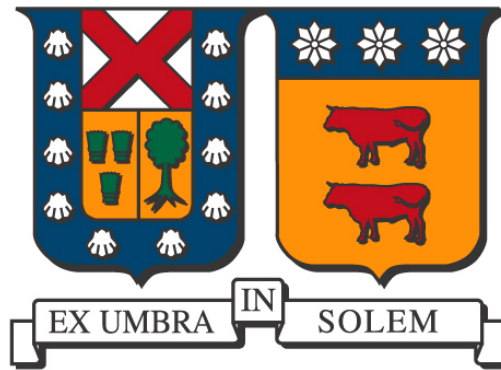


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA
VALPARAÍSO - CHILE



“PREDICCIÓN AUTOMATIZADA DE SOSPECHA DE
CÁNCER DE MAMA EN BASE AL ANÁLISIS DE
EXÁMENES DE MAMOGRAFÍA”

IAN PAUL ROBERTS VALENZUELA

MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE INGENIERO CIVIL
TELEMÁTICO

PROFESOR GUÍA

MAURICIO ARAYA LÓPEZ

PROFESOR CORREFERENTE

MARCOS ZÚÑIGA BARRAZA

MARZO - 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: Predicción Automatizada de Sospecha de Cáncer de Mama en Base al Análisis de Exámenes de Mamografía

Nombre del candidato(a): Ian Paul Roberts Valenzuela

Carrera / Grado: Ingeniería Civil Telemática

Campus: Casa Central

Departamento: Departamento de 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 (marcar una opción)

☒ 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 (**marcar una opción**):

☐ 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: 01/12/2025

Firma:

Estudiante o Candidato(a):

Fecha: 01/12/2025

Firma:

Agradecimientos

Detrás de este trabajo hay agradecimientos a todos aquellos que me han apoyado y acompañado durante todos estos años. Quisiera agradecer especialmente a mis padres, por todo lo que han hecho por mí, y por inculcarme los valores y principios que me hicieron quien soy hoy día. A todos mis amigos, por todas las experiencias que me han aportado, por los buenos y malos momentos, y por estar ahí siempre. Agradezco también a mi profesor guía, Mauricio Araya, por todas las puertas que me abrió en la Universidad, y por toda la paciencia que me tuvo.

Resumen

El cáncer de mama es una de las principales causas de mortalidad entre las mujeres en todo el mundo, y su detección temprana y precisa desempeña un papel fundamental para mejorar los resultados de las pacientes. Esta tesis se centra en la predicción de los estadios del cáncer mediante imágenes de mamografía y técnicas de aprendizaje profundo, aprovechando las redes neuronales para automatizar y mejorar la precisión diagnóstica. Este trabajo emplea un conjunto de datos de mamografías anotadas, con las cuales se entrenan modelos basados en redes neuronales convolucionales (CNN) para clasificar los estadios del cáncer, incorporando arquitecturas avanzadas y aprendizaje por transferencia para mejorar la precisión y la solidez. El rendimiento de los modelos propuestos se evalúa según métricas establecidas, con las cuales se demuestra la capacidad de estos para distinguir eficazmente entre los estadios del cáncer con alta sensibilidad y especificidad. Los resultados muestran el potencial de los enfoques impulsados por inteligencia artificial para ayudar a los radiólogos en la toma de decisiones clínicas, reduciendo el error humano y permitiendo un diagnóstico más rápido. Esta investigación contribuye al creciente campo de la imagen médica y la inteligencia artificial, proporcionando una solución escalable para mejorar el diagnóstico del cáncer de mama y la atención al paciente.

Palabras clave: redes neuronales convolucionales, aprendizaje de máquinas, mamografías

Abstract

Breast cancer is a leading cause of mortality among women worldwide, and its early and accurate detection plays a critical role in improving patient outcomes. This thesis focuses on cancer stage prediction using mammography images and deep learning techniques, leveraging neural networks to automate and improve diagnostic accuracy. This work employs an annotated mammography dataset, on which convolutional neural network (CNN)-based models are trained to classify cancer stages, incorporating advanced architectures and transfer learning to improve accuracy and robustness. The performance of the proposed models is evaluated against established metrics, with which the ability of these models to effectively distinguish between cancer stages with high sensitivity and specificity is demonstrated. The results show the potential of AI-driven approaches to assist radiologists in clinical decision making, reducing human error and enabling faster diagnosis. This research contributes to the growing field of medical imaging and artificial intelligence, providing a scalable solution to improve breast cancer diagnosis and patient care.

Keywords: convolutional neural networks, machine learning, mammography

Índice general

1. Introducción	1
1.1. Propósito	1
1.2. Alcance	2
1.3. Objetivos de la Memoria	2
1.3.1. General	2
1.3.2. Específicos	2
1.3.3. Requerimientos Funcionales	3
1.3.4. Requerimientos No Funcionales	3
1.4. Estructura de la memoria	3
2. Estado del Arte	5
2.1. Cáncer	5
2.1.1. Tumores	5
2.1.2. Estadios	6
2.1.3. Cáncer de Mama	6
2.2. Datasets	7
2.3. Trabajos Relacionados	8
2.3.1. Basados en Datos Tabulados	8
2.3.2. Trabajos con Datos Mixtos	9
2.3.3. Trabajos con Imágenes	10
2.4. ResNet	11
2.5. Xception	12
2.6. EfficientNetV2	13
2.7. Ensemble Learning	14

3. Propuesta	17
3.1. Definición del Problema	17
3.2. Propuesta de Solución	17
3.2.1. Contexto General del Prototipo	18
3.2.2. Esquema General de Componentes	18
3.3. Materiales	19
3.3.1. Stack de Tecnologías	19
3.3.2. Ambiente de Desarrollo	19
3.3.3. Dataset: CBIS-DDSM	20
4. Diseño de la solución	23
4.1. Exploración de datos	23
4.1.1. Anomalías	23
4.1.2. Patologías	24
4.1.3. Evaluación	25
4.2. Preprocesamiento de los Datos	25
4.2.1. Cajas Englobantes	27
4.3. Generación de los Datasets	28
4.3.1. Dataset S1	29
4.3.2. Dataset S10	30
4.4. Red Neuronal de Parches	30
4.4.1. Preprocesamiento y Aumentación de Datos	31
4.4.2. Entrenamiento	31
4.5. Red Neuronal de Imágenes Completas	31
4.5.1. Preprocesamiento y Aumentación de Datos	32
4.5.2. Entrenamiento	32
4.6. Metamodelo Ensemble Learning	32
5. Resultados	35
5.1. Métricas	35
5.2. Resultados EfficientNetV2	36
5.2.1. Parches	36
5.2.2. Mamografías	37
5.2.3. Testing	38
5.3. Resultados ResNet50	40
5.3.1. Parches	40

5.3.2. Mamografías	41
5.3.3. Testing	42
5.4. Resultados Xception	44
5.4.1. Parches	44
5.4.2. Mamografías	45
5.4.3. Testing	46
5.5. Resultados Ensemble Learning	47
6. Conclusiones	49
6.1. Conclusiones generales	49
6.2. Conclusiones específicas	49
6.2.1. Sobre los Datasets	49
6.2.2. Sobre las Redes Neuronales	50
6.3. Trabajo Futuro	50

Índice de figuras

2.1. Sistema TNM para cáncer de mama.	6
2.2. Bloque residual ResNet.	12
2.3. Bloque Inception	13
2.4. Bloque XCception	14
3.1. Diagrama del esquema general del sistema.	19
3.2. Ejemplo de mamografía CBIS-DDSM (CC).	21
4.1. Tipo de anomalía	23
4.2. Cantidad de anomalías	24
4.3. Patologías	24
4.4. Evaluaciones	25
4.5. Ejemplo máscaras de segmentación estandarizadas.	26
4.6. Cajas englobantes generadas.	27
4.7. Histogramas de altura y ancho para las cajas englobantes generadas.	28
4.8. Ejemplo de <i>sampling</i> de parches para dataset S1. Los cuadrados en blanco corresponden al <i>ground-truth</i> de la lesión. Los cuadrados en gris claro corresponden a los parches generados para las lesiones, y los cuadrados en gris oscuro corresponden a los parches de fondo.	29
4.9. Parches generados.	29
4.10. Ejemplo de <i>sampling</i> de parches para dataset S10.	30
5.1. <i>Loss</i> y <i>Accuracy</i> sobre el conjunto de entrenamiento para EfficientNetV2 sobre los datasets de parches.	36
5.2. <i>Loss</i> y <i>Accuracy</i> sobre el conjunto de validación para EfficientNetV2 sobre los datasets de parches.	37

5.3. <i>Loss y Accuracy</i> sobre el conjunto de entrenamiento para EfficientNetV2 sobre los datasets de mamografías.	37
5.4. <i>Loss y Accuracy</i> sobre el conjunto de validación para EfficientNetV2 sobre los datasets de mamografías.	38
5.5. Matrices de confusión para EfficientNetV2 sobre imágenes de mamografía.	39
5.6. <i>Loss y Accuracy</i> sobre el conjunto de entrenamiento para ResNet50 sobre los datasets de parches.	40
5.7. <i>Loss y Accuracy</i> sobre el conjunto de validación para ResNet50 sobre los datasets de parches.	41
5.8. <i>Loss y Accuracy</i> sobre el conjunto de entrenamiento para ResNet50 sobre los datasets de mamografías.	41
5.9. <i>Loss y Accuracy</i> sobre el conjunto de validación para ResNet50 sobre los datasets de mamografías.	42
5.10. Matrices de confusión para ResNet50 sobre imágenes de mamografía.	43
5.11. <i>Loss y Accuracy</i> sobre el conjunto de entrenamiento para Xception sobre los datasets de parches.	44
5.12. <i>Loss y Accuracy</i> sobre el conjunto de validación para Xception sobre los datasets de parches.	44
5.13. <i>Loss y Accuracy</i> sobre el conjunto de entrenamiento para Xception sobre los datasets de mamografía.	45
5.14. <i>Loss y Accuracy</i> sobre el conjunto de validación para Xception sobre los datasets de mamografía.	46
5.15. Matrices de confusión para Xception sobre imágenes de mamografía.	47

Índice de tablas

2.1. Estadios de cáncer	6
2.2. Niveles de BIRADS	7
3.1. Descripción datos CBIS-DDSM	21
4.1. Cantidad de imágenes y máscaras de segmentación por conjunto de datos.	26
4.2. Cantidad de imágenes por conjunto de datos.	27
5.1. Métricas obtenidas por EfficientNetV2 para entrenamiento sobre parches. Se presentan únicamente los resultados del modelo seleccionado (el que presenta <i>validation loss</i> más bajo).	36
5.2. Métricas obtenidas por EfficientNetV2 para entrenamiento sobre imágenes de mamografía. Se presentan únicamente los resultados del modelo seleccionado (el que presenta <i>validation loss</i> más bajo).	38
5.3. Métricas obtenidas por EfficientNetV2 sobre el conjunto de testing para imágenes de mamografía.	39
5.4. Métricas obtenidas por ResNet50 para entrenamiento sobre parches. Se presentan únicamente los resultados del modelo seleccionado (el que presenta <i>validation loss</i> más bajo).	40
5.5. Métricas obtenidas por ResNet50 para entrenamiento sobre imágenes de mamografía. Se presentan únicamente los resultados del modelo seleccionado (el que presenta <i>validation loss</i> más bajo).	42
5.6. Métricas obtenidas por ResNet50 sobre el conjunto de <i>testing</i> para imágenes de mamografía.	42
5.7. Métricas obtenidas por Xception para entrenamiento sobre parches. Se presentan únicamente los resultados del modelo final seleccionado (el que presenta <i>validation loss</i> más bajo).	45

5.8. Métricas obtenidas por XCeption para entrenamiento sobre imágenes de mamografía. Se presentan únicamente los resultados del modelo seleccionado (el que presenta <i>validation loss</i> más bajo).	45
5.9. Métricas obtenidas por XCeption sobre el conjunto de <i>testing</i> para imágenes de mamografía.	46
5.10. Métricas obtenidas por los metamodelos de <i>ensemble learning</i> sobre el conjunto de <i>testing</i> . 47	
5.11. Métricas obtenidas por los metamodelos de <i>ensemble learning</i> sobre el conjunto de <i>testing</i> . 48	

Capítulo 1

Introducción

Actualmente, el cáncer de mama es el tipo de cáncer más comúnmente diagnosticado a nivel global, correspondiendo a 1 de cada 8 (12,5 %) de los casos diagnosticados. Esta enfermedad es, a su vez, la segunda causa de muerte en mujeres, siendo responsable del 2.5 % de fallecimientos.

Para combatir el cáncer de mama, se establece un control anual en que se realiza una mamografía al paciente, de forma que la enfermedad pueda ser detectada y posteriormente tratada en sus etapas tempranas, y evitar poner en riesgo a quien la padece. Esto conlleva a un gran volumen de exámenes realizados, los cuales deben ser analizados por un especialista. Dada la cantidad de exámenes, este proceso puede llegar a ser bastante lento, llegando a tomar hasta meses antes de que un paciente reciba una respuesta.

Esta naturaleza generalizada del cáncer de mama requiere una re-evaluación de las metodologías de detección convencionales, con especial énfasis en el diagnóstico temprano para el tratamiento del paciente. Es en este cambio de paradigmas en donde se encuentra la integración con redes neuronales e inteligencia artificial con la medicina, nexo que durante los últimos años ha tomado cada vez más fuerza.

Con el objetivo de mejorar esta problemática, es que la FALP ha propuesto un desafío para generar una solución de carácter predictivo utilizando redes neuronales, mediante la exploración exhaustiva de amplios conjuntos de datos, de forma que esta sea capaz de discernir patrones útiles para poder clasificar estas imágenes dentro de tres estados de presencia de cáncer: “Sano”, “Benigno” o “Maligno”.

1.1. Propósito

El presente trabajo tiene como propósito el poder entregar una herramienta de diagnóstico asistido, que sirva como un instrumento de priorización para el gran volumen de exámenes de mamografías que debe manejar el sistema de salud, permitiendo que aquellos casos con alta sospecha de padecer cáncer, puedan ser vistos por un especialista oportunamente.

Para esto, se desarrollará una solución enfocada al uso de inteligencia artificial, principalmente redes neuronales, que son capaces de aprender a identificar características a partir de imágenes de entrada, en este caso, mamografías, para después entregar una probabilidad sobre la pertenencia a alguna clase. Se espera que un proyecto de este tipo sea capaz de aprender a identificar los posibles estados de presencia de cáncer anteriormente definidos.

1.2. Alcance

El alcance de esta memoria es contribuir al diagnóstico temprano de casos de cáncer de mama dentro de la FALP, en el contexto de exámenes de mamografía, mediante el desarrollo de una herramienta capaz de generar una predicción sobre la sospecha de cáncer a partir de una prueba realizada por un médico.

Para lograr esto, se conforma un equipo compuesto por:

- Benjamín Carrasco: memorista de Ing. Civil Telemática, encargado de la integración de la red neuronal generada con la interfaz usuaria, para proveer un sistema completo utilizable por los médicos.
- Ian Roberts: memorista de Ing. Civil Telemática, responsable de la exploración de datos, y del diseño e implementación de la red neuronal, mediante uso de metodologías de minería de datos y técnicas de aprendizaje de máquinas.
- Juan Pablo Sánchez: memorista de Ing. Civil Telemática, a cargo de la interfaz usuaria y de proveer el acceso de la herramienta generada a los médicos de FALP.

1.3. Objetivos de la Memoria

A partir del desafío propuesto por la FALP, complementado con reuniones que se tuvieron con la contraparte, se definen los siguientes objetivos para la memoria:

1.3.1. General

Implementar una herramienta de diagnóstico asistido, basada en técnicas de *Machine Learning*, para priorizar exámenes de mamografía según el grado de sospecha de presencia de cáncer.

1.3.2. Específicos

1. **Diseñar y entrenar** un sistema, basado en redes neuronales, capaz de realizar predicciones sobre la presencia de cáncer a partir de una imagen de examen de mamografía.

2. **Diseñar e implementar** una interfaz de usuario intuitiva, de forma que el sistema pueda ser utilizado por médicos, quienes posiblemente no tienen un alto nivel de conocimientos computacionales.
3. **Validar** el funcionamiento del sistema, mediante métricas y retroalimentación de la contraparte.

1.3.3. Requerimientos Funcionales

1. Ser capaz de recibir el examen de mamografía en su formato original (DICOM).
2. Entregar una probabilidad para indicar la presencia de cáncer en el examen recibido por el sistema.
3. Proveer una interfaz gráfica de interacción al usuario.
4. Levantar una alerta o notificación en caso de que el examen analizado sea sospechoso de padecer cáncer de mama.

1.3.4. Requerimientos No Funcionales

1. Poseer una interfaz accesible, sencilla e intuitiva al usuario.
2. Ejecutarse de manera rápida, demorándose mucho menos que el análisis de un radiólogo en entregar una respuesta.
3. Ser confiable, en cuanto a la confidencialidad de la información.

1.4. Estructura de la memoria

Este escrito está estructurado de la siguiente manera: El capítulo 2 se centra en presentar el trabajo desarrollado en problemas similares por otros autores, además de explicar la teoría que se utiliza para la realización de la memoria. En el capítulo 3, se define el problema y se presenta el esquema de la solución. También se detallan los materiales y el ambiente de trabajo a utilizar durante el desarrollo de la memoria. En el capítulo 4, se desarrolla el diseño de la solución y su implementación. A continuación, en el capítulo 5 se presentan los resultados obtenidos por la solución detallada anteriormente. Finalmente, en el capítulo 6, se muestra un conjunto de conclusiones, generales y específicas, en torno al problema y solución desarrollados, además de las proyecciones del trabajo de memoria.

Capítulo 2

Estado del Arte

2.1. Cáncer

El cáncer es el crecimiento descontrolado de células anormales en alguna parte del cuerpo, las cuales luego pueden esparcirse a otros órganos y tejidos. En un cuerpo sano, las células crecen y se multiplican normalmente, en lo que se conoce como “proceso de división celular”, formando nuevas células a medida que el cuerpo las necesita. Las células viejas o dañadas mueren, permitiendo que las nuevas tomen su lugar. Este proceso natural puede ser interrumpido por células dañadas o anormales, que sufrieron ciertos cambios en los genes que controlan este ciclo, lo que hace que estas crezcan y se multipliquen cuando no deben. La agrupación de estas células descontroladas pasa a formar una masa o protuberancia, conocida como tumor, el cual puede ser clasificado como benigno o maligno.

2.1.1. Tumores

Un tumor benigno es una masa de células anormales que crece de manera controlada y localizada, sin invadir tejidos cercanos ni propagarse a otras partes del cuerpo. Por lo general, los tumores benignos no representan un riesgo grave para la salud (salvo que se encuentren presionando alguna parte importante del cuerpo, como el cerebro), y pueden ser removidos mediante cirugía con bajas probabilidades de recurrencia.

Por otra parte, un tumor maligno se caracteriza por un crecimiento descontrolado y agresivo, con la capacidad de invadir tejidos vecinos y esparcirse a otras áreas del cuerpo a través del sistema linfático o sanguíneo, proceso conocido como metástasis. Este comportamiento hace que los tumores malignos sean potencialmente mortales y requieran tratamientos más complejos, como quimioterapia, radioterapia u operaciones complejas.

2.1.2. Estadios

Para la clasificación de estadios, los oncólogos utilizan el sistema TNM, el cual nos permite determinar una categoría para el tumor (T), los nódulos linfáticos (N) y la presencia de metástasis (M), cuya combinación se utiliza para encasillar al paciente en un estadio entre 0 y 4.

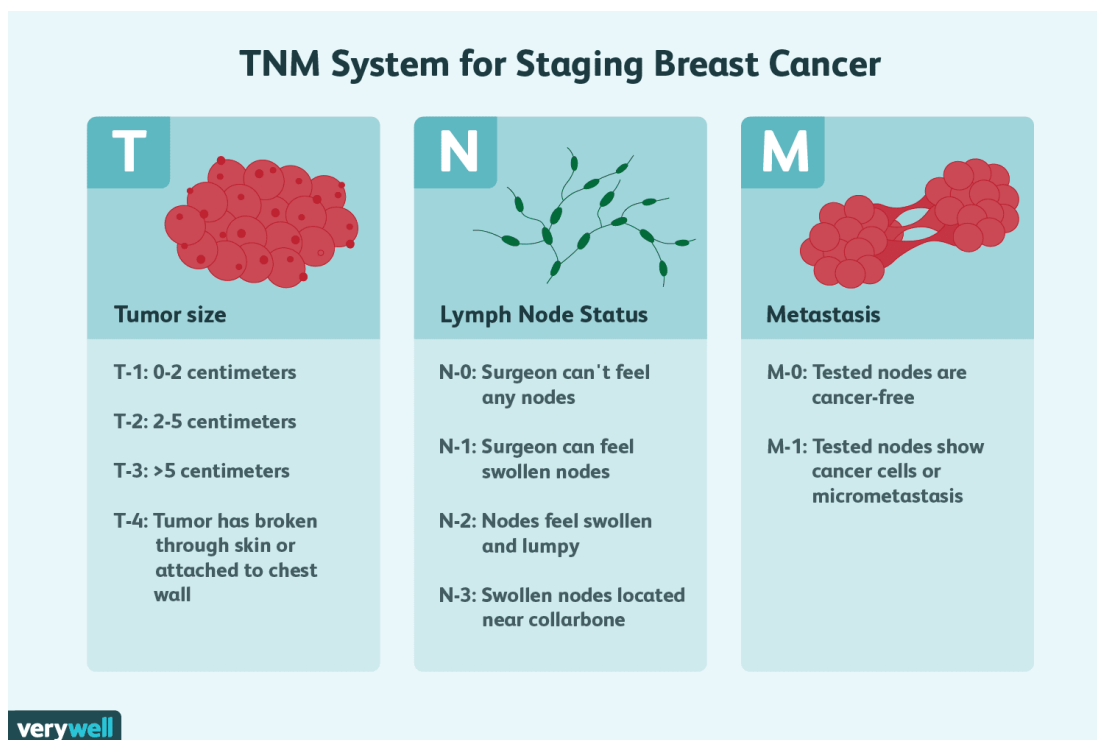


Figura 2.1: Sistema TNM para cáncer de mama.

El sistema TNM ayuda a describir el cáncer en mucho detalle. Pero para muchos tipos de cáncer, las combinaciones de TNM se agrupan en cinco estadios más generales.

Estadio	
Estadio 0	Hay células anormales pero no se diseminaron al tejido cercano. Este estadio también se llama <i>carcinoma in situ</i> (CIS). El CIS no es cáncer, pero podría volverse canceroso.
Estadio I, II y III	Hay cáncer. Mientras más alto es el número, más grande es el tumor canceroso y más extendido está en los tejidos cercanos.
Estadio IV	El cáncer se diseminó a otros órganos del cuerpo.

Tabla 2.1: Estadios de cáncer

2.1.3. Cáncer de Mama

El cáncer de mama es una enfermedad caracterizada por el crecimiento descontrolado de células malignas en el tejido mamario, principalmente en los conductos que transportan la leche (carcinoma ductal)

o en los lóbulos donde se produce la leche (carcinoma lobulillar). Se da principalmente en mujeres, aunque puede darse con menor frecuencia en hombres. 1 de cada 8 diagnósticos de cáncer se atribuyen a cáncer de mama, dándose un total de 2.3 millones de casos nuevos en el año 2020. [9]

Para el caso del cáncer de mama en específico, existe además el sistema BI-RADS (Breast Imaging Reporting and Data System), que consiste en una herramienta estandarizada desarrollada por el Colegio Americano de Radiología (ACR) para clasificar y describir hallazgos en estudios de imágenes mamarias, como mamografías, ecografías y resonancias magnéticas. Su objetivo principal es facilitar la comunicación entre los profesionales de la salud, mejorar la interpretación de los resultados y guiar la toma de decisiones clínicas. BI-RADS clasifica los hallazgos en siete categorías numeradas del 0 al 6, cuya descripción puede verse en la Tabla 2.2

BIRADS	
BIRADS 0: No concluyente	Significa que el radiólogo pudo haber observado alguna lesión sospechosa pero que no está definida con claridad. Por ese motivo puede ser necesario realizar una mamografía focalizada o magnificada del área de interés, o bien una ecografía adicional.
BIRADS 1: Mama normal	Significa que no se identifica ninguna imagen de sospecha, las estructuras son normales y no hay ningún comentario que agregar. El estudio se considera negativo porque no hay ningún hallazgo anormal.
BIRADS 2: Hallazgos benignos	Significa que el estudio muestra hallazgos benignos y no hay signos de cáncer.
BIRADS 3: Hallazgos probablemente benignos	Los hallazgos en esta categoría tienen muy alta probabilidad (más de 98 %) de ser benignos (no cancerosos).
BIRADS 4: Anormalidad sospechosa	Significa que hay algún hallazgo mamográfico sospechoso. No parece indicar que se trate de cáncer pero, debido a que existe alguna sospecha, se debería realizar una biopsia mamaria.
BIRADS 5: Anormalidad que sugiere un hallazgo maligno	Significa que los hallazgos tienen apariencia de cáncer y hay alta probabilidad de que sea cáncer. La imagen puede ser una masa con contornos extraños, microcalcificaciones irregulares con disposición lineal, ductal o arboriforme.
BIRADS 6: Lesión maligna ya demostrada	Esta categoría se utiliza únicamente para los hallazgos en una mamografía que ya se demostró que son cancerosos, según una biopsia realizada anteriormente.

Tabla 2.2: Niveles de BIRADS

2.2. Datasets

Muchos estudios se han realizado en el área del cáncer de mama, para estudiar la posibilidad de obtener un diagnóstico asistido por computadora, mediante diversas técnicas de *Machine Learning*. Estos trabajos no se limitan únicamente a las imágenes de mamografía, sino que también pueden utilizar características extraídas por un especialista a partir de biopsias, o los documentos del historial clínico

del paciente. La inmensa mayoría de estos suele utilizar conjuntos de datos disponibles en internet de manera pública. De estos, los más comunes son:

- **Wisconsin Breast Cancer Dataset (WBCD):** Consiste en un conjunto de datos tabulados, que presentan características como el radio, textura o área de una lesión, las cuales se obtuvieron mediante el análisis de un especialista a partir de biopsias de cáncer de mama. Estos conjuntos son clasificados dentro de las categorías “benignas” o “malignas”.
- **Breast Cancer Coimbra Dataset (BCCD):** Breast Cancer Coimbra Dataset corresponde a un dataset tabulado que contiene información sobre factores biológicos y anatómicos de pacientes, como la edad, BMI, niveles de glucosa, insulina, entre otros. El dataset clasifica a los pacientes en dos posibles grupos, “sanos” o “cáncer”, sin hacer distinción entre los tipos de cáncer. Cabe mencionar que es el único dataset público que incluye características de pacientes sanos.
- **Cohort Dataset:** El Cohort dataset es un conjunto de datos tabulados obtenidos a partir del historial clínico y de tratamiento de pacientes de cáncer de mama, e incluye información como la edad al momento de diagnóstico, la cantidad de nodos linfáticos, la cantidad de tratamientos de quimioterapia, y el historial familiar.
- **Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM):** Corresponde a un conjunto de imágenes de mamografía, en su formato original (DICOM), obtenidas a partir de diversos hospitales de Estados Unidos. Este dataset consiste en una versión curada y estandarizada del dataset Digital Database for Screening Mammography (DDSM), que presentaba problemas en cuanto a la calidad de imágenes y anotaciones. Este es el dataset más utilizado para trabajos que utilizan las imágenes de mamografía.
- **INbreast:** Similar a CBIS-DDSM, INbreast es un conjunto de imágenes de mamografía. Estas fueron obtenidas del *Centro Hospitalar de S. Joao*, Portugal. Si bien es un conjunto de datos más moderno que CBIS-DDSM, presenta una cantidad bastante reducida de imágenes en comparación.

2.3. Trabajos Relacionados

2.3.1. Basados en Datos Tabulados

La mayoría de los estudios que buscan abordar el problema de la clasificación del cáncer de mama lo hacen principalmente a partir de datos tabulados, con características previamente extraídas por un especialista.

En el documento “*An expert system for detection of breast cancer based on association rules and neural network*” [7], se presenta un sistema de diagnóstico automático para la detección de cáncer de mama basado en reglas de asociación y una red neuronal. Bajo este enfoque, se utilizan reglas de asociación para reducir el espacio de características, las cuales luego son clasificadas por una red neuronal simple. El modelo es probado con un esquema de validación cruzada, utilizando el dataset WBCD, obteniendo una precisión del 95.6 %.

Utilizando el mismo conjunto de datos, en [2] se presenta una comparación de seis algoritmos de aprendizaje de máquinas: GRU-SVM, Regresión Lineal, Multilayer Perceptron, Nearest Neighbor, Regresión Softmax y Support Vector Machines (SVM). Cada una de estas técnicas es probada por separado para el dataset WBCD. Los resultados presentados indican que todas las técnicas son efectivas para la clasificación de estos datos tabulados en las clases “benigno” o “maligno”, alcanzando una precisión superior al 90 % en todos los casos, siendo el Multilayer Perceptron el más destacado, con una precisión de 99.04 %.

En [13], se continúa con la idea presentada en el trabajo anterior, mediante la aplicación de nuevas técnicas al conjunto de datos WBCD. En este trabajo se estudia el uso de K-Nearest Neighbors, Regresión Logística, Árboles de Decisión, Random Forest, SVM y Aprendizaje Profundo. De manera similar al estudio anterior, todas las técnicas utilizadas alcanzan una precisión superior al 90 %, siendo el Aprendizaje Profundo el que obtiene los mejores resultados, con una precisión del 98.24 %.

En [12] se presenta un enfoque de *Ensemble Learning* para el conjunto de datos tabulados Cohort. En este trabajo, se comienza realizando una discriminación de características útiles utilizando métodos de selección estadísticos, para luego realizar una clasificación mediante árboles de decisión, cuyas respuestas son combinadas utilizando *Bootstrap Aggregation*. El método se compara con SVMs y *Multilayer Perceptron*, alcanzando los mejores resultados al utilizar el método de *ensemble*, que obtiene una precisión del 96.5 %.

2.3.2. Trabajos con Datos Mixtos

En el estudio “*A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction*” [17] se generan tres modelos para la predicción del riesgo de cáncer a partir de un conjunto de datos privado: un modelo de regresión logística, que utiliza biomarcadores y factores de riesgo para analizar el riesgo de cáncer en el paciente; un modelo de aprendizaje profundo, basado en la arquitectu-

ra ResNet18, que emplea exclusivamente exámenes visuales de mamografías para estimar el riesgo de cáncer; y un modelo híbrido que combina ambos enfoques. Este modelo híbrido utiliza ResNet18 como extractor de características para generar un *embedding*, el cual se concatena con las características de factores de riesgo utilizadas por el modelo de regresión logística. El vector resultante se emplea para entrenar un *Multilayer Perceptron*, que alcanza los mejores resultados entre los tres modelos, logrando un AUC de 0.70.

2.3.3. Trabajos con Imágenes

En el contexto de utilización de exámenes imagenológicos, en el documento “*Mammogram Analysis Using Feed-Forward Back Propagation and Cascade-Forward Back Propagation Artificial Neural Network*” [15], se utiliza el algoritmo de *Grey-Level Cooccurrence Matrix* (GLCM) para extraer características de las imágenes de mamografía. El algoritmo es comúnmente utilizado para la detección de texturas en imágenes, por lo que se espera que sea capaz de identificar texturas anómalas, en este caso, calcificaciones, dentro de las imágenes de mamografía. Estas características se utilizan luego para entrenar una red neuronal, alcanzando una precisión del 87.5 % sobre un conjunto de datos privado.

Continuando la búsqueda para clasificar calcificaciones, en [18] se genera un modelo de diagnóstico asistido, capaz de extraer posibles regiones de interés a partir de imágenes de mamografía utilizando un algoritmo de detección *fuzzy*. Estas regiones de interés luego son procesadas para generar un vector de características, el cual es entregado a una red neuronal para realizar una predicción sobre la malignidad del área detectada. El modelo propuesto alcanza una precisión del 83.3 % sobre un conjunto de datos privado.

En [4] se hace uso de imágenes de ultrasonido para clasificación en las categorías de “benigno” o “maligno”. Para esto, el modelo propone un algoritmo de segmentación para regiones de interés dentro de la imagen, además de la extracción de características a partir de estas utilizando *wavelets*. Los vectores extraídos por los algoritmos son luego utilizados para entrenar un *Multilayer Perceptron*, alcanzando una precisión del 72.73 % sobre un conjunto de datos privado.

En el documento “*Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer*” [10] se hace una revisión de la literatura para la clasificación de cáncer de mama utilizando técnicas de redes neuronales, para los distintos tipos de exámenes disponibles: mamografías, MRI, ecografías y termografías.

En [8] Se presenta un enfoque utilizando redes neuronales convolucionales para la predicción de mamografías en las categorías “benigno” y “maligno”. En este trabajo, se emplea la arquitectura ResNet50 para la clasificación de lesiones (calcificaciones y masas) según su grado de malignidad (benigno o maligno), a partir de “parches” extraídos de las mamografías anotadas. Estos modelos entrenados con parches son luego transferidos a modelos de imágenes completas mediante *transfer learning* para la predicción de mamografías. El trabajo utiliza el dataset CBIS-DDSM, reportando un AUC de 0.88 para el mejor modelo.

Finalmente, en “*CoroNet: Deep Neural Network-Based End-to-End Training for Breast Cancer Diagnosis*” [11] se presenta un modelo de aprendizaje profundo basado en la arquitectura Xception para la clasificación de lesiones presentes en mamografías, considerando 2 clases (calcificaciones y masas), alcanzando una precisión del 88.67 % sobre el dataset CBIS-DDSM.

Dentro de la literatura también es posible encontrar trabajos con imágenes digitalizadas de biopsias [1], que corresponden a una porción del tejido infectado visto por microscopio, tecnología que no se encuentra actualmente disponible de manera masiva en nuestro país.

No se pudo encontrar estado del arte relacionado con sistemas que, además de la detección de cáncer, tengan un énfasis en la clasificación del estadio. Tampoco existe, hasta donde sabemos, una propuesta que haya intentado integrar diversos exámenes para la predicción del estadio del paciente.

2.4. ResNet

ResNet [6] corresponde a una familia de redes neuronales convolucionales que hacen uso de bloques residuales, los cuales buscan abordar el desafío del entrenamiento de redes neuronales profundas. A diferencia de una red neuronal clásica, en las que la entrada es transformada por un conjunto de capas convolucionales y luego por la función de activación, en un bloque residual, la entrada se agrega a la salida de las capas convolucionales, creando esencialmente un “atajo”. Este permite que el gradiente pueda fluir de mejor manera, permitiéndonos crear redes más profundas sin ser afectadas por el problema del gradiente desvaneciente.

La salida de un bloque residual puede denotarse por:

$$H(x) = F(x) + x \quad (2.1)$$

La arquitectura de un bloque residual puede verse en la Figura 2.2

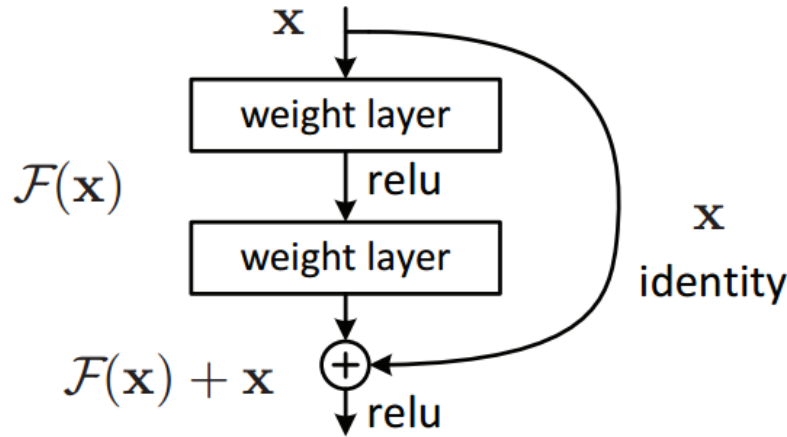


Figura 2.2: Bloque residual ResNet.

Esta adición de la entrada directamente a la salida de la capa garantiza que el rendimiento de la red no se degrade a medida que se agregan más capas, permitiéndonos crear una arquitectura más profunda.

Una red de tipo ResNet consiste en unir múltiples bloques residuales juntos, pudiendo identificar modelos como ResNet18, ResNet34, ResNet50 o ResNet128 según la cantidad de capas del modelo.

2.5. Xception

Xception (Extreme Inception) [5] es una arquitectura de red neuronal convolucional que se basa en los principios de la red Inception, pero introduce una mejora significativa al utilizar convoluciones separables en profundidad. En el módulo Inception tradicional (Figura 2.3), la extracción de características se realiza combinando varias operaciones convolucionales con diferentes tamaños de *kernel* para capturar jerarquías espaciales de distintas escalas, que luego se concatenan para formar el resultado.

Xception reemplaza estas operaciones convolucionales con convoluciones separables en profundidad, que separan el filtrado espacial de las operaciones por canal (Figura 2.4). Esto permite que el modelo procese de forma independiente la información espacial y de profundidad (canal), lo que mejora significativamente la eficiencia computacional y la potencia de representación.

En una convolución separable en profundidad, se aplica una única convolución espacial de forma independiente a cada canal, seguida de una convolución puntual (*kernel* 1x1) para combinar las características resultantes en todos los canales. Este enfoque reduce la cantidad de parámetros y el costo computacional en comparación con las convoluciones estándar, al tiempo que preserva la capacidad del modelo para aprender patrones complejos.

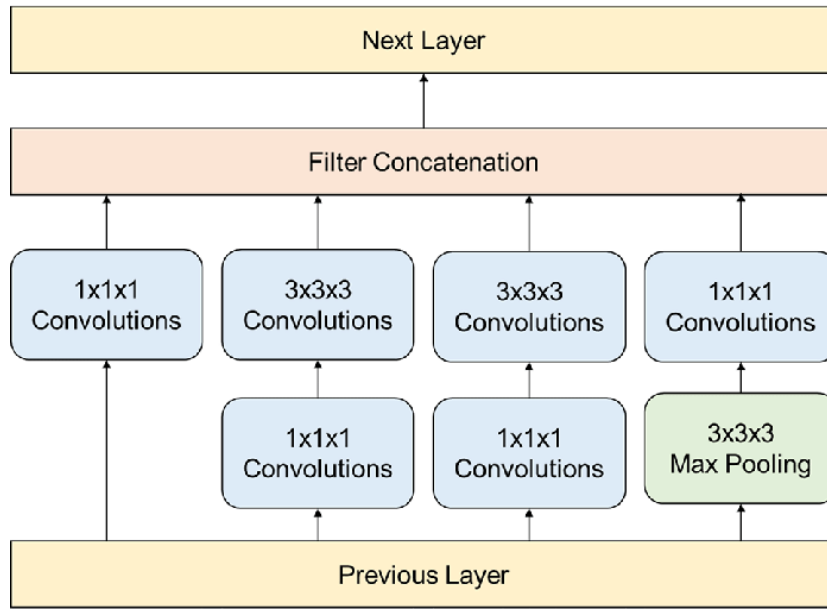


Figura 2.3: Bloque Inception

El resultado de una convolución separable en profundidad se puede describir como:

$$H(x) = \text{Pointwise}(\text{Depthwise}(x)) \quad (2.2)$$

Esta arquitectura está organizada en una serie de módulos similares a los bloques residuales de ResNet, con atajos agregados para mejorar el flujo de gradiente y facilitar el entrenamiento de redes más profundas.

2.6. EfficientNetV2

EfficientNetV2 [16] representa una familia de redes neuronales convolucionales diseñadas para lograr una alta precisión con una eficiencia computacional mejorada mediante la combinación de escalado de modelos y optimización de diseño. A diferencia de las redes tradicionales, que simplemente aumentan la cantidad de capas o parámetros para mejorar el rendimiento, EfficientNetV2 utiliza escalado compuesto, equilibrando la profundidad (cantidad de capas), el ancho (cantidad de canales por capa) y la resolución (tamaño de la imagen) para optimizar la red para tareas específicas. Además, introduce bloques de convolución fusionados, que reemplazan las convoluciones separables en profundidad tradicionales en las primeras etapas de la red, combinando la convolución y la normalización por lotes en un solo paso para reducir la sobrecarga y mejorar la velocidad de entrenamiento.

La red emplea bloques residuales invertidos de estilo MobileNetV2 en sus últimas etapas, que consis-

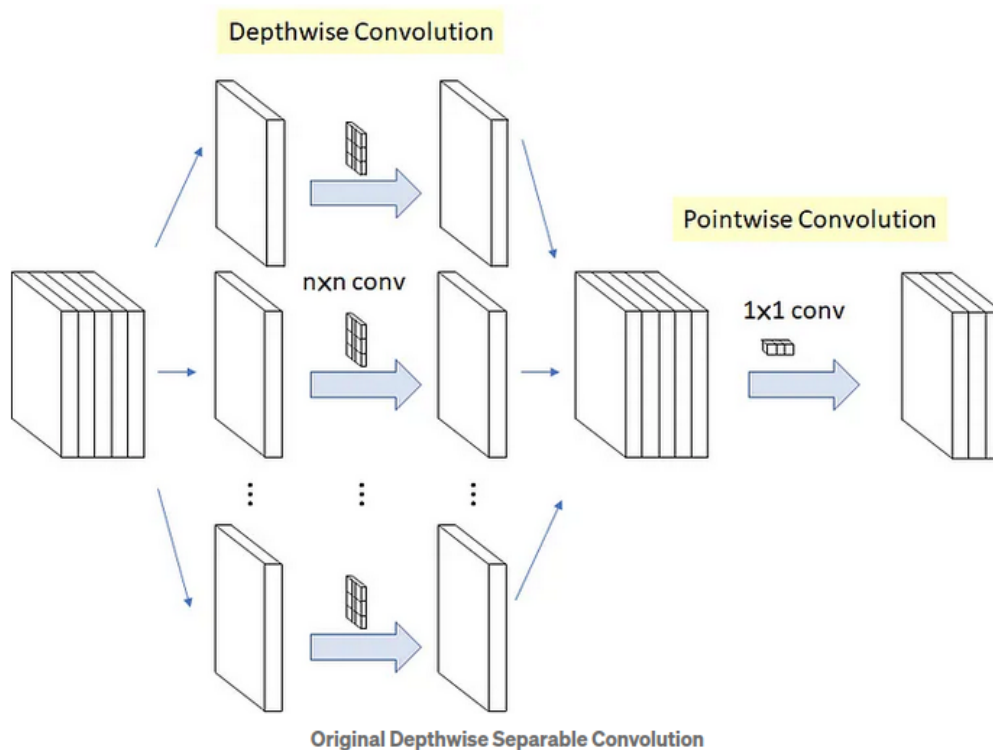


Figura 2.4: Bloque Xception

ten en una secuencia de convoluciones separables en profundidad livianas combinadas con una conexión de acceso directo para facilitar el flujo de gradiente y la reutilización de características. Esta combinación de convoluciones fusionadas y bloques residuales invertidos permite a EfficientNetV2 procesar conjuntos de datos de imágenes a gran escala mientras mantiene requisitos computacionales y de memoria más bajos en comparación con arquitecturas anteriores.

El diseño de EfficientNetV2 pone énfasis en el entrenamiento progresivo, que comienza con resoluciones de imagen más pequeñas y las aumenta gradualmente durante el entrenamiento, mejorando aún más tanto la precisión como la eficiencia.

Así, se obtiene la familia de modelos EfficientNetV2-S, EfficientNetV2-M y EfficientNetV2-L, que reflejan diferentes configuraciones de tamaño y complejidad computacional para adaptarse a varios entornos de hardware.

2.7. Ensemble Learning

Ensemble Learning, o aprendizaje por conjuntos es una técnica que combina varios modelos para mejorar la precisión y reducir la probabilidad de errores en comparación con el uso de un solo modelo.

En lugar de depender de un solo modelo para hacer predicciones, los métodos de conjunto agregan los resultados de varios modelos, aprovechando sus fortalezas y compensando las debilidades individuales. Estos modelos pueden ser del mismo tipo (homogéneos) o de diferentes tipos (heterogéneos). El aprendizaje por conjuntos es particularmente eficaz porque reduce la varianza, el sesgo o ambos, según el enfoque utilizado.

Existen diversas formas de lograr esta combinación de resultados, entre las que podemos encontrar:

- **Bagging:** El *bagging*, o Bootstrap Aggregation, es una técnica en donde se entrenan varios modelos de forma independiente en diferentes subconjuntos de los datos y sus predicciones se combinan mediante votación por mayoría.
- **Boosting:** En *boosting*, se entrenan los modelos de forma secuencial, y cada modelo se centra en los errores de su predecesor para mejorar el rendimiento de forma iterativa.
- **Stacking:** Aplicar una técnica de *stacking* implica entrenar un metamodelo para combinar las predicciones de varios modelos base y aprender la forma óptima de combinar sus resultados.

Al combinar múltiples modelos, el aprendizaje en conjunto mitiga eficazmente las limitaciones de cada modelo individual, lo que da como resultado un sistema de predicción robusto y preciso que se usa ampliamente en tareas que van desde el reconocimiento de imágenes hasta el procesamiento del lenguaje natural.

Capítulo 3

Propuesta

3.1. Definición del Problema

El problema surge en hospitales y centros clínicos del país, donde se recomienda a todas las mujeres mayores de 40 años realizarse una mamografía preventiva de forma anual. En caso de existir antecedentes familiares, este proceso puede iniciarse aún más temprano. Esta dinámica genera un enorme volumen de exámenes, cuyos resultados pueden tardar entre 30 y 45 días en ser analizados por un radiólogo. En enfermedades de esta gravedad, cada 30 días de retraso en el inicio del tratamiento disminuyen en un 9 % las probabilidades de supervivencia del paciente [3]. Esta situación resulta inaceptable, por lo que se hace imprescindible mejorar el proceso de diagnóstico.

Actualmente, este se da mediante un sistema de filas, en donde la primera persona en realizarse el examen obtiene su resultado antes que el resto. En esta memoria, se busca la generación de un sistema de software capaz de predecir los estados de sospecha de cáncer, herramienta con la cual se podría posteriormente crear una fila priorizada de exámenes, en la cual aquellos que se sospeche tengan una alta probabilidad de padecer la enfermedad, se analicen antes que aquellos clasificados como un examen sano.

3.2. Propuesta de Solución

Como propuesta de solución, se busca el desarrollar un sistema capaz de realizar la predicción automatizada para los posibles diagnósticos de cáncer de mama, que pueden ser: “Sano”, “Benigno” o “Maligno”. El sistema generado debe ser capaz de recibir como entrada distintas imágenes de mamografía, las cuales tendrán un formato tipo DICOM (Digital Imaging and Communications in Medicine), que corresponde al tipo de examen médico más utilizado para el diagnóstico de cáncer de mama. Una

vez recibidos uno o más de estos archivos, el sistema procederá a realizar una predicción sobre la etapa de la enfermedad.

Para lograr esto, se propone un esquema que hace uso de un conjunto de redes neuronales con distintas arquitecturas, implementando diversos trabajos del estado del arte. Así, se entrenan estos modelos utilizando imágenes y subsecciones de estas en donde se presentan lesiones, las cuales son utilizadas en el diagnóstico de esta enfermedad. Una vez entrenadas, se realizará un proceso de conjunto de redes, buscando que la combinación sea capaz de detectar las lesiones aprendidas anteriormente en las imágenes.

3.2.1. Contexto General del Prototipo

El prototipo implementado cumple con la funcionalidad de generar la detección del estado de cáncer de mama a partir del análisis de una imagen de mamografía, junto al diseño e iteración inicial de las interfaces usuarias.

A continuación, se detalla cada uno de los componentes del prototipo inicial, que corresponden a:

- **Preprocesamiento de Imágenes:** Componente encargado de estandarizar las dimensiones de la imagen y validar el formato en estándar DICOM. Luego, procede a transformar la imagen a un formato entendible por la red neuronal para poder realizar una predicción.
- **Modelo de detección de cáncer:** Corresponde al componente encargado de detectar los posibles estados de presencia de cáncer en la imagen entregada por el doctor. Estos estados pueden ser “Sano”, “Benigno” o “Maligno”. Se comunica con la API para poder entregar una respuesta al sistema, y depende del módulo de Preprocesamiento de Imágenes para poder funcionar.
- **API REST:** Es el componente encargado de comunicar las interfaces usuarias con la red neuronal, permitiendo el presentar al usuario la predicción para una imagen de mamografía.
- **Interfaz Web:** La aplicación web esta enfocada en generar los componentes base para la conexión con el modelo de detección de cáncer. Estos consisten en:
 - Barra de navegación
 - Carga de archivos de mamografía
 - Traspaso de documentos a API

3.2.2. Esquema General de Componentes

A continuación, se presenta un esquema general de cómo funcionan en conjunto los componentes descritos para generar el sistema.

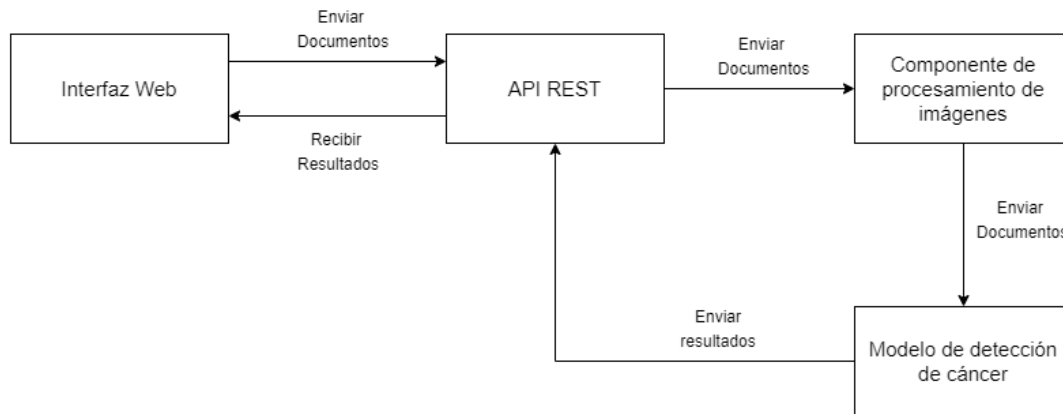


Figura 3.1: Diagrama del esquema general del sistema.

En esta memoria, se procederá a detallar la formulación, creación y aplicación de los componentes del **Modelo de detección de cáncer** y del **Componente de procesamiento de imágenes**. Los demás componentes, desarrollados por otros miembros del grupo de memoria multidisciplinaria, serán tratados como una caja negra, asumiendo que funcionan de forma correcta. Para más detalles sobre estos, se puede consultar el informe de memoria individual de los otros integrantes del grupo.

3.3. Materiales

3.3.1. Stack de Tecnologías

Para el desarrollo de la red neuronal, se utilizará Python, específicamente la librería de *Machine Learning* e inteligencia artificial PyTorch. Además, para el preprocesamiento y la visualización de los datos, se empleará un entorno de Jupyter Notebook, dado que ofrece facilidades para presentar imágenes.

3.3.2. Ambiente de Desarrollo

A continuación, se presentan los detalles del hardware y software utilizados durante el desarrollo del sistema.

Hardware de Desarrollo

- 16 GB de memoria RAM
- 200 GB de almacenamiento de disco duro (HDD)
- GPU NVidia GeForce GTX 1060 6GB

Software de Desarrollo

- Python 3.11.2
 - NumPy
 - Pandas
 - OpenCV
 - PyDicom
 - PyTorch

3.3.3. Dataset: CBIS-DDSM

Toda aplicación imagenológica que haga uso de redes neuronales requiere de una gran cantidad de imágenes anotadas, de forma que la solución obtenida sea capaz de generalizar correctamente a casos que no ha visto anteriormente. Para el desarrollo del prototipo inicial se utiliza el dataset “Curated Breast Imaging Subset of Digital Database for Screening Mammography” [14] (CBIS-DDSM), disponible públicamente en Internet. Este se encuentra compuesto de imágenes de mamografías digitalizadas provenientes de distintas universidades e instituciones de salud de Estados Unidos.

El dataset cuenta con un total de 3568 imágenes de mamografías correspondientes a 1566 pacientes distintos. La variación en la cantidad de imágenes y pacientes se debe a que un paciente puede tener imágenes tanto del seno izquierdo como del derecho, además de que existen distintas vistas desde las cuales se puede capturar la imagen: Cráneo-caudal (CC) y Medio Lateral Oblicua (MLO). Complementando las imágenes, el dataset incluye una máscara de segmentación que permite identificar en la imagen la ubicación de la lesión que llevó al especialista a determinar la benignidad o malignidad de un tumor.

Además de las imágenes, el dataset contiene un archivo CSV en donde se presentan datos respecto a las masas o calcificaciones encontradas en las imágenes. Un mayor detalle respecto a esto puede verse en la Tabla 3.1

Un punto importante a destacar, y que se detallará con mayor precisión en las secciones posteriores, corresponde a la definición con la que se presentan las imágenes, como se puede observar en la Figura 3.2. El dataset CBIS-DDSM corresponde a un conjunto de imágenes publicadas en 1997, por lo que las mamografías datan de algunos años antes. La diferencia en la calidad de imagen obtenida con los equipos de radiología de aquella época en comparación con los actuales podría ser un factor clave al momento de transferir el modelo desarrollado a una aplicación real.

Columna	Descripción
Patient ID	ID del paciente. Permite distinguir que imágenes de mamografía corresponden a que estudio.
Breast Density	Indica la densidad del seno del cual se tomó la mamografía.
Left or Right Breast	Es una variable que indica si la imagen corresponde a una mamografía del seno izquierdo o derecho.
Image View	Indica a que vista corresponde la imagen de mamografía. Puede ser CC o MLO.
Abnormality Type	Indica a que tipo de lesión corresponde la máscara marcada en la mamografía. Puede ser 'Masa' o 'Calcificación'.
Assessment	Detalla el nivel de BIRADS que se le asigno a la lesión por un radiólogo.
Pathology	Indica si la lesión corresponde a un tipo de cáncer Benigno o Maligno.

Tabla 3.1: Descripción datos CBIS-DDSM

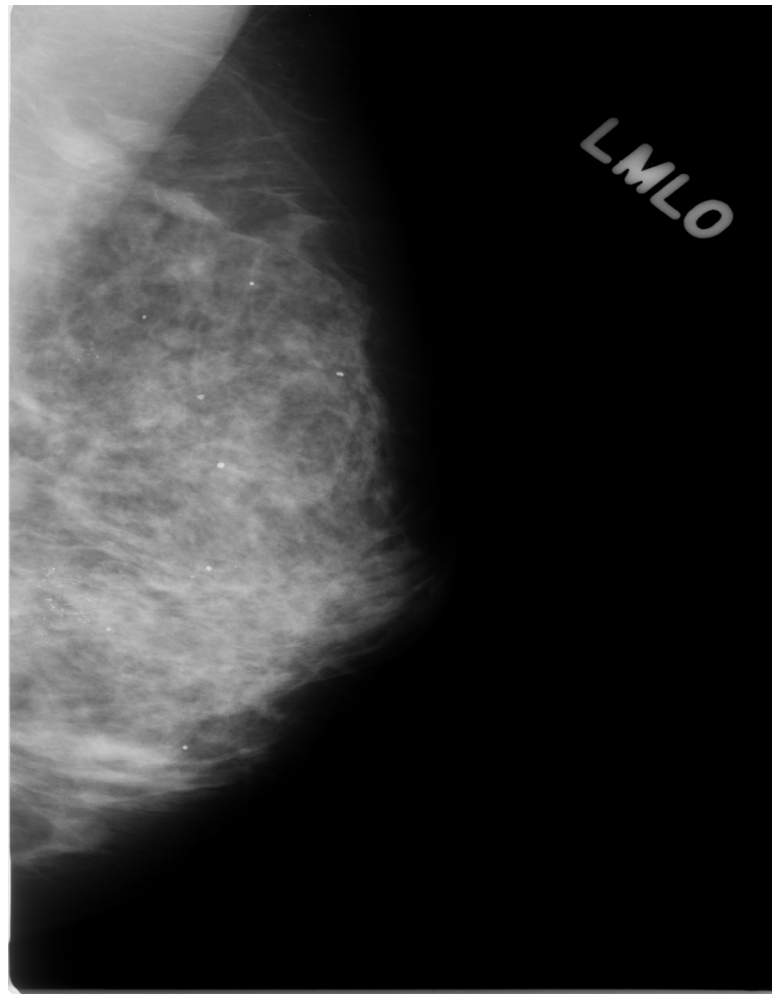


Figura 3.2: Ejemplo de mamografía CBIS-DDSM (CC).

Capítulo 4

Diseño de la solución

4.1. Exploración de datos

Se comienza con una exploración inicial de los datos obtenidos de CBIS-DDSM. Los datos vienen ya separados en conjuntos de *training* y *testing*, por lo que se mantendrá este orden para la exploración inicial de estos.

4.1.1. Anomalías

Comenzando por las anomalías presentes en los datos, se pueden distinguir dos características de interés dentro del conjunto: el tipo de anomalía y la cantidad de estas por paciente.

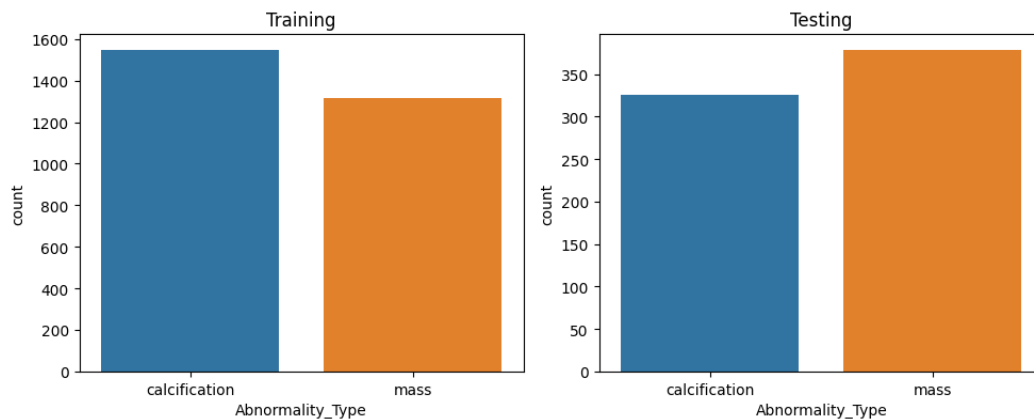


Figura 4.1: Tipo de anomalía

Comenzando por el tipo de anomalía, se observa que ambos conjuntos están relativamente balanceados, con una proporción similar de calcificaciones y masas en ambas particiones. Si bien lo ideal sería

que la distribución fuera idéntica, esto no siempre es posible, ya que se debe mantener a los pacientes separados entre las particiones para evitar fugas de información entre los conjuntos, lo que hace que una separación estratificada clásica no sea factible.

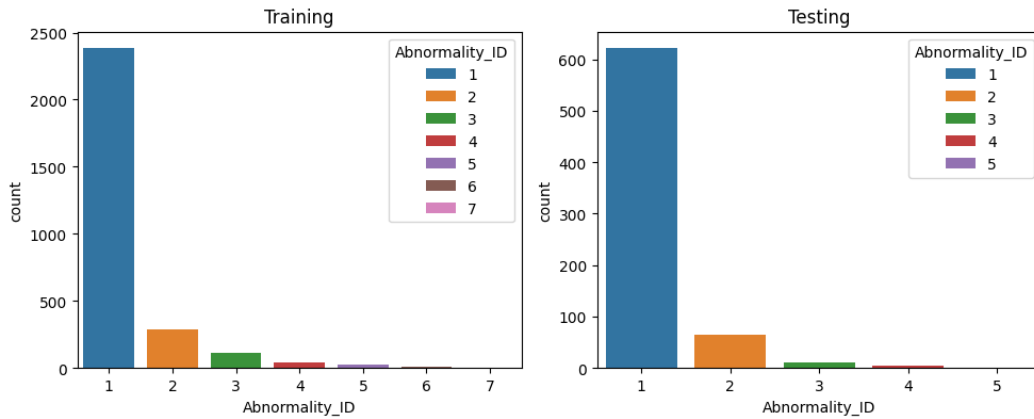


Figura 4.2: Cantidad de anomalías

Por otra parte, el “Abnormality ID” nos indica la cantidad de anomalías presentes por paciente. Esto quiere decir que para la inmensa mayoría de los casos de CBIS-DDSM, las mamografías presentan una única lesión (calcificación o masa), y los casos con más de 2 lesiones para un mismo paciente son bastante reducidos para ambos conjuntos.

4.1.2. Patologías

Además del tipo de anomalía con la que se puede trabajar, cada lesión encontrada dentro de las mamografías está asociada a una patología. Esta patología corresponde al diagnóstico realizado por un especialista y también a las clases que el sistema desarrollado buscará predecir una vez entrenado.

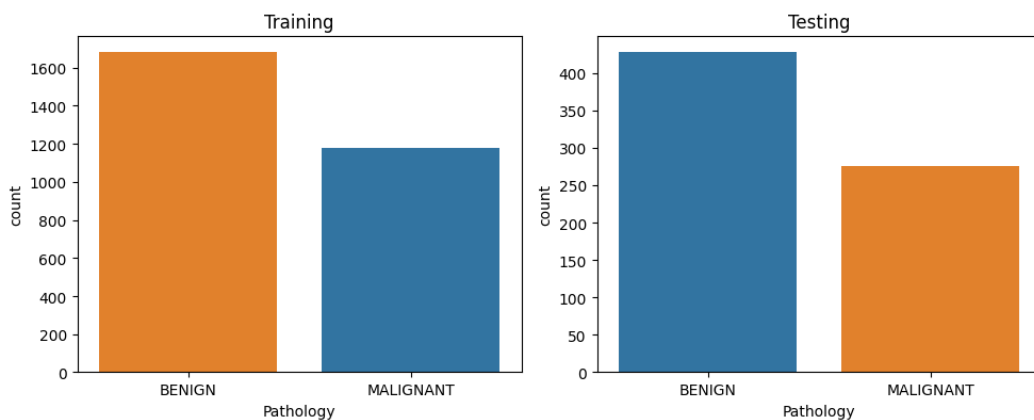


Figura 4.3: Patologías

Se puede observar que las patologías están estratificadas, buscando mantener la misma distribución en los conjuntos de *training* y *testing*. Además, es evidente que existen más casos benignos que malignos, lo cual es coherente con lo esperado, dado el progreso natural de la enfermedad.

4.1.3. Evaluación

Además de la patología, cada especialista entrega un nivel de BIRADS para cada mamografía. El nivel de BIRADS es una estandarización para el reporte de resultados de cáncer de mama, como se mencionó en secciones anteriores.

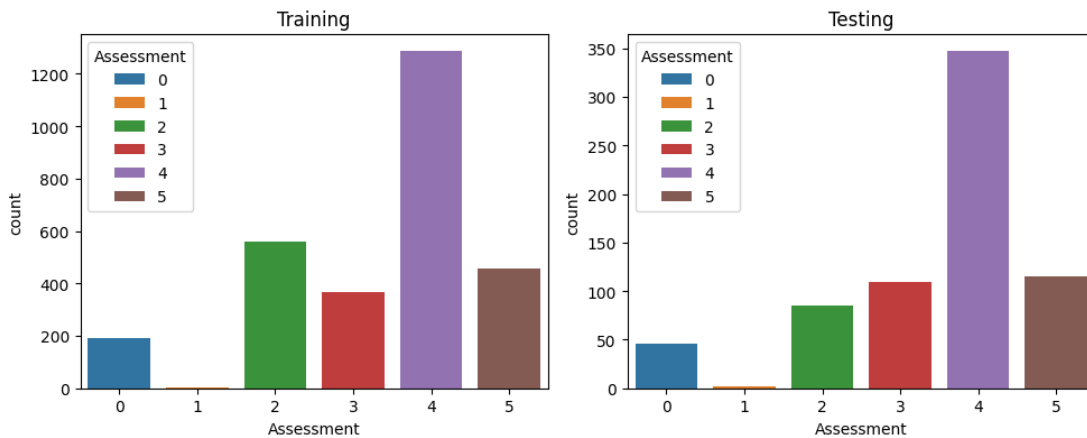


Figura 4.4: Evaluaciones

Cabe mencionar que esta evaluación es hecha antes de tener una patología confirmada, es decir, es el primer paso hacia el análisis que hace un radiólogo, de ahí la prevalencia del nivel 4 (sospechoso) en el dataset. Una vez la zona sospechosa fue analizada por otros métodos, se obtiene la patología del examen. Es interesante también notar que el nivel 1 (negativo) presenta una cantidad no descartable de casos, indicando que el análisis inicial del radiólogo fue erróneo, y un sistema como el desarrollado podría ser una herramienta útil para el diagnóstico asistido.

4.2. Preprocesamiento de los Datos

El trabajo desarrollado se basa en las imágenes disponibles en CBIS-DDSM, por lo que las siguientes secciones se refieren estrictamente al preprocesamiento de este dataset.

Se comienza descartando todas aquellas imágenes que tienen distinto tamaño para el examen de mamografía y la máscara, indicando que existe algún problema con la máscara de segmentación de estas. De esta forma, se descarta un total de 80 imágenes.

Luego se juntan todas las imágenes de calcificaciones y masas, que originalmente vienen separadas, pero manteniendo la separación original de *training* y *testing*. Esto nos entrega un total de 2485 imágenes de mamografía distintas, con un total de 2799 máscaras de segmentación para el conjunto de *training*, y 631 imágenes con 689 máscaras en el caso del conjunto de *testing*. La discrepancia entre la cantidad de exámenes y las máscaras viene dada por el hecho de que en algunos exámenes se pueden apreciar 2 o más lesiones, entregando más de un solo archivo de máscaras en estos casos.

	Cantidad de Imágenes	Mascaras de Segmentación
Training	2485	2799
Testing	631	689

Tabla 4.1: Cantidad de imágenes y máscaras de segmentación por conjunto de datos.

Habiendo ya unificado el dataset, se procede a tomar todas aquellas mamografías con más de una máscara de segmentación y se juntan todas en un solo archivo DICOM, cuyo resultado puede apreciarse en la Figura 4.5

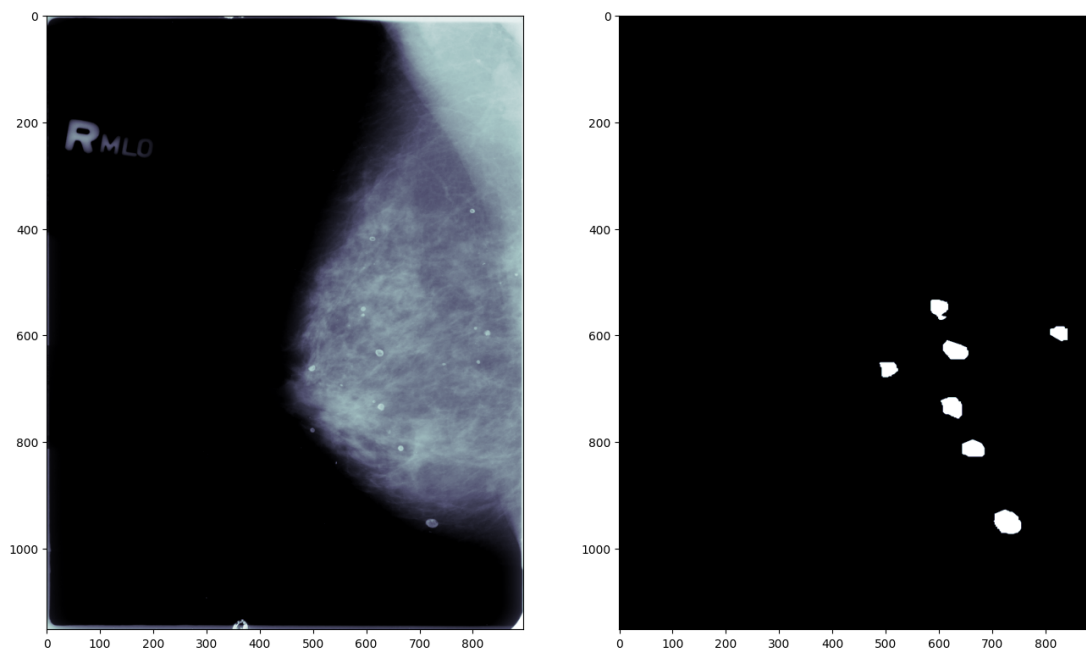


Figura 4.5: Ejemplo máscaras de segmentación estandarizadas.

Finalmente, se procede a separar las imágenes en conjuntos de *training*, *testing* y *validation*. Los conjuntos de *training* y *testing* ya vienen definidos en el CBIS-DDSM. Para el trabajo desarrollado, mantendremos esta separación estándar. Para el conjunto de validación, se divide el conjunto de *training* en una proporción de 80:20 para entrenamiento y validación, respectivamente. La división se realiza de forma estratificada, manteniendo la distribución de patologías que presenta el conjunto de entrenamiento.

La cantidad de imágenes en cada conjunto puede apreciarse en la Tabla 4.2

	Cantidad de Imágenes
Training	1938
Validation	547
Testing	631

Tabla 4.2: Cantidad de imágenes por conjunto de datos.

Finalmente, se procede a estandarizar el tamaño de las imágenes, ajustándolas a 1152x896 píxeles. Este proceso también se aplica a las máscaras de segmentación. Luego, se guardan las imágenes redimensionadas en formato PNG para su uso en las etapas posteriores. Una vez completado este proceso, se obtienen las cajas englobantes para cada lesión a partir de las máscaras de segmentación que proporciona el dataset. Estas cajas englobantes serán utilizadas posteriormente para la generación de los datasets de parches. El resultado de este proceso puede observarse en la Figura 4.6.

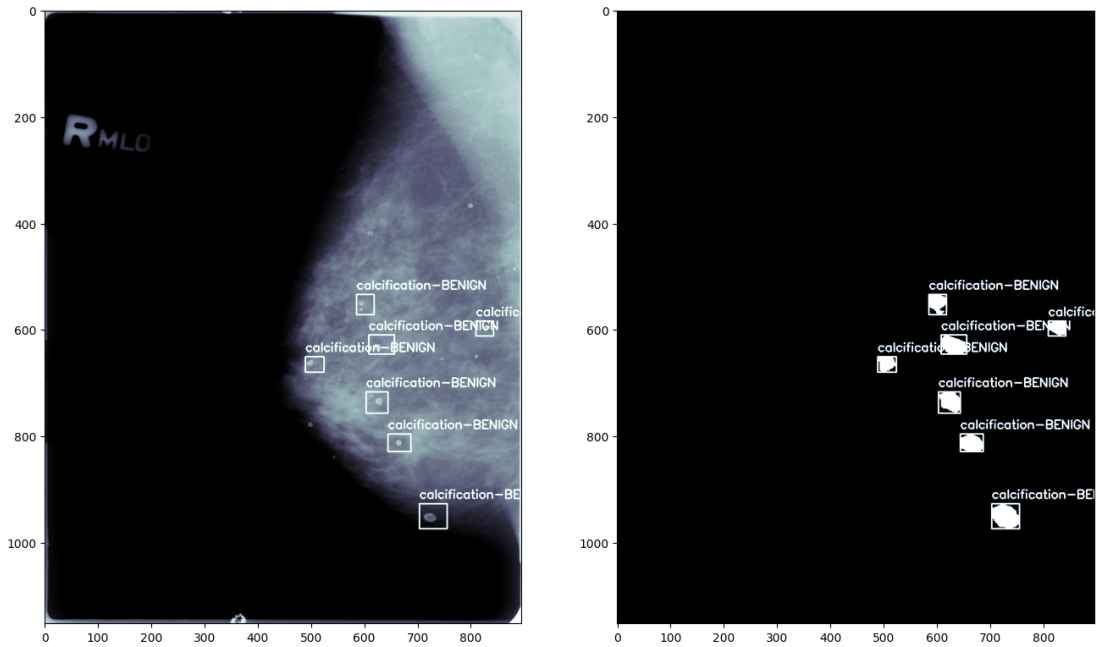


Figura 4.6: Cajas englobantes generadas.

4.2.1. Cajas Englobantes

Antes de generar los datasets de parches, es aconsejable estudiar las cajas englobantes generadas en el paso anterior. Esto es esencial para seleccionar un tamaño de parche adecuado para los datos disponibles. A partir de las cajas englobantes generadas durante el preprocesamiento de los datos, se obtienen los histogramas presentados en la Figura 4.7, que muestran la distribución de alturas y anchos de

las lesiones, para los conjuntos de *training* y *testing*, respectivamente. La línea roja muestra el promedio, y las líneas punteadas muestran la desviación estándar.

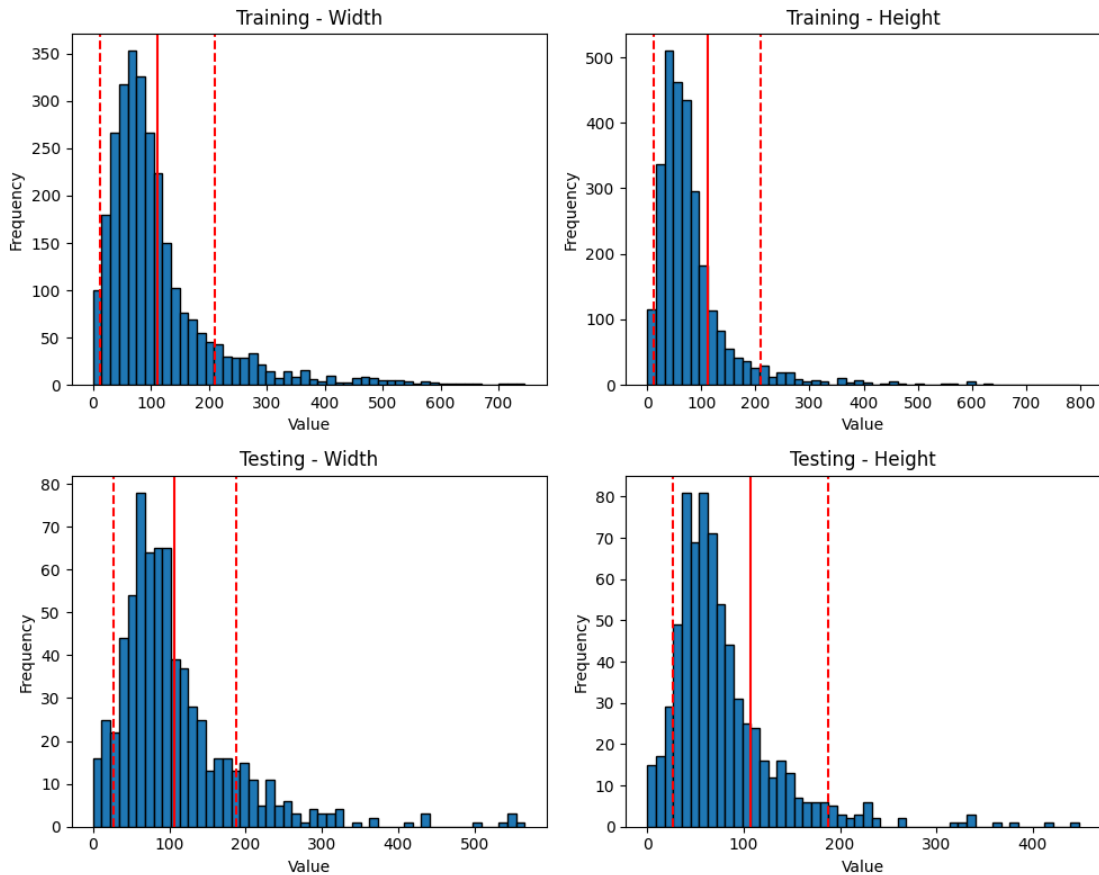


Figura 4.7: Histogramas de altura y ancho para las cajas englobantes generadas.

Es a partir de este análisis que se elige un tamaño de parches de 224x224.

4.3. Generación de los Datasets

Siguiendo el trabajo realizado en [8], primero se entrenará una red utilizando parches más pequeños obtenidos a partir de las cajas englobantes. Luego, se procederá a entrenar la red utilizando las imágenes completas. Esta sección explicará en mayor detalle el método de obtención de estos parches, los cuales pueden pertenecer a una de las siguientes clases: fondo, calcificación benigna, calcificación maligna, masa benigna y masa maligna. En este contexto, 'fondo' se refiere a una porción del examen en donde no se encuentra ninguna lesión.

4.3.1. Dataset S1

El primer dataset, S1, genera un total de 2 parches por cada lesión. El primero de estos parches busca centrar el área de interés (ROI), que corresponde a la ubicación de la lesión, obteniéndola lo más centrada posible. El segundo parche corresponde a una zona fuera del ROI, para obtener un parche de fondo. Los parches generados tienen un tamaño de 224x224. Un ejemplo de los parches generados se puede ver en la Figura 4.8.

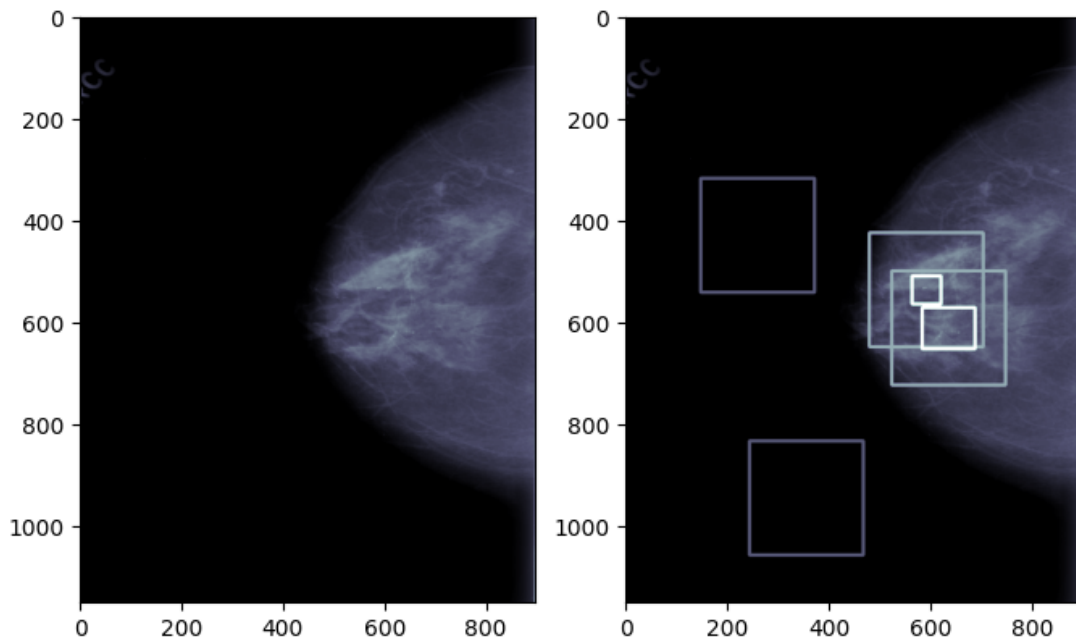


Figura 4.8: Ejemplo de *sampling* de parches para dataset S1. Los cuadrados en blanco corresponden al *ground-truth* de la lesión. Los cuadrados en gris claro corresponden a los parches generados para las lesiones, y los cuadrados en gris oscuro corresponden a los parches de fondo.

Un ejemplo de las imágenes resultantes de la generación de parches puede verse en la Figura 4.9

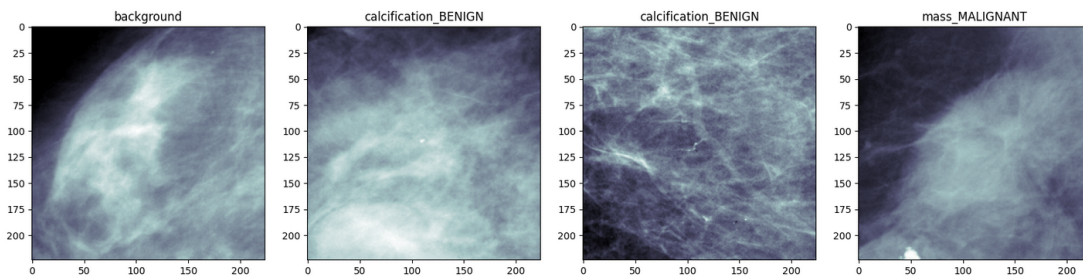


Figura 4.9: Parches generados.

4.3.2. Dataset S10

También se crea el dataset S10, el cual consiste en tomar un total de 10 parches de 224x224 por cada lesión. Estos se toman en el área de la lesión, similar al dataset S1, con la diferencia de que ahora no se encuentran centrados, sino que se toman aleatoriamente dentro de la región de interés, y se permite que incluyan cierta cantidad de fondo, manteniendo un mínimo de 90 % de *overlap* para el parche y el ROI. Además, se toman 10 parches de fondo aleatoriamente en la imagen.

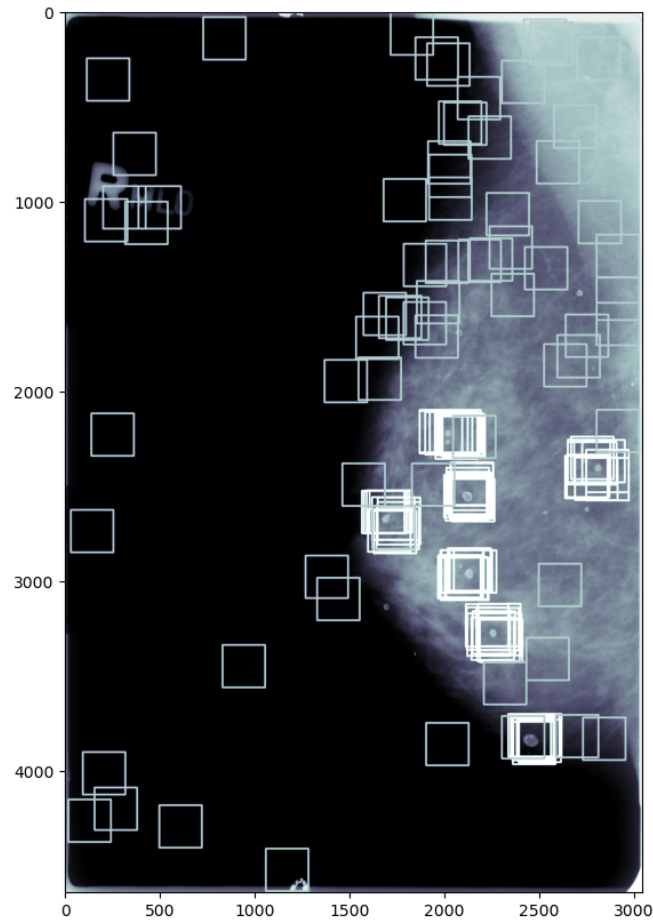


Figura 4.10: Ejemplo de *sampling* de parches para dataset S10.

4.4. Red Neuronal de Parches

Con los datasets ya definidos, en esta sección se detalla la creación y proceso de entrenamiento para las redes neuronales que se utilizan en el desarrollo de la solución. Siguiendo el trabajo realizado en [8], se entrenan los modelos utilizando los parches generados bajo las técnicas descritas para S1 y S10. Estos

modelos se entrenan inicialmente en los parches, permitiendo que la red aprenda a clasificar las lesiones en sus respectivas clases. Luego, estos modelos son re-entrenados mediante *transfer learning* utilizando las imágenes completas de mamografía. El objetivo es aprovechar el aprendizaje previo de las redes para la clasificación de parches y aplicarlo al análisis de imágenes completas de mamografía. De esta manera, se crea una solución que no depende de las anotaciones de un radiólogo y que puede proporcionar un diagnóstico preliminar más rápido y eficiente.

Siguiendo este esquema se procede a entrenar 3 modelos con arquitecturas ResNet50 [8], Xception [11] y EfficientNetV2.

4.4.1. Preprocesamiento y Aumentación de Datos

Para el preprocesamiento, se resta la imagen promedio del conjunto de ImageNet a todas las imágenes, lo que nos permite “centrar” nuestros datos al modelo pre-entrenado. No se aplican otras técnicas de preprocesamiento para los datos. Además, al momento de entrenar la red se aplican las siguientes transformaciones para la aumentación de datos:

- Volteado Horizontal
- Volteado Vertical

4.4.2. Entrenamiento

Para el entrenamiento, se utiliza el optimizador Adam, y las redes se cargan con sus respectivos pesos pre-entrenados en ImageNet. Cada modelo es entrenado durante un total de 50 épocas, con una reducción del *learning rate* cuando la pérdida de validación no disminuye por un total de 5 épocas consecutivas. Se implementa un *early stopping* con una paciencia de 10 épocas, lo que significa que si la pérdida de validación no mejora durante 10 épocas consecutivas, el entrenamiento se detiene prematuramente. Los mejores pesos obtenidos durante el entrenamiento se conservan y se utilizan para realizar las predicciones finales. Para el entrenamiento, se define una época como una pasada completa por todos los datos de entrenamiento, y se utiliza un *batch size* de 64 para el entrenamiento de los modelos.

4.5. Red Neuronal de Imágenes Completas

A partir de los modelos de parches previamente entrenados, se procede a realizar *transfer learning* para que el modelo pueda ser utilizado con imágenes de mamografía completas.

En este proceso, se utilizan únicamente las imágenes completas de mamografías, sin hacer uso de máscaras de segmentación o cajas englobantes.

Para realizar el *transfer learning*, se aprovecha la naturaleza de las redes convolucionales, las cuales son agnósticas al tamaño de la imagen de entrada. Esto significa que, para una imagen más grande, simplemente se aplicarán los filtros convolucionales una mayor cantidad de veces. El verdadero desafío al realizar este cambio hacia imágenes completas radica en las capas densas de la red, las cuales deben recibir una entrada de tamaño fijo.

Por lo tanto, para implementar este cambio, se descartan las capas densas que fueron entrenadas para los parches, manteniendo únicamente la sección convolucional de cada modelo. Luego, se agregan nuevas capas densas adaptadas para manejar la entrada de las imágenes completas de mamografía. Esta estrategia permite que el modelo aproveche las características aprendidas en los parches y las transfiera eficazmente a la tarea de clasificación de imágenes completas.

4.5.1. Preprocesamiento y Aumentación de Datos

En este caso, se vuelve a restar la imagen promedio de ImageNet a todas las imágenes, lo que nos permite “centrar” nuestros datos. Además, se aplican las siguientes transformaciones para la aumentación de datos:

- Volteado Horizontal

No se pueden utilizar técnicas más agresivas de aumentación de datos debido a la naturaleza de los datos con los que estamos trabajando. Las imágenes de mamografía son estándares, lo que quiere decir que nunca encontraremos imágenes de mamografía rotadas o con cambios de intensidad. Un volteado horizontal es una aumentación más realista al caso de aplicación en el que se utiliza el modelo desarrollado.

4.5.2. Entrenamiento

Para el entrenamiento, se utiliza el optimizador Adam, y las redes se cargan con sus respectivos pesos pre-entrenados en los datos de parches (S1 y S10). Nuevamente, cada modelo es entrenado durante un total de 50 épocas, con una reducción del *learning rate* y *early stopping*. Los mejores pesos obtenidos durante el entrenamiento se conservan y se utilizan para realizar las predicciones finales. Dada una limitación en poder de cómputo, se utiliza un *batch size* de 8 para el entrenamiento de los modelos en esta etapa.

4.6. Metamodelo Ensemble Learning

Con los modelos previamente entrenados utilizando imágenes completas de mamografías, se selecciona el mejor modelo para cada arquitectura y se combinan en un metamodelo basado en la técnica

de *stacking*. Este enfoque se inspira en el proceso de diagnóstico clínico, donde las radiografías suelen ser revisadas por un grupo de especialistas, como radiólogos, médicos y oncólogos, cuya experiencia y opiniones contribuyen al diagnóstico final.

De forma similar, cada modelo entrenado en este trabajo actúa como un “especialista” que ofrece una perspectiva única, y el metamodelo actúa como un “panel de expertos”. Este panel analiza las predicciones individuales y las combina para generar una predicción final más robusta y precisa.

Para implementar el *ensemble learning*, se consideraron las siguientes estrategias:

- **Promedio Simple:** Para una imagen de mamografía cualquiera, cada modelo produce una predicción independiente. Las salidas se combinan calculando el promedio aritmético para obtener la predicción final.
- **Promedio Ponderado:** Se introduce una capa adicional entrenable en el metamodelo para combinar las predicciones individuales de cada modelo base. Esta capa actúa como un integrador, aprendiendo los pesos óptimos que deben asignarse a las salidas de los modelos base. Esto permite que el metamodelo ajuste dinámicamente la contribución de cada red neuronal según su desempeño y su capacidad para manejar las diferentes características de los datos.

Capítulo 5

Resultados

5.1. Métricas

Se pretende introducir brevemente las métricas y herramientas que se utilizan para medir la precisión y calidad de los modelos de clasificación generados para la solución.

1. **Precision:** La precisión busca determinar qué proporción de las clasificaciones fue correcto. Matemáticamente se define como $\frac{TP}{TP+FP}$
2. **Recall:** Nos permite buscar la cantidad de positivos que fueron clasificados correctamente. En el contexto médico esta métrica toma especial importancia al considerar las clasificaciones como diagnósticos de una enfermedad. Se define de la forma $\frac{TP}{TP+FN}$
3. **AUC:** Se define como el área bajo una curva ROC, permitiendo medir de forma simple el desempeño de nuestro clasificador a través de todos los posibles umbrales para la curva ROC.
4. **Matriz de Confusión:** Consiste en una presentación en forma de tabla o matriz, que permite la visualización de errores en la clasificación de clases. Cada fila de la matriz representa las instancias de clases correctas, mientras que las columnas representan las clases predichas, o vice-versa. Esto permite ver rápidamente si el sistema está confundiendo clases entre sí, permitiendo identificar posibles puntos débiles. Una matriz de confusión ideal tendría un valor de 1 en su diagonal.

5.2. Resultados EfficientNetV2

5.2.1. Parches

A continuación se detallan los resultados obtenidos por la arquitectura EfficientNetV2 para los datasets S1 y S10.

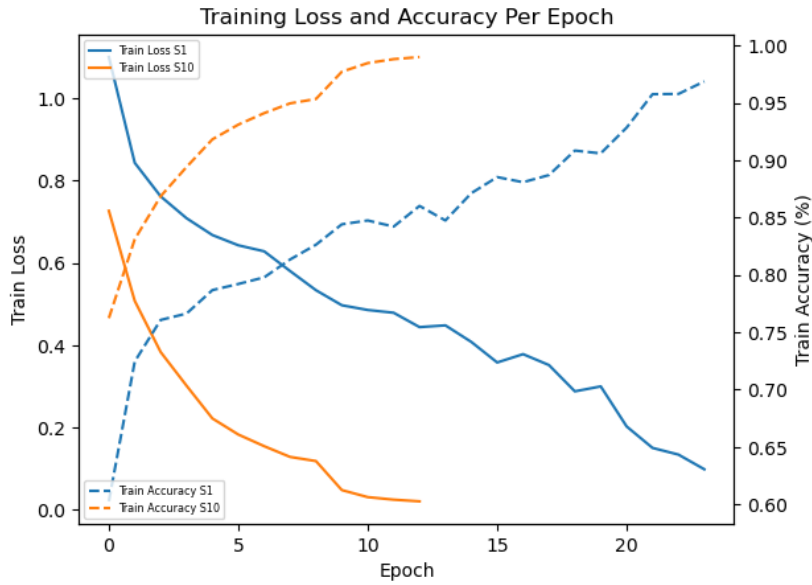


Figura 5.1: *Loss* y *Accuracy* sobre el conjunto de entrenamiento para EfficientNetV2 sobre los datasets de parches.

Se puede apreciar que en ambos experimentos la red es capaz de aprender con bastante precisión a identificar lesiones de mamografía. Además, vemos que el entrenamiento con S10 hace que la red aprenda estas características mejor y más rápido. Esto es esperable ya que este dataset tiene una mayor cantidad y variedad de lesiones comparado a S1.

A partir de las métricas obtenidas sobre el conjunto de validación, es posible observar que para el dataset S10, la red muestra signos claros de sobreajuste al conjunto de entrenamiento, evidenciado por un incremento constante en la pérdida de validación, incluso cuando la pérdida de entrenamiento continúa disminuyendo. En la Tabla 5.1, se presentan en detalle las métricas obtenidas por cada modelo.

		Loss	Accuracy	AUC	Precision	Recall
S1	Training	0.4483	0.8476	0.9768	0.7828	0.8230
	Validation	0.5543	0.8048	0.9533	0.7163	0.7209
S10	Training	0.3835	0.8686	0.9820	0.8140	0.8540
	Validation	0.5859	0.7998	0.9528	0.7167	0.7424

Tabla 5.1: Métricas obtenidas por EfficientNetV2 para entrenamiento sobre parches. Se presentan únicamente los resultados del modelo seleccionado (el que presenta *validation loss* más bajo).

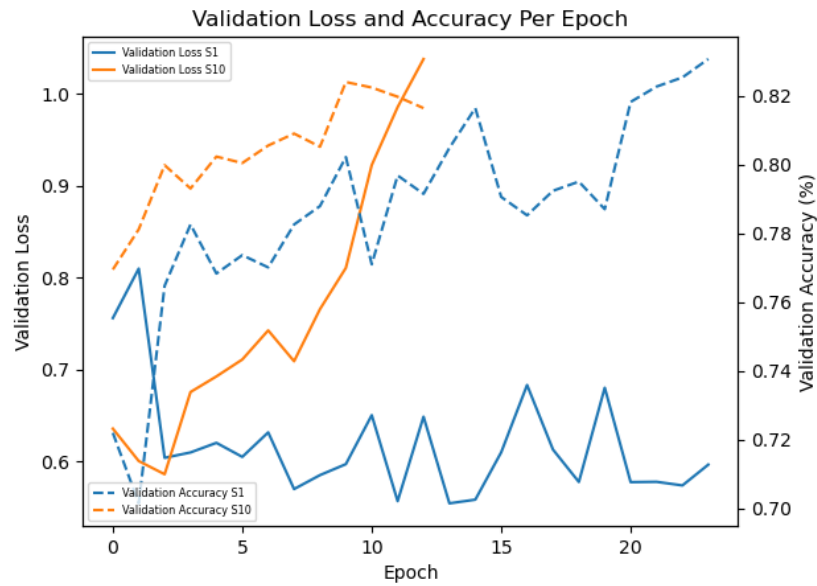


Figura 5.2: *Loss* y *Accuracy* sobre el conjunto de validación para EfficientNetV2 sobre los datasets de parches.

5.2.2. Mamografías

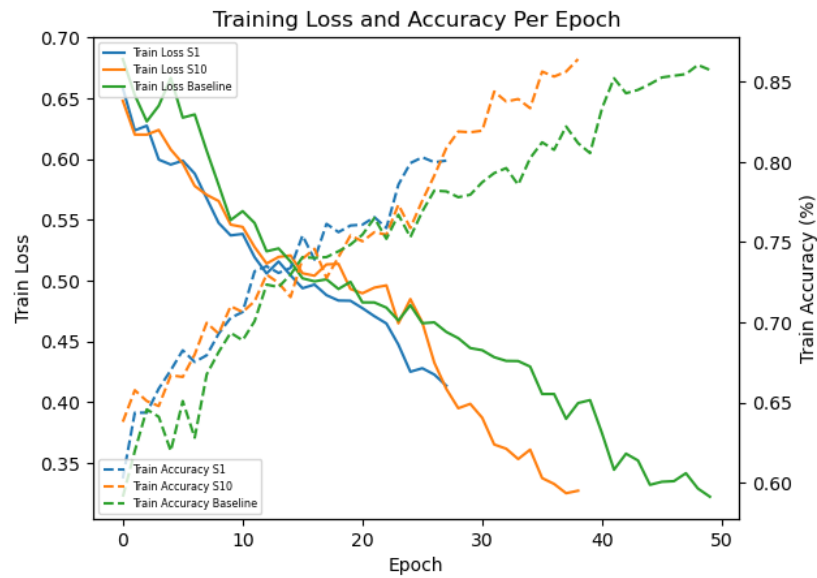


Figura 5.3: *Loss* y *Accuracy* sobre el conjunto de entrenamiento para EfficientNetV2 sobre los datasets de mamografías.

Para los resultados obtenidos sobre el conjunto de imágenes de mamografías, es posible apreciar que tanto los modelos pre-entrenados en parches, como el modelo *baseline*, que corresponde a una red

EfficientNetV2 entrenada directamente con las imágenes de mamografía, alcanzan resultados similares, aunque las redes pre-entrenadas convergen más rápido.

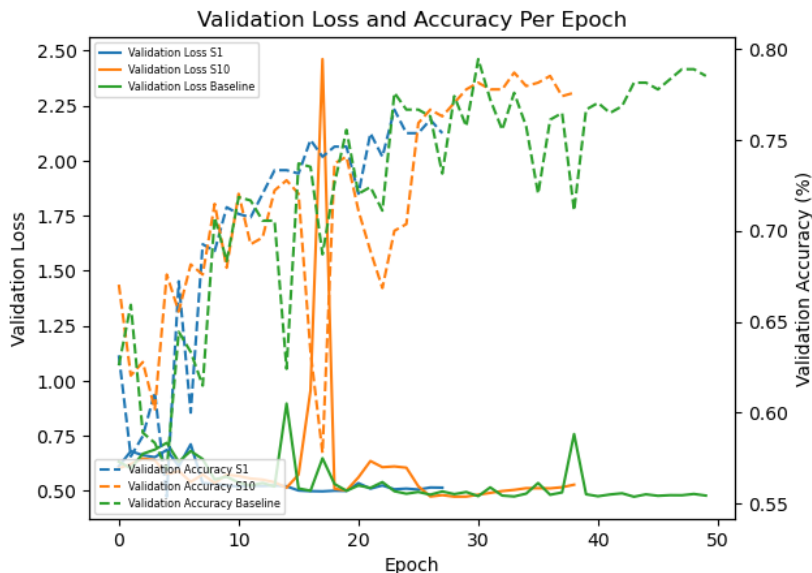


Figura 5.4: *Loss* y *Accuracy* sobre el conjunto de validación para EfficientNetV2 sobre los datasets de mamografías.

A diferencia de lo ocurrido durante el entrenamiento de parches, es posible apreciar que las redes entrenadas en imágenes completas no se sobreajustan a los datos de entrenamiento, en ninguno de los 3 casos.

		Loss	Accuracy	AUC	Precision	Recall
Pre entrenado en S1	Training	0.4881	0.7614	0.8415	0.7608	0.7633
	Validation	0.4975	0.7407	0.8336	0.7343	0.7331
Pre entrenado en S10	Training	0.3952	0.8190	0.9015	0.8178	0.8205
	Validation	0.4735	0.7704	0.8560	0.7651	0.7623
Baseline	Training	0.4019	0.8055	0.8972	0.8038	0.8058
	Validation	0.4842	0.7667	0.8498	0.7610	0.7597

Tabla 5.2: Métricas obtenidas por EfficientNetV2 para entrenamiento sobre imágenes de mamografía. Se presentan únicamente los resultados del modelo seleccionado (el que presenta *validation loss* más bajo).

5.2.3. Testing

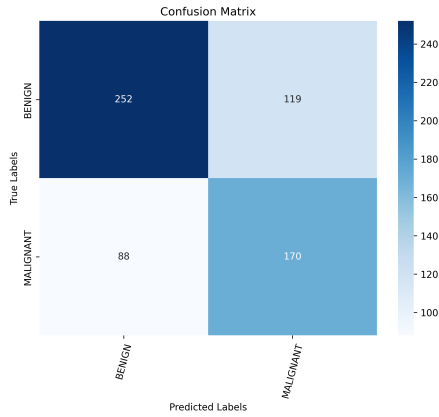
En la Tabla 5.3 se presentan los resultados obtenidos por cada modelo sobre el conjunto de *testing*.

	Loss	Accuracy	AUC	Precision	Recall
EfficientNetV2 S1 (Parches)	0.6220	0.7781	0.9507	0.6831	0.6834
EfficientNetV2 S10 (Parches)	0.6818	0.7737	0.9409	0.6808	0.6942
EfficientNetV2 (Pretrained S1)	0.6176	0.6709	0.7516	0.6647	0.6691
EfficientNetV2 (Pretrained S10)	0.6372	0.6804	0.7683	0.6719	0.6748
EfficientNetV2 Baseline	0.7054	0.6582	0.7411	0.6546	0.6595

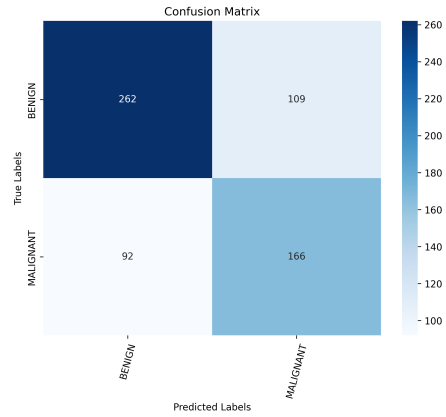
Tabla 5.3: Metricas obtenidas por EfficientNetV2 sobre el conjunto de testing para imagenes de mamografia.

Es posible apreciar que los modelos pre-entrenados en parches obtienen un mejor desempeño que el *baseline*, que consiste en un modelo entrenado directamente en imágenes de mamografía, lo que indica que el traspaso de lo aprendido por las redes sobre los parches funciona.

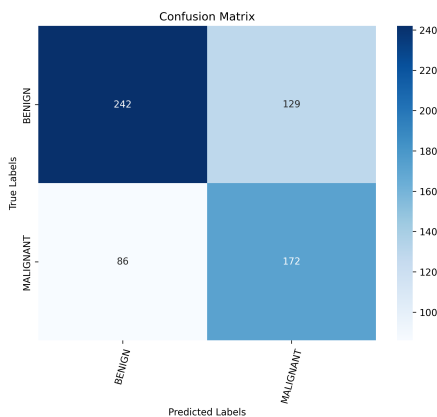
Finalmente, se presentan las matrices de confusión para los 3 modelos generados:



(a) Matriz de confusión para EfficientNetV2 pre-entrenado en S1.



(b) Matriz de confusión para EfficientNetV2 pre-entrenado en S10.



(c) Matriz de confusión para EfficientNetV2 *baseline*.

Figura 5.5: Matrices de confusión para EfficientNetV2 sobre imágenes de mamografía.

5.3. Resultados ResNet50

5.3.1. Parches

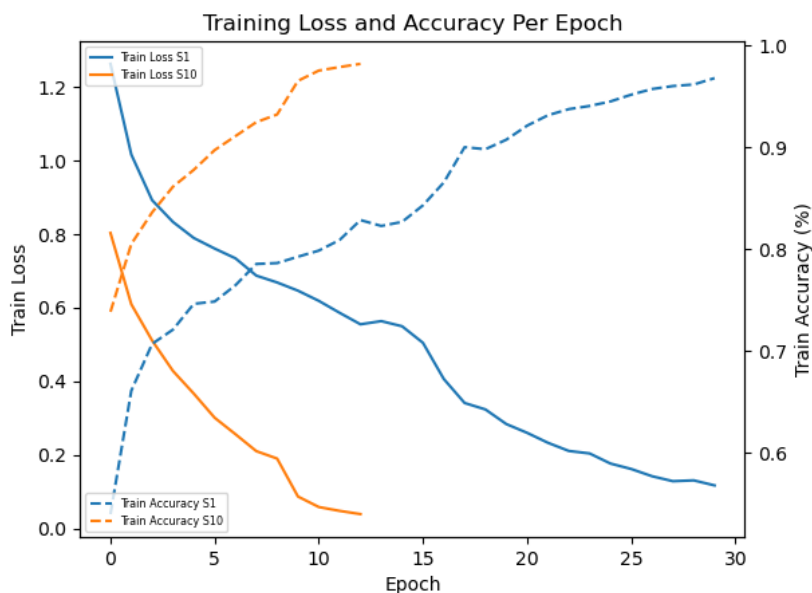


Figura 5.6: *Loss* y *Accuracy* sobre el conjunto de entrenamiento para ResNet50 sobre los datasets de parches.

Para el entrenamiento sobre parches de ResNet50, se puede apreciar el mismo patrón de entrenamiento que el visto en EfficientNetV2. Ambos modelos alcanzan una alta precisión, en donde se puede ver que el entrenado en S10 converge más rápido dada la naturaleza y cantidad de los datos.

Nuevamente es posible apreciar que ambas redes convergen cerca del 80 % de precisión, presentando casi el mismo comportamiento de sobreajuste para S10 que se pudo apreciar en EfficientNetV2.

		Loss	Accuracy	AUC	Precision	Recall
S1	Training	0.2834	0.9077	0.9908	0.8650	0.8889
	Validation	0.5334	0.8066	0.9577	0.7144	0.7205
S10	Training	0.5103	0.8362	0.9718	0.7669	0.8009
	Validation	0.5471	0.8002	0.9524	0.7208	0.7182

Tabla 5.4: Métricas obtenidas por ResNet50 para entrenamiento sobre parches. Se presentan únicamente los resultados del modelo seleccionado (el que presenta *validation loss* más bajo).

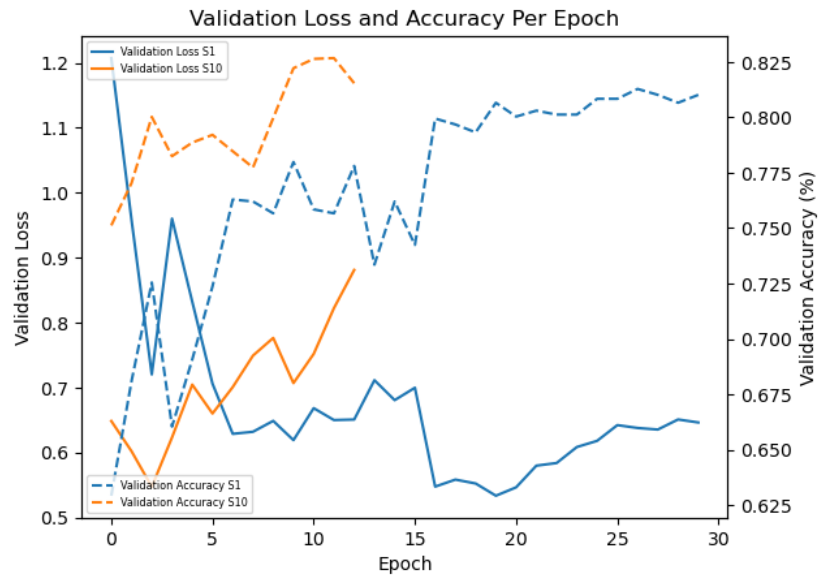


Figura 5.7: *Loss* y *Accuracy* sobre el conjunto de validación para ResNet50 sobre los datasets de parches.

5.3.2. Mamografías

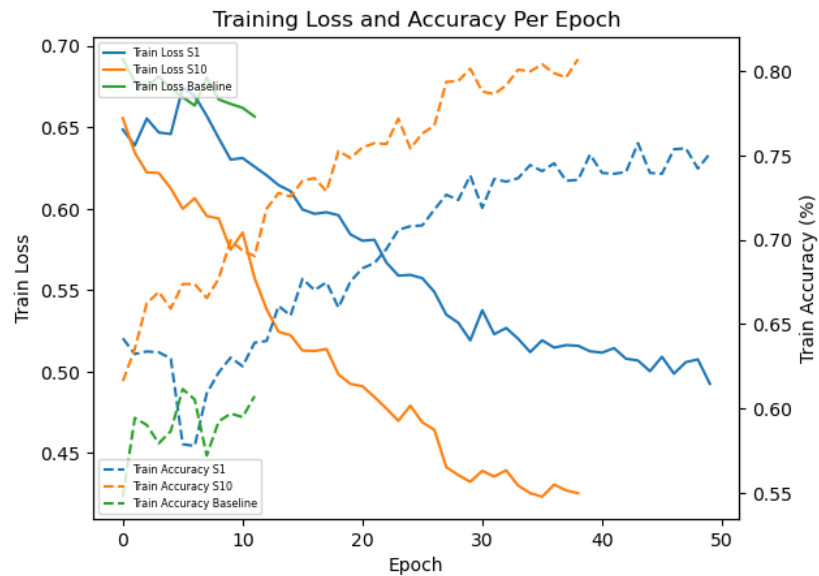


Figura 5.8: *Loss* y *Accuracy* sobre el conjunto de entrenamiento para ResNet50 sobre los datasets de mamografías.

Para ResNet50, vemos que la inclusión del pre-entrenamiento en parches hace una gran diferencia en lo que puede aprender la red. El modelo *baseline* se queda bastante detrás de los modelos pre-entrenados, indicando que es un problema demasiado complejo para una red ResNet50 sin la técnica de parches.

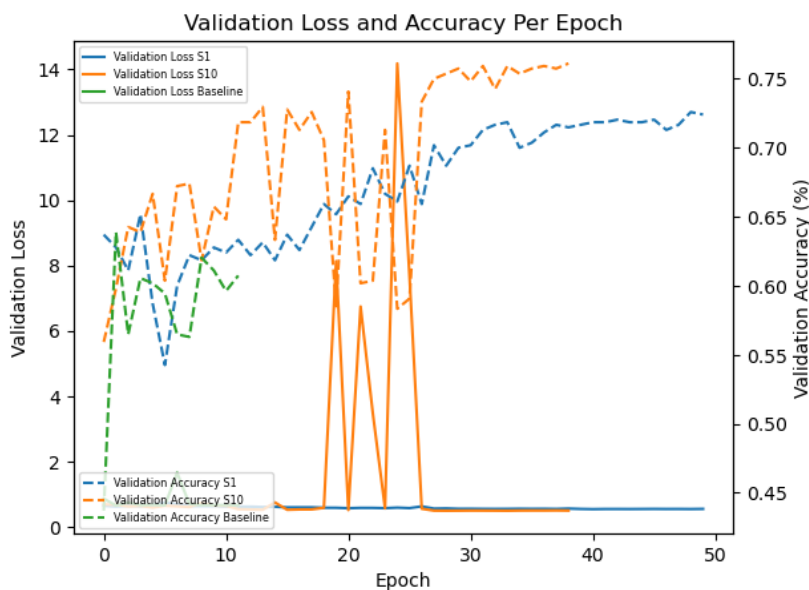


Figura 5.9: *Loss* y *Accuracy* sobre el conjunto de validación para ResNet50 sobre los datasets de mamografías.

		Loss	Accuracy	AUC	Precision	Recall
Pre entrenado en S1	Training	0.5125	0.7505	0.8239	0.7483	0.7494
	Validation	0.5490	0.7167	0.7865	0.7096	0.7093
Pre entrenado en S10	Training	0.4365	0.7941	0.8794	0.7925	0.7948
	Validation	0.4943	0.7537	0.8390	0.7479	0.7502
Baseline	Training	0.6778	0.5943	0.6075	0.5879	0.5846
	Validation	0.6514	0.6389	0.6671	0.6458	0.5936

Tabla 5.5: Métricas obtenidas por ResNet50 para entrenamiento sobre imágenes de mamografía. Se presentan únicamente los resultados del modelo seleccionado (el que presenta *validation loss* más bajo).

5.3.3. Testing

En la Tabla 5.3 se presentan los resultados obtenidos por cada modelo sobre el conjunto de testing.

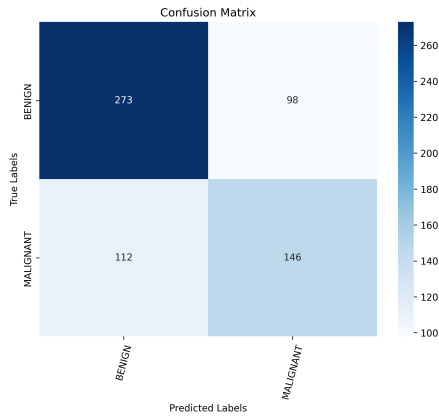
	Loss	Accuracy	AUC	Precision	Recall
ResNet50 S1 (Parches)	0.6756	0.7730	0.9450	0.6660	0.6762
ResNet50 S10 (Parches)	0.6482	0.7509	0.9365	0.6486	0.6552
ResNet50 (Pretrained S1)	0.6021	0.6661	0.7276	0.6537	0.6509
ResNet50 (Pretrained S10)	0.6440	0.6804	0.7537	0.6784	0.6842
ResNet50 Baseline	0.7035	0.6216	0.5824	0.6107	0.5641

Tabla 5.6: Métricas obtenidas por ResNet50 sobre el conjunto de *testing* para imágenes de mamografía.

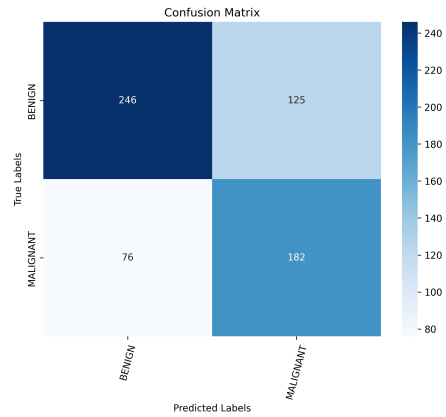
Aquí, se hace aún más evidente la diferencia entre el pre-entrenamiento hecho con parches por sobre el entrenar directamente el modelo con imágenes de mamografía, obteniendo mejores resultados para el

primer caso, tanto con el dataset S1 como S10.

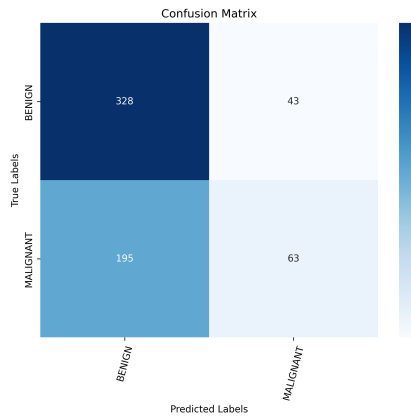
Finalmente, se presentan las matrices de confusión para los 3 modelos generados:



(a) Matriz de confusión para ResNet50 pre-entrenado en S1.



(b) Matriz de confusión para ResNet50 pre-entrenado en S10.



(c) Matriz de confusión para ResNet50 *baseline*.

Figura 5.10: Matrices de confusión para ResNet50 sobre imágenes de mamografía.

A partir de las matrices de confusión, es posible observar que para el caso de *Baseline* el modelo predice casi todas las imágenes como “Benignas”, quedando en evidencia que el modelo no fue capaz de aprender las características que diferencian a ambas clases, y demostrando que el enfoque basado en parches es necesario para una arquitectura como ResNet50.

5.4. Resultados Xception

5.4.1. Parches

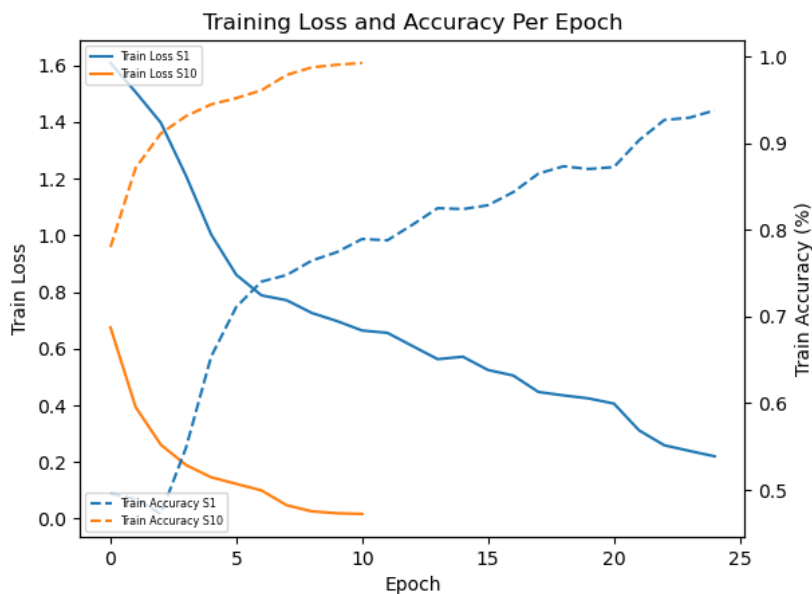


Figura 5.11: *Loss* y *Accuracy* sobre el conjunto de entrenamiento para Xception sobre los datasets de parches.

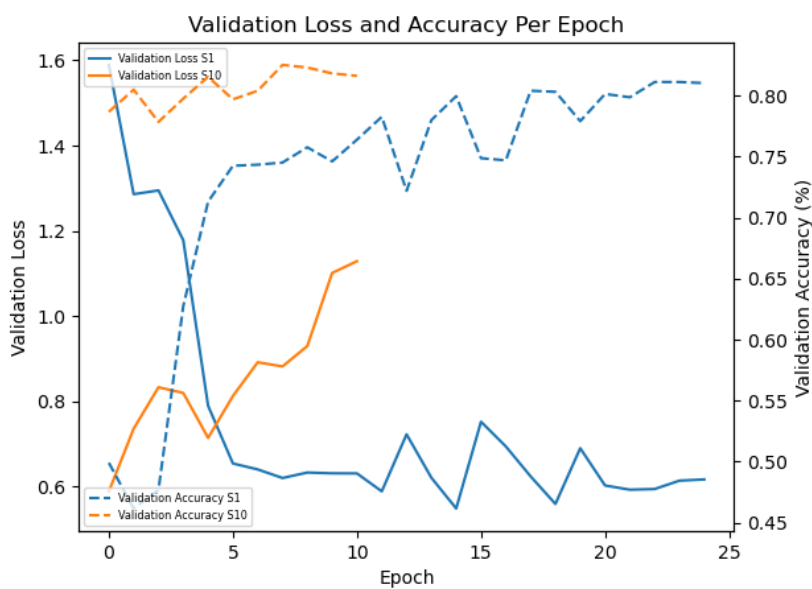


Figura 5.12: *Loss* y *Accuracy* sobre el conjunto de validación para Xception sobre los datasets de parches.

Finalmente, para Xception vemos que se vuelve a repetir el patrón de entrenamiento que hemos visto para ambos modelos anteriores en el caso de parches, pudiendo alcanzar valores para la pérdida y la precisión muy similares a las arquitecturas anteriormente presentadas.

		Loss	Accuracy	AUC	Precision	Recall
S1	Training	0.5714	0.8238	0.9662	0.7486	0.7771
	Validation	0.5481	0.7995	0.9531	0.7194	0.7256
S10	Training	0.6748	0.7799	0.9508	0.6984	0.7385
	Validation	0.5880	0.7867	0.9471	0.7087	0.7343

Tabla 5.7: Métricas obtenidas por Xception para entrenamiento sobre parches. Se presentan únicamente los resultados del modelo final seleccionado (el que presenta *validation loss* más bajo).

5.4.2. Mamografías

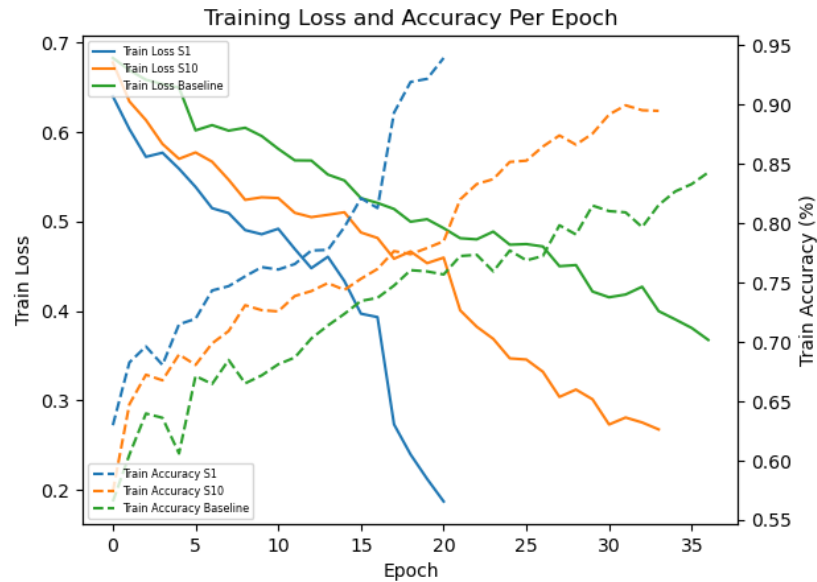


Figura 5.13: *Loss* y *Accuracy* sobre el conjunto de entrenamiento para Xception sobre los datasets de mamografía.

		Loss	Accuracy	AUC	Precision	Recall
Pre entrenado en S1	Training	0.4918	0.7609	0.8422	0.7594	0.7616
	Validation	0.4950	0.7426	0.8298	0.7383	0.7282
Pre entrenado en S10	Training	0.3688	0.8371	0.9183	0.8354	0.8364
	Validation	0.5011	0.7759	0.8481	0.7776	0.7589
Baseline	Training	0.4719	0.7723	0.8546	0.7706	0.7726
	Validation	0.4837	0.7556	0.8404	0.7497	0.7471

Tabla 5.8: Métricas obtenidas por Xception para entrenamiento sobre imágenes de mamografía. Se presentan únicamente los resultados del modelo seleccionado (el que presenta *validation loss* más bajo).

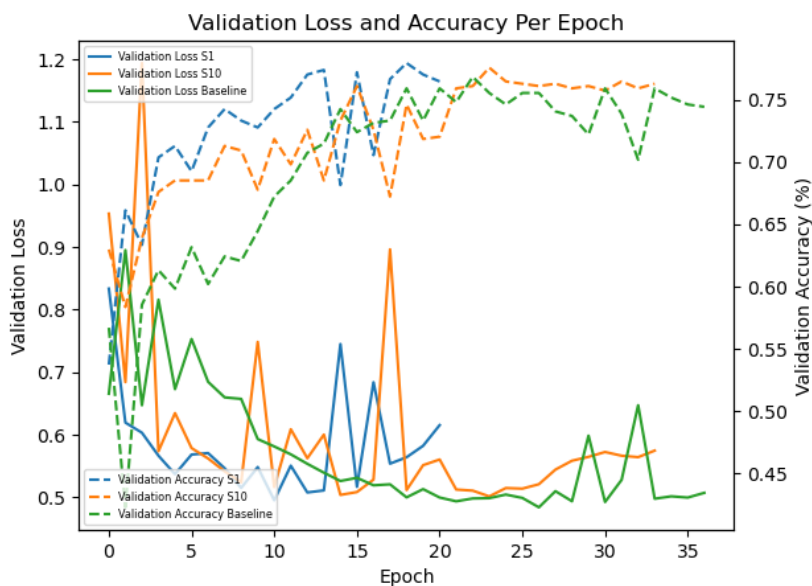


Figura 5.14: *Loss* y *Accuracy* sobre el conjunto de validación para Xception sobre los datasets de mamografía.

En la arquitectura Xception, es posible apreciar que el modelo *Baseline* no solamente es capaz de aprender el problema, sino que incluso obtiene mejores métricas que los modelos pre-entrenados en ciertos casos. Esto puede deberse al hecho de que Xception es una arquitectura diseñada con imágenes médicas en mente, por lo que entrenarlo directamente en el problema puede llevar a buenos resultados sin necesidad del pre-entrenamiento en parches.

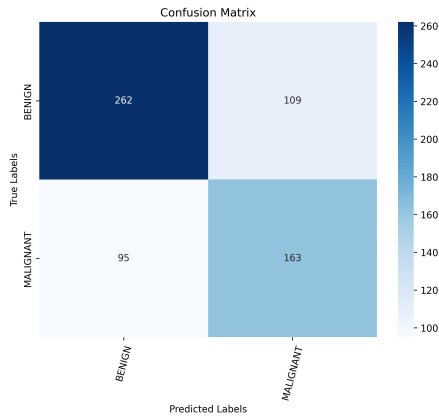
5.4.3. Testing

En la Tabla 5.9 se presentan los resultados obtenidos por cada modelo sobre el conjunto de *testing*.

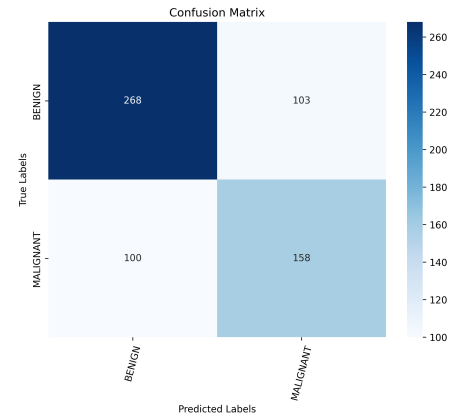
	Loss	Accuracy	AUC	Precision	Recall
Xception S1 (Parches)	0.6477	0.7614	0.9446	0.6665	0.6762
Xception S10 (Parches)	0.6159	0.7641	0.9432	0.6757	0.6970
Xception (Pretrained S1)	0.5989	0.6756	0.7629	0.6666	0.6690
Xception (Pretrained S10)	0.6773	0.6773	0.7594	0.6668	0.6674
Xception Baseline	0.5846	0.6948	0.7739	0.6863	0.6893

Tabla 5.9: Métricas obtenidas por Xception sobre el conjunto de *testing* para imágenes de mamografía.

Finalmente, se presentan las matrices de confusión para los 3 modelos generados:



(a) Matriz de confusión para Xception pre-entrenado en S1.



(b) Matriz de confusión para Xception pre-entrenado en S10.

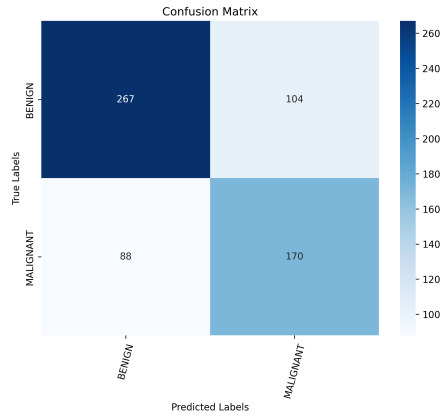
(c) Matriz de confusión para Xception *Baseline*.

Figura 5.15: Matrices de confusