dropout: Clasificador=0,4; Capas Ocultas=0,3; Capas de atención=0,3.
- Loss function: Weighted CrossEntropyLoss (Calcificaciones=1,0; Masas=1,5; SinHallazgos=1.0).

### 3.5.2. Clasificador de Malignidad de Masas

En cuanto a este clasificador, siguiendo la estrategia anterior, se utiliza el modelo GOOGLE/VIT-BASE-PATCH16-224 antes mencionado. Para la fase de entrenamiento nuevamente se descongelan las últimas 7 capas, pero esta vez no se realiza un aumento adicional de datos, sino que se ocupan los ya disponibles en el subset descrito en la tabla 3.5.

De esta forma se cuenta con un total de 2.502 imágenes en el conjunto final de datos, con una división del 80 % para training, 10 % para validation y 10 % para testing. En la tabla 3.8 se detalla la distribución por clase.

Tabla 3.8: Conjunto de entrenamiento para el clasificador de malignidad correspondiente a la clase de masas

División	Benigno	Maligno	Sospechoso	Total
Train	672	674	655	2.001
Validation	87	83	80	250
Test(Evaluation)	75	77	99	251
<b>Total</b>	834	834	834	2.502

La configuración de hiperparámetros utilizada es la siguiente:

- Learning rate:  $1 \cdot 10^{-4}$ .
- Batch size: 16.
- Epochs: 30.
- Dropout: Clasificador=0,3; Capas Ocultas=0,2; Capas de atención=0,2.
- Loss function: CrossEntropyLoss.

### 3.5.3. Clasificador de Malignidad de Calcificaciones

Al igual que los anteriores clasificadores se vuelve a usar transfer learning, pero en este caso se selecciona el modelo GOOGLE/VIT-LARGE-PATCH16-224, disponible igualmente en *Hugging Face*. El modelo cuenta con 24 capas o bloques transformer, de los cuales se descongelan los últimos 5 para el proceso de entrenamiento, de esta forma se obtienen 62.984.195 parámetros entrenables de un total de 303.304.707.

Para el entrenamiento de este modelo se usa directamente el subset presentado en la tabla 3.6, sin un aumento extra de datos. El conjunto final de datos a utilizar se describe en la tabla 3.9, el cual consiste en un total de 2.502 imágenes, ocupando el 80 % para training, 10 % para validation y 10 % para testing.

Tabla 3.9: Conjunto de entrenamiento para el clasificador de malignidad correspondiente a la clase de calcificaciones

<b>División</b>	<b>Benigno</b>	<b>Maligno</b>	<b>Sospechoso</b>	<b>Total</b>
Train	684	638	679	2.001
Validation	80	93	77	250
Test(Evaluation)	70	103	78	251
<b>Total</b>	834	834	834	2.502

La configuración de hiperparámetros utilizada es la siguiente:

- Learning rate:  $3 \cdot 10^{-4}$ .
- Batch size: 16.
- Epochs: 30.
- Dropout: Clasificador=0,0; Capas Ocultas=0,0; Capas de atención=0,0.
- Loss function: CrossEntropyLoss.

### 3.6. Ambiente de Desarrollo

Para el desarrollo del módulo de clasificación, es decir, el entrenamiento y evaluación de los modelos de ViT, se usó un computador con las siguientes especificaciones de hardware: Procesador Apple Silicon M1 Pro (10-core CPU y 16-core GPU MPS), con 16 GB de memoria RAM y 1 TB de almacenamiento SSD.

El sistema fue implementado utilizando Python (3.11.6), junto con la biblioteca de alto nivel de modelos pre-entrenados Transformers (4.45.1) de Hugging Face y PyTorch (2.4.1) como backend. Adicionalmente, para el manejo de imágenes y metadatos se usa Pydicom (3.0.1), Open CV (4.10), NumPy (2.1.1) y Pandas (2.2.3).

# Capítulo 4

## Resultados

En esta sección se presentan los resultados obtenidos por cada clasificador sobre el conjunto de test del dataset utilizado. Este conjunto no es utilizado durante el proceso de entrenamiento, lo cual es útil para saber la capacidad del modelo de generalizar sobre imágenes no vistas.

Para analizar el rendimiento de los modelos, las siguientes métricas son utilizadas: Accuracy, Precision, Recall y F1 Score (más detalles y cómo se calculan en el Anexo A). Para las últimas 3 métricas se obtiene un valor promedio utilizando *Macro-averaging*, el cual consiste en obtener la métrica de cada clase y luego calcular la media aritmética de todas las clases. Por lo tanto, se otorga el mismo peso a cada clase, independientemente de la cantidad de instancias.

### 4.1. Clasificador Principal

Después del proceso de entrenamiento se seleccionó el mejor modelo utilizando como criterio la mejor *precision* alcanzada, debido a que en un contexto médico se necesita que los resultados presenten una alta certeza, es decir, con la menor cantidad de falsos positivos, ya que esto puede traducirse en tratamientos invasivos y riesgosos que son innecesarios. En este caso, corresponde al obtenido en la epoch número 26. Las

métricas de rendimiento sobre el conjunto de test se presentan en la tabla 4.1 y la matriz de confusión junto con su versión normalizada en las figuras 4.1 y 4.2 respectivamente.

Tabla 4.1: Métricas de rendimiento para el clasificador principal sobre los datos de test

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>
75,4 %	75,2 %	75,3 %	74,9 %

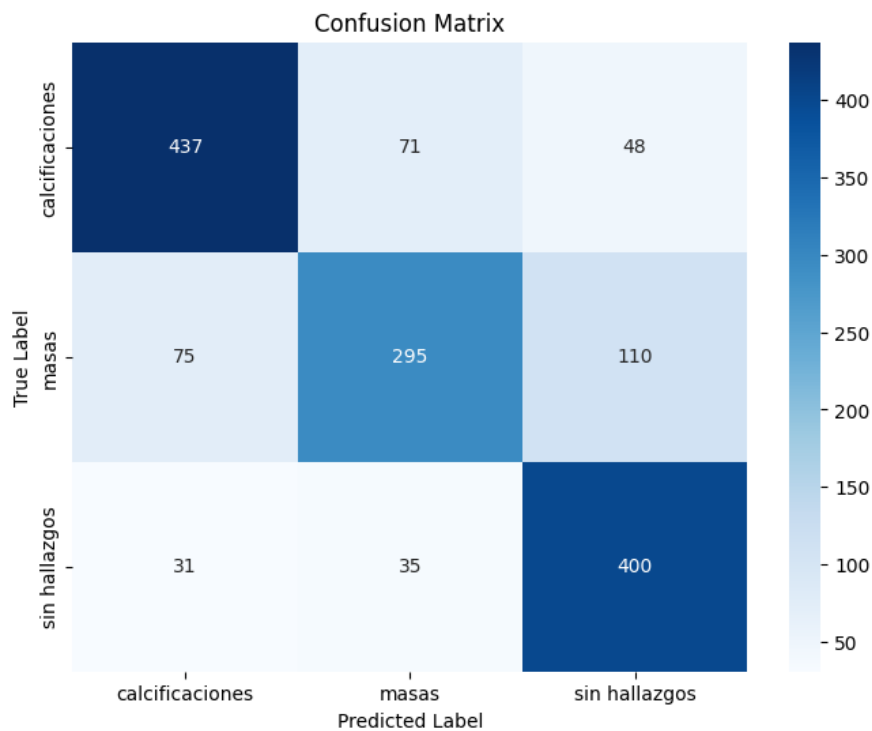


Figura 4.1: Matriz de confusión - Clasificador principal

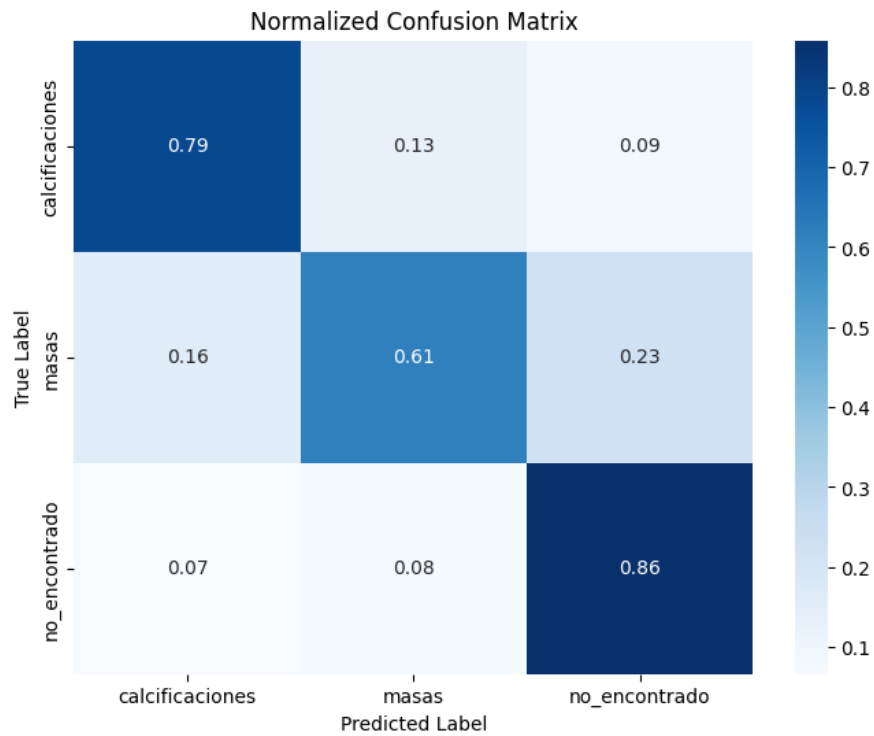


Figura 4.2: Matriz de confusión normalizada - Clasificador principal

De forma complementaria, con el fin de interpretar el comportamiento del modelo y analizar su capacidad para identificar regiones relevantes en las imágenes, se presentan algunos ejemplos cualitativos del conjunto de test: se muestra la imagen original junto con la etiqueta verdadera y la predicción obtenida, además del modelo de atención generado por el modelo, permitiendo evaluar la coherencia entre las áreas destacadas por el modelo y las lesiones realmente presentes.

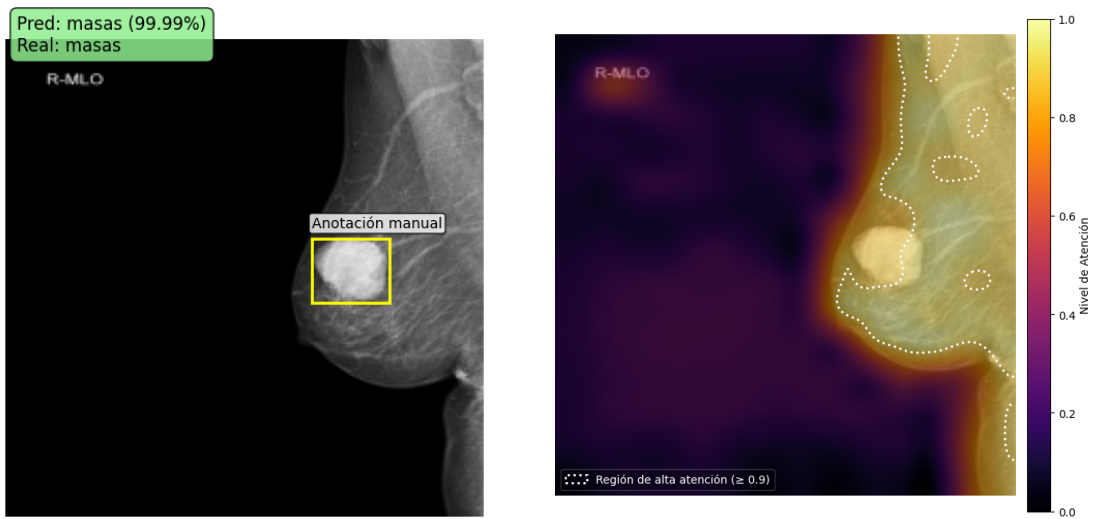


Imagen original.

Mapa de atención del modelo.

Figura 4.3: Ejemplo de predicción del modelo sobre el conjunto de test.

**Predicción:** *Masas* (99.99 %) **Etiqueta real:** *Masas*.

Se observa que el modelo concentra su atención en el tejido mamario, pero no particularmente en donde se encuentra la masa, sino que al rededor de esta.

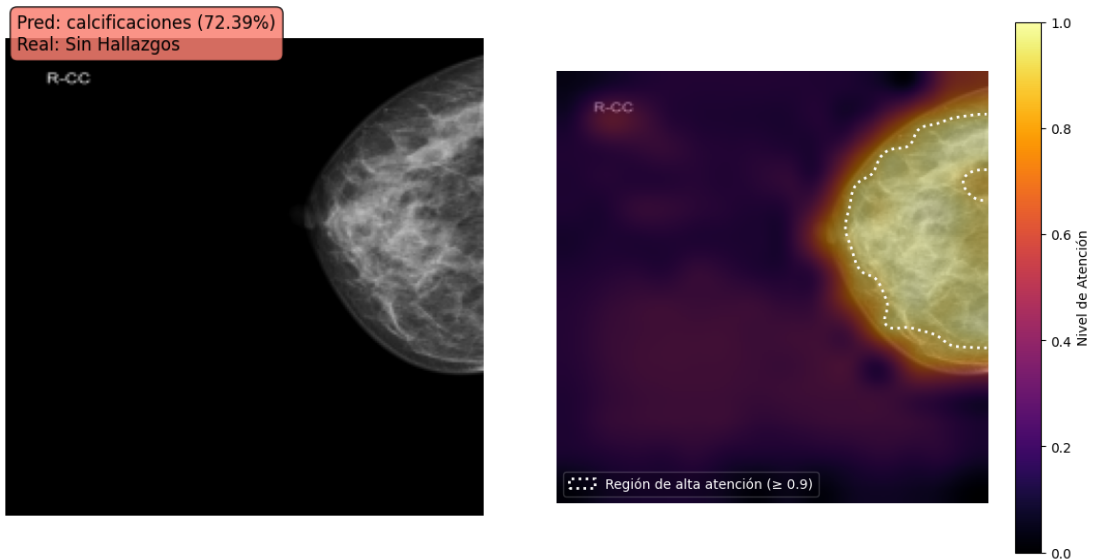


Imagen original con predicción y etiqueta real.

Mapa de atención del modelo.

Figura 4.4: Ejemplo de predicción del modelo sobre el conjunto de test.

**Predicción:** *Calcificaciones* (72.39 %) **Etiqueta real:** *Sin Hallazgos*.

En este caso el modelo centra la atención en el interior del seno, sin enfocarse en un área en particular.

De lo anterior se evidencia que el modelo tiene un buen desempeño general al clasificar las mamografías de acuerdo a los hallazgos presentes, especialmente en diferenciar aquellas sin hallazgos con respecto a las que presentan alguna lesión. La mayor dificultad se presenta con la clase de masas, la que tiene un nivel inferior de verdaderos positivos, lo que reduce el nivel de confiabilidad del clasificador.

## 4.2. Clasificador de Malignidad de Masas

Siguiendo la metodología anterior, se selecciona al mejor modelo correspondiente al resultante en la epoch número 29. De manera similar, las métricas de desempeño se presentan en la Tabla 4.2, acompañadas por la matriz de confusión mostrada en la Figura 4.5 y su representación normalizada en la Figura 4.6.

Tabla 4.2: Métricas de rendimiento para el clasificador de malignidad de masas sobre los datos de test

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>
89,2 %	89,2 %	89,7 %	89,2 %

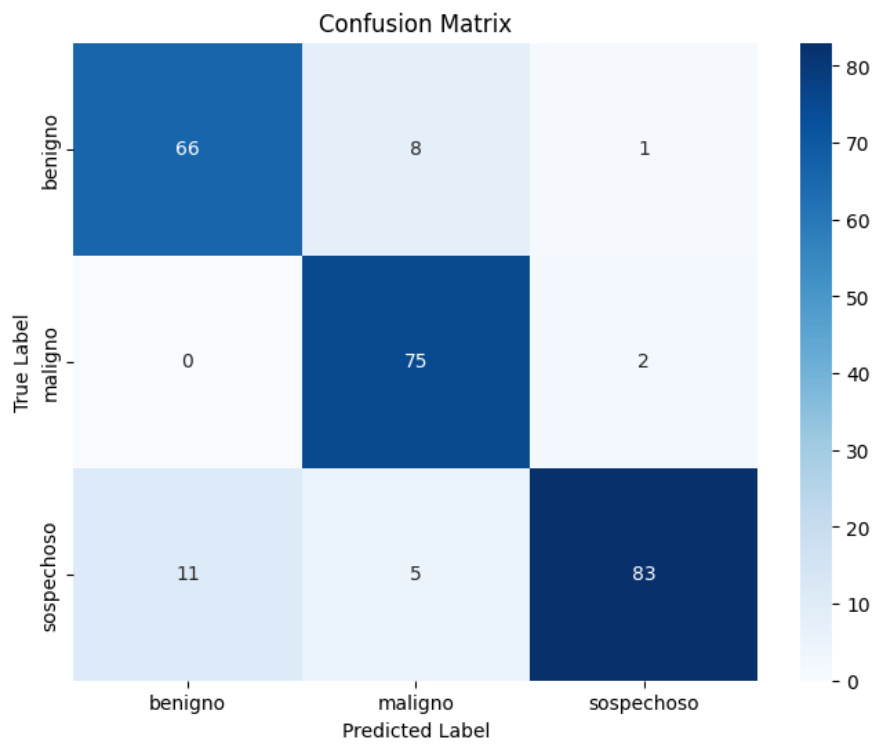


Figura 4.5: Matriz de confusión - Clasificador Masas

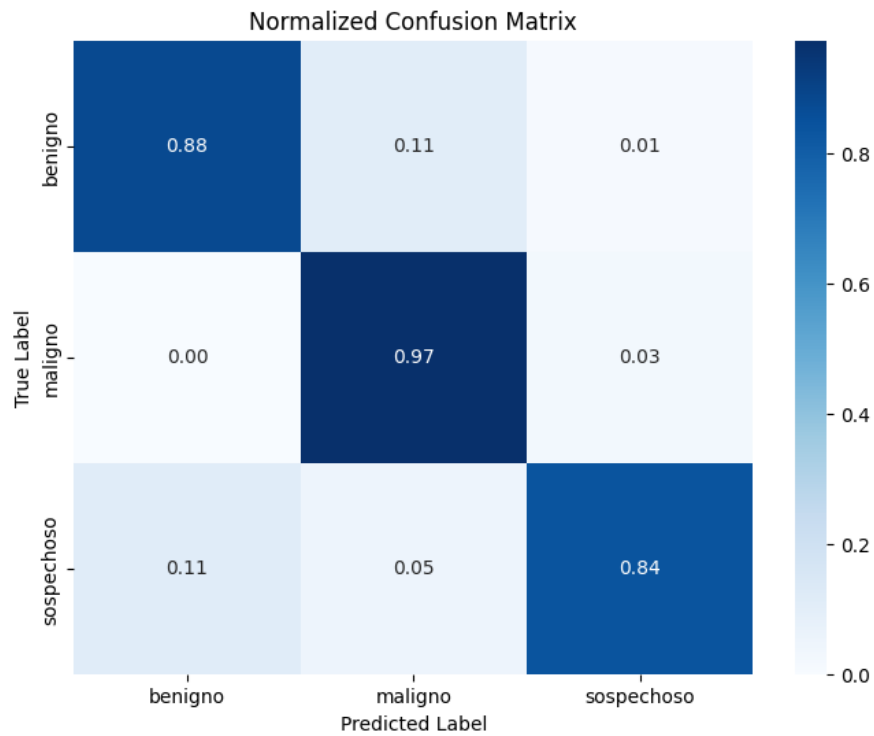


Figura 4.6: Matriz de confusión normalizada - Clasificador Masas

Para realizar un análisis cualitativo sobre el modelo, se presentan algunas muestras del conjunto de test, acompañadas de su clase real y predicha, así como del mapa de atención entregado por el vision transformer.

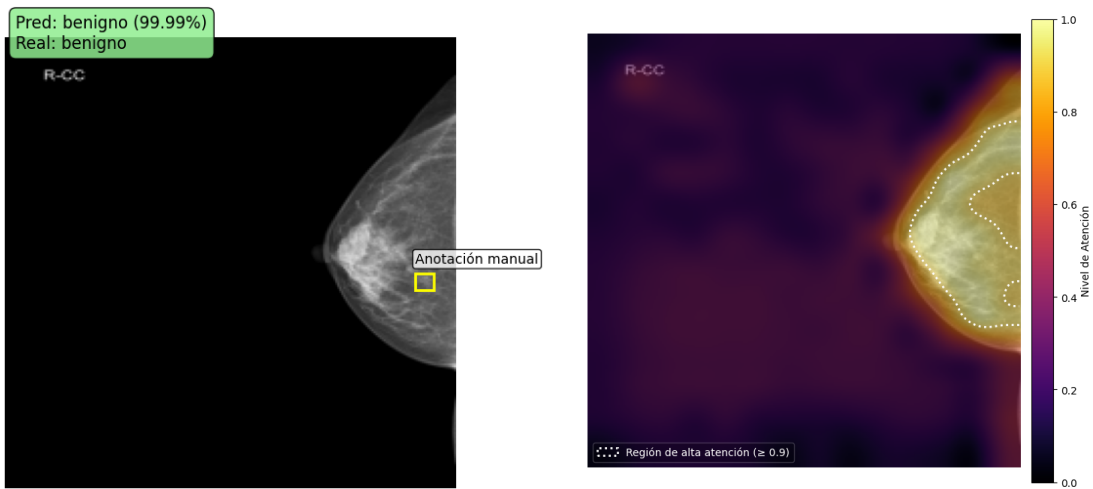


Imagen original.

Mapa de atención del modelo.

Figura 4.7: Ejemplo de predicción del modelo sobre el conjunto de test.

**Predicción:** *Benigno* (99.99 %) **Etiqueta real:** *Benigno*.

El modelo centra la atención en la zona interior del seno, sin enfocarse particularmente en la lesión presente.

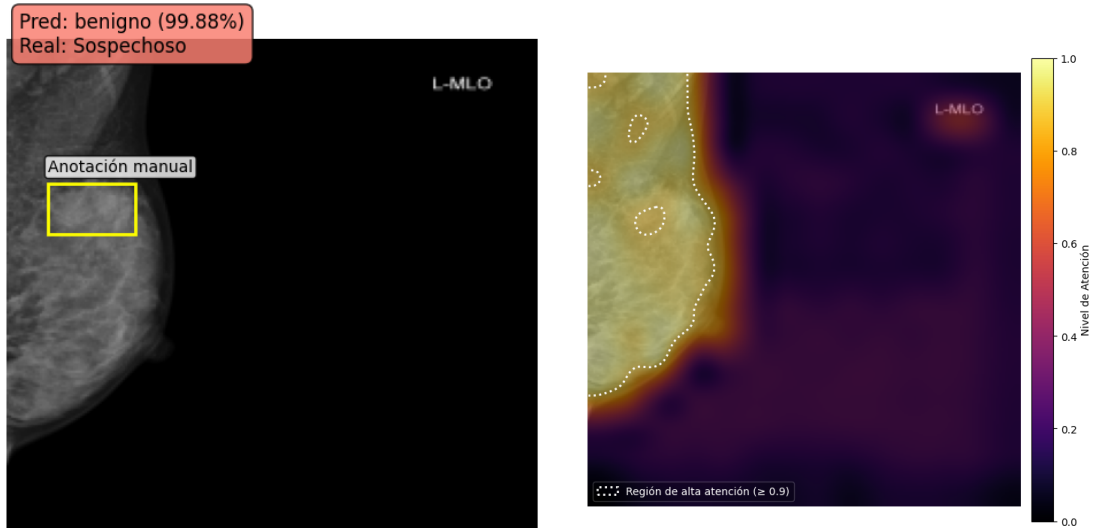


Imagen original.

Mapa de atención del modelo.

Figura 4.8: Ejemplo de predicción del modelo sobre el conjunto de test.

**Predicción:** *Benigno* (99.88 %) **Etiqueta real:** *Sospechoso*.

Se observa que las regiones que presentan un nivel de atención mayor se encuentran al rededor de la masa y no precisamente sobre esta.

A partir de los resultados observados, el clasificador presenta un gran desempeño, con métricas cercanas al 90 %, evidenciando por sobre todo los bajos falsos positivos presentes en las clases ‘benigno’ y ‘maligno’. Cabe destacar que la clase ‘sospechoso’ presenta un rendimiento menor. Esto se debe a que esta clase es difícil de clasificar por definición, abarcando un rango de malignidad muy amplio, que va desde el 2 al 95 %.

### 4.3. Clasificador de Malignidad de Calcificaciones

Repetiendo el enfoque empleado en los dos clasificadores anteriores, se selecciona el modelo obtenido en la última epoch, es decir el número 30. Los indicadores de eficacia alcanzados se encuentran en la tabla 4.3. Las figuras 4.9 y 4.10 muestran la matriz de confusión junto con su correspondiente matriz normalizada.

Tabla 4.3: Métricas de rendimiento para el clasificador de malignidad de calcificaciones sobre los datos de test

<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>
84,5 %	83,8 %	83,8 %	83,7 %

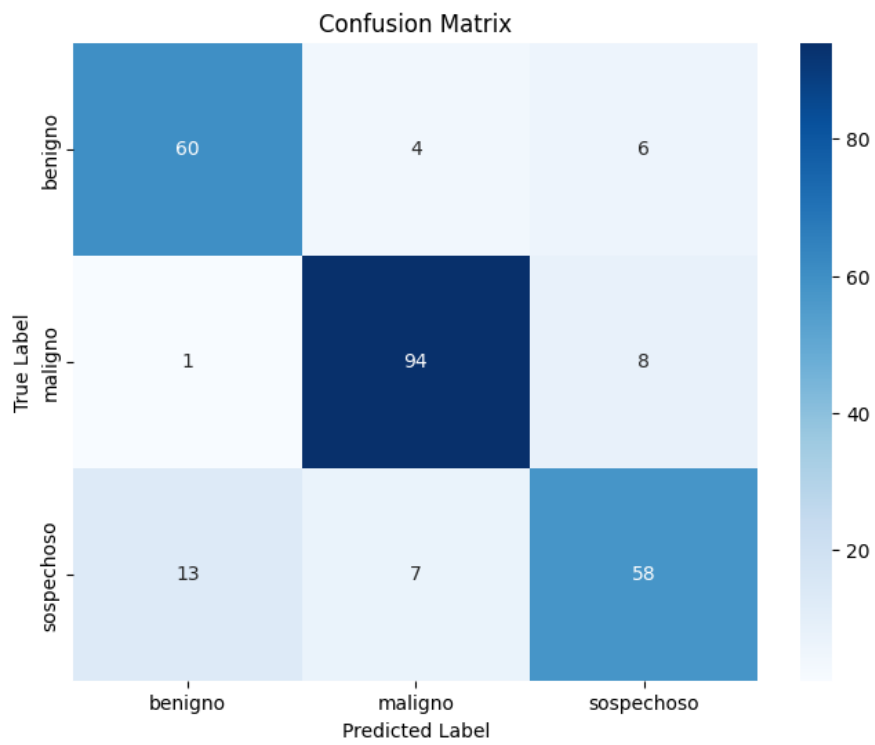


Figura 4.9: Matriz de confusión - Clasificador calcificaciones

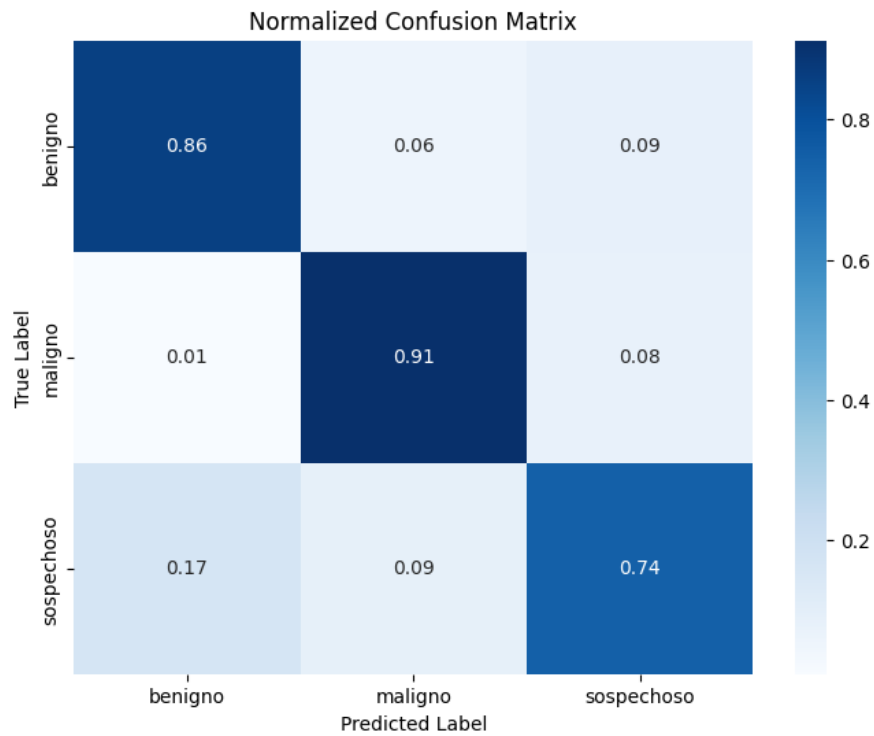


Figura 4.10: Matriz de confusión normalizada - Clasificador calcificaciones

Para complementar los resultados de este clasificador, se incluye la revisión de casos proveniente del conjunto de test, en donde se muestra su clase real y predicha, junto con el mapa de atención correspondiente al clasificador.

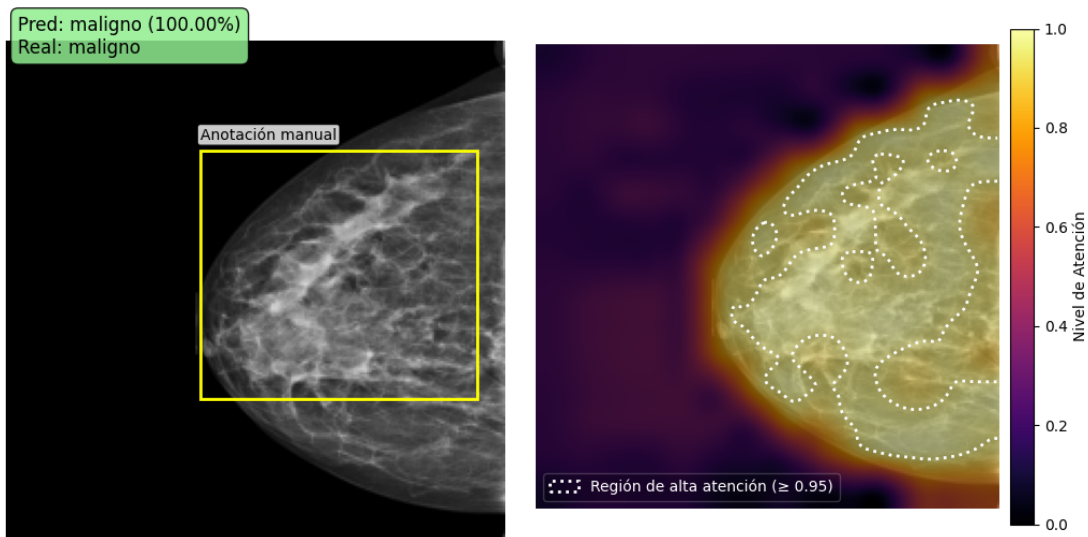


Imagen original.

Mapa de atención del modelo.

Figura 4.11: Ejemplo de predicción del modelo sobre el conjunto de test.

**Predicción:** *Maligno* (100.00 %) **Etiqueta real:** *Maligno*.

La anotación manual es extensa, abarcando una gran región del seno, esta se encuentra dentro de la zona de mayor atención del modelo.

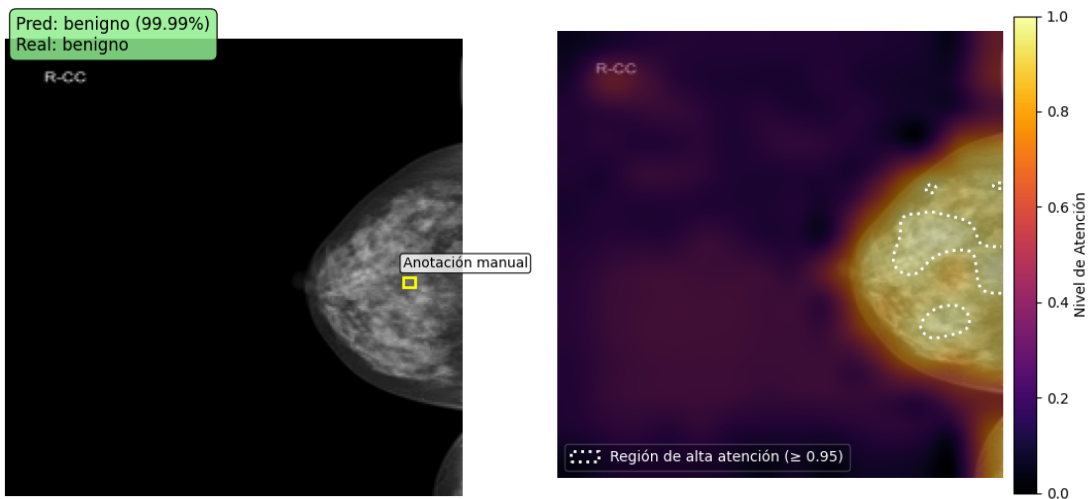


Imagen original.

Mapa de atención del modelo.

Figura 4.12: Ejemplo de predicción del modelo sobre el conjunto de test.

**Predicción:** *Benigno* (99.99 %) **Etiqueta real:** *Benigno*.

En este caso, la anotación es pequeña y las zonas con alta atención igual son de menor tamaño y se encuentran alrededor de la anotación real.

En base a lo expuesto, el desempeño del clasificador es adecuado, con indicadores por sobre el 80 %, sobre todo cuando se trata de clasificar imágenes benignas y malignas. De manera similar al caso previo, la clase ‘sospechoso’ muestra una menor eficacia debido a una mayor confusión, especialmente con la clase ‘benigna’. Esto se debe a las mismas razones expuestas anteriormente para el clasificador de masas.



# Capítulo 5

## Factibilidad de la Solución

En el trabajo realizado en esta memoria se presenta la efectividad para la clasificación de imágenes médicas por parte de la arquitectura de vision transformer, como alternativa a los modelos basados en CNN. En particular, la arquitectura planteada es capaz de entregar un resultado en base a la imagen completa, sin necesidad de un análisis previo donde se deba seleccionar un recorte de la imagen.

A diferencia de otras soluciones propuestas, en que solo entrega un resultado binario (benigno o maligno) para un solo tipo de anomalía establecida, el modelo construido realiza una clasificación multi-clase, siendo capaz de clasificar si la imagen presenta o no un hallazgo, indicando si pertenece a la categoría de masas o calcificaciones. Más aún, la solución planteada permite clasificar la malignidad de dicha lesión en 3 niveles correspondientes al estándar BI-RADS utilizado por los médicos, brindando un mejor apoyo en el proceso de diagnóstico de la enfermedad.

Los resultados conseguidos demuestran que, pese a las limitaciones conocidas en torno al desarrollo de sistemas CAD sobre la falta de datasets públicos con una estructura consistente, es posible sobrellevar este problema usando estrategias como transfer-learning junto con fine-tuning, lo que permite aprovechar el conocimiento adquirido por

modelos previamente entrenados en grandes conjuntos de datos y usarlo como punto de partida para aprender tareas más específicas a partir de conjuntos de datasets más pequeños, como los utilizados en este trabajo. Otra estrategia utilizada en este proyecto para mejorar la capacidad de generalización de los modelos es data augmentation para incrementar la diversidad y enriquecer el conjunto de datos aumentando su tamaño de forma artificial.

En línea con lo anterior, también se valida el nivel de madurez alcanzado por la arquitectura Transformer, existiendo diversas soluciones de alto nivel para trabajar con esta familia de modelos. Un ejemplo destacado es Hugging Face, el cual no solo actúa como un repositorio de modelos pre-entrenados, sino que también brinda un framework completo y fácil de usar para el entrenamiento, validación y despliegue de modelos.

Este enfoque presenta una menor complejidad computacional y un resultado más completo, lo que convierte a **Smart Mammo** en una solución más atractiva para situaciones donde es necesario tener un resultado preliminar de forma inmediata de múltiples exámenes de mamografías (fig. 5.1), permitiendo alertar sobre casos que necesitan mayor atención, gracias a la generación de reportes (fig. 5.2), donde se prioriza por la malignidad de los hallazgos. Esto es altamente beneficioso en entornos donde hay baja disponibilidad de radiólogos, donde se suelen usar mamógrafos móviles para llegar a comunas rurales o alejadas de las zonas céntricas del país.

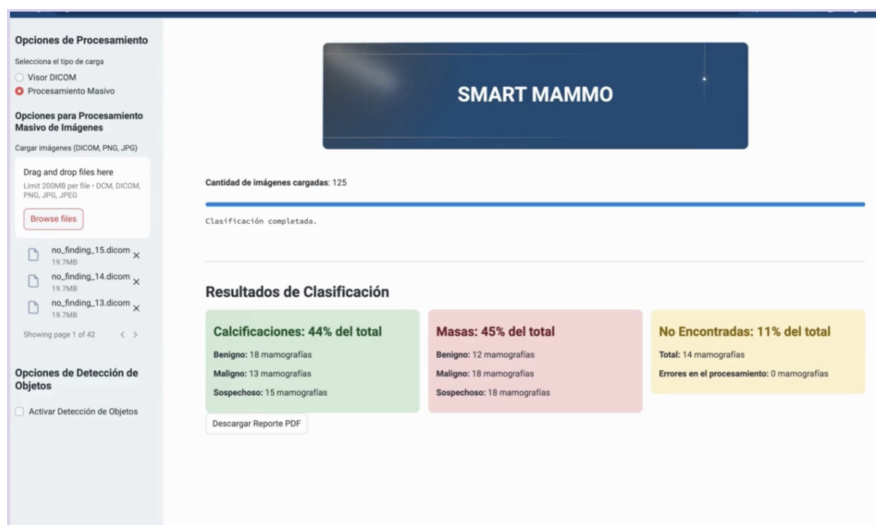


Figura 5.1: Vista de revisión de múltiples imágenes usando Smart Mammo

Reporte de Clasificación Masiva		
Nombre de Imagen	Categoría Primaria	Categoría Secundaria
calc_maligno_1.dicom	Calcificaciones	Maligno
calc_maligno_2.dicom	Calcificaciones	Maligno
calc_maligno_3.dicom	Calcificaciones	Maligno
calc_maligno_4.dicom	Calcificaciones	Maligno
calc_maligno_5.dicom	Calcificaciones	Maligno
calc_maligno_6.dicom	Calcificaciones	Maligno
calc_maligno_7.dicom	Calcificaciones	Maligno
calc_maligno_10.dicom	Calcificaciones	Maligno
calc_maligno_12.dicom	Calcificaciones	Maligno
calc_maligno_14.dicom	Calcificaciones	Maligno
calc_maligno_15.dicom	Calcificaciones	Maligno
calc_maligno_16.dicom	Calcificaciones	Maligno
calc_maligno_17.dicom	Calcificaciones	Maligno
mass_benigno_4.dicom	Masas	Maligno
mass_benigno_14.dicom	Masas	Maligno
mass_maligno_1.dicom	Masas	Maligno
mass_maligno_3.dicom	Masas	Maligno
mass_maligno_4.dicom	Masas	Maligno
mass_maligno_5.dicom	Masas	Maligno
mass_maligno_6.dicom	Masas	Maligno

Resumen	
Total de imágenes procesadas: 125	
Masas: 56 (45%)	
Calcificaciones: 55 (44%)	
No Encontradas: 14 (11%)	
Errores en el procesamiento: 0 mamografías	

Detalle por Categoría Primaria	
<b>Calcificaciones</b>	
Benigno: 18 mamografías	
Maligno: 13 mamografías	
Sospechoso: 15 mamografías	
<b>Masas</b>	
Benigno: 12 mamografías	
Maligno: 18 mamografías	
Sospechoso: 18 mamografías	

Figura 5.2: Vista de un reporte generado por Smart Mammo



# Capítulo 6

## Conclusiones

El cáncer de mama es una de las principales muertes en mujeres en Chile y el mundo, por lo que contar con herramientas para una detección temprana es de suma importancia para incrementar las posibilidades de un tratamiento exitoso. Por lo tanto, el desarrollo de sistemas de ayuda al diagnóstico basados en modelos de machine learning resulta fundamental para asistir a los profesionales de la salud en la revisión de imágenes de mamografías.

El trabajo presentado se orienta en esa dirección, explorando el uso de nuevas tecnologías prometedoras, como los modelos basados en transformer. Estas arquitecturas destacan por su menor complejidad computacional en comparación con otros enfoques, lo que las hace especialmente adecuadas para su implementación en entornos clínicos.

En particular, la metodología desarrollada consiste en separar las tareas de clasificación mediante el entrenamiento de tres modelos independientes. Los resultados obtenidos son alentadores, especialmente en los clasificadores que evalúan la malignidad de las lesiones, alcanzando un rendimiento cercano al 90 %. No obstante, el clasificador principal encargado de detectar hallazgos aún presenta métricas por debajo del 80 %, lo que evidencia áreas de mejora.

## 6.1. Trabajo Futuro

En cuanto a trabajos futuros, se identifican dos líneas de desarrollo. La primera se centra en mejorar la confiabilidad de los clasificadores actuales, donde una opción consiste en mejorar la calidad de los datos de entrenamiento, utilizando más datasets disponibles para poder conseguir un conjunto de datos más variados y no depender de un aumento artificial de los datos. Otra opción es probar distintas arquitecturas de transformer existentes y disponibles para la tarea de clasificar imágenes, evaluando distintos flujos de procesamiento de la imágenes, cómo por ejemplo, explorar el uso de un doble clasificador binario: primero para identificar si hay o no un hallazgo y luego, en caso afirmativo, se clasifica en ‘Masas’ o en ‘Calcificaciones’, en vez del clasificador multiclases presentado.

La segunda línea de desarrollo consiste en llevar los resultados un paso más allá de solo entregar los resultados en base a las observaciones BI-RADS mediante mamografías y en su lugar, emplear resultados de exámenes de biopsias, de esta forma se tendría resultados más precisos sobre si un hallazgo es de tipo benigno o maligno, acotando los resultados posibles y facilitando entregar un diagnóstico más minucioso de cara al paciente. Sin embargo, este enfoque para que sea viable requiere una colaboración más directa con instituciones médicas, debido a la limitada disponibilidad de datos públicos que además de imágenes incluyan información de resultados de biopsia. Con esta metodología, los datos de entrenamiento no quedan sujetos a la interpretación de los radiólogos, sino que se basan en el estudio del tejido sospechoso, reduciendo el error humano.

# Anexos

## Anexo A: Descripción de las Métricas de Evaluación

Las métricas utilizadas para analizar el rendimiento de los modelos se describen a continuación:

- **Accuracy:** mide la proporción de casos clasificados correctamente respecto del número total de objetos del conjunto de datos. Para calcularlo se divide el número de predicciones correctas por el número total de predicciones realizadas por el modelo. Se define como  $Accuracy = \frac{\text{Predicciones Correctas}}{\text{Total de Predicciones}}$
- **Precision:** en la clasificación de múltiples clases es la fracción de instancias correctamente clasificadas como pertenecientes a una clase específica de todas las instancias que el modelo predijo que pertenecerían a esa clase. Se define como  $Precision = \frac{TP_{ClassA}}{TP_{ClassA} + FP_{ClassA}}$
- **Recall:** en la clasificación de múltiples clases es la fracción de instancias en una clase que el modelo clasificó correctamente de todas las instancias en esa clase. Se define como  $Recall = \frac{TP_{ClassA}}{TP_{ClassA} + FN_{ClassA}}$
- **F1 Score:** Corresponde a la media armónica de la precisión y el recall, y se utiliza para evaluar el balance entre estas dos métricas en problemas de clasificación. En el caso multiclase no se calcula de forma general, sino que se realiza de forma individual por clase. El F1 score se define como  $F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$



# Bibliografía

- [1] World Health Organization. *Breast Cancer*. Accessed: 2025-03. Mar. de 2024. URL: <https://www.who.int/es/news-room/fact-sheets/detail/breast-cancer>.
- [2] Freddie Bray et al. “Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries”. En: *CA: A Cancer Journal for Clinicians* 74.3 (2024), págs. 229-263. DOI: <https://doi.org/10.3322/caac.21834>. eprint: <https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21834>. URL: <https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21834>.
- [3] Gloria Icaza, Loreto Núñez y Herna Bugueño. “Descripción epidemiológica de la mortalidad por cáncer de mama en mujeres.” En: *Revista Médica de Chile* 145.1 (oct. de 2016). URL: <https://www.revistamedicadechile.cl/index.php/rmedica/article/view/4763>.
- [4] Solange Parra-Soto et al. “Cáncer en Chile y en el mundo: una mirada actual y su futuro escenario epidemiológico”. En: *Revista Médica de Chile* 148.10 (jul. de 2020). URL: <https://www.revistamedicadechile.cl/index.php/rmedica/article/view/8058>.
- [5] Carlos Rodríguez-Garcés, Denisse Espinosa-Valenzuela y Geraldo Padilla-Fuentes. “Cáncer y acción preventiva en Chile: perfilando la abstención a la mamografía y papanicolaou”. En: *Revista Médica de Chile* 149.8 (jun. de 2021). URL: <https://www.revistamedicadechile.cl/index.php/rmedica/article/view/8058>.

//www.revistamedicadechile.cl/index.php/rmedica/article/view/8642.

- [6] Ma. Fca. Maldonado Wilson. *La importancia de la detección temprana del cáncer de mama: sobrevida es cercana al 99 por ciento*. Accessed: 2025-03. Oct. de 2023. URL: <https://uchile.cl/noticias/210435/mes-del-cancer-de-mama-la-importancia-de-una-deteccion-temprana>.
- [7] American Cancer Society. *Understanding Your Mammogram Report*. Accessed: 2025-03. American Cancer Society. Ene. de 2022. URL: <https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/mammograms/understanding-your-mammogram-report.html>.
- [8] American Cancer Society. *Breast Density and Your Mammogram Report*. Accessed: 2025-03. American Cancer Society. Sep. de 2024. URL: <https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/mammograms/breast-density-and-your-mammogram-report.html>.
- [9] Kosmia Loizidou, Rafaella Elia y Costas Pitris. “Computer-aided breast cancer detection and classification in mammography: A comprehensive review”. En: *Computers in Biology and Medicine* 153 (2023), pág. 106554. ISSN: 0010-4825. DOI: <https://doi.org/10.1016/j.compbiomed.2023.106554>. URL: <https://www.sciencedirect.com/science/article/pii/S0010482523000197>.
- [10] Hiba Chougrad, Hamid Zouaki y Omar Alheyane. “Multi-label transfer learning for the early diagnosis of breast cancer”. En: *Neurocomputing* 392 (2020), págs. 168-180. ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2019.01.112>. URL: <https://www.sciencedirect.com/science/article/pii/S0925231219304710>.

- [11] Brian L. Sprague et al. “Prevalence of Mammographically Dense Breasts in the United States”. En: *JNCI: Journal of the National Cancer Institute* 106.10 (sep. de 2014), dju255. ISSN: 0027-8874. DOI: 10.1093/jnci/dju255. eprint: <https://academic.oup.com/jnci/article-pdf/106/10/dju255/18614661/dju255.pdf>. URL: <https://doi.org/10.1093/jnci/dju255>.
- [12] Brian L. Sprague et al. “Trends in Clinical Breast Density Assessment From the Breast Cancer Surveillance Consortium”. En: *J Natl Cancer Inst* 111.6 (jun. de 2019), págs. 629-632. DOI: 10.1093/jnci/djy210. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6579740/>.
- [13] Gelan Ayana et al. “Vision-Transformer-Based Transfer Learning for Mammogram Classification”. En: *Diagnostics* 13.2 (2023). ISSN: 2075-4418. DOI: 10.3390/diagnostics13020178. URL: <https://www.mdpi.com/2075-4418/13/2/178>.
- [14] Shi Gengtian, Bai Bing y Zhang Guoyou. “EfficientNet-Based Deep Learning Approach for Breast Cancer Detection With Mammography Images”. En: *2023 8th International Conference on Computer and Communication Systems (ICCCS)*. 2023, págs. 972-977. DOI: 10.1109/ICCCS57501.2023.10151156.
- [15] R. Sawyer-Lee et al. *Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) [Data set]*. 2016. URL: <https://doi.org/10.7937/K9/TCIA.2016.7002S9CY>.
- [16] Khalil ur Rehman et al. “Computer Vision-Based Microcalcification Detection in Digital Mammograms Using Fully Connected Depthwise Separable Convolutional Neural Network”. En: *Sensors* 21.14 (2021). ISSN: 1424-8220. DOI: 10.3390/s21144854. URL: <https://www.mdpi.com/1424-8220/21/14/4854>.

- [17] University of South Florida. *DDSM: Digital Database for Screening Mammography*. Accessed: 2025-03. URL: <http://www.eng.usf.edu/cvprg/Mammography/Database.html>.
- [18] Felipe A. Zeiser et al. “Segmentation of Masses on Mammograms Using Data Augmentation and Deep Learning”. En: *Journal of Digital Imaging* 33 (2020), págs. 858-868. DOI: 10.1007/s10278-020-00330-4. URL: <https://doi.org/10.1007/s10278-020-00330-4>.
- [19] Mugahed A. Al-antari, Seung-Moo Han y Tae-Seong Kim. “Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms”. En: *Computer Methods and Programs in Biomedicine* 196 (2020), pág. 105584. ISSN: 0169-2607. DOI: <https://doi.org/10.1016/j.cmpb.2020.105584>. URL: <https://www.sciencedirect.com/science/article/pii/S0169260720314176>.
- [20] Inês C. Moreira et al. “INbreast: Toward a Full-Field Digital Mammographic Database”. En: *Academic Radiology* 19.2 (feb. de 2012). Epub 2011 Nov 10, págs. 236-248. DOI: 10.1016/j.acra.2011.09.014.
- [21] Alexey Dosovitskiy et al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”. En: *ArXiv abs/2010.11929* (2021). URL: <https://doi.org/10.48550/arXiv.2010.11929>.
- [22] Ting-Yu Lin y Mei-Ling Huang. *Dataset of Breast Mammography Images with Masses*. Ver. V2. 2020. DOI: 10.17632/ywsbh3ndr8.2.
- [23] Hugging Face. *Vision Transformer (base-sized model)*. Accessed: 2024. URL: <https://huggingface.co/google/vit-base-patch16-224>.
- [24] Hugging Face. *Vision Transformer (large-sized model)*. Accessed: 2024. URL: <https://huggingface.co/google/vit-large-patch16-224>.

- [25] John Suckling et al. “Mammographic Image Analysis Society (MIAS) database v1.21”. En: (2015). DOI: 10.17863/CAM.105113. URL: <https://www.repository.cam.ac.uk/handle/1810/250394>.
- [26] Hieu T. Nguyen et al. “VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography”. En: *medRxiv* (2022). DOI: 10.1101/2022.03.07.22272009. URL: <https://www.medrxiv.org/content/early/2022/03/10/2022.03.07.22272009>.
- [27] European Commission et al. *European guidelines for quality assurance in breast cancer screening and diagnosis – Fourth edition, supplements*. Ed. por L Karsa et al. Publications Office, 2013. DOI: 10.2772/13196. URL: <https://data.europa.eu/doi/10.2772/13196>.
- [28] McCausland Center. *The DICOM Standard*. Accessed: 2024. URL: <https://web.archive.org/web/20150920230923/http://www.mccauslandcenter.sc.edu/mricro/dicom/index.html>.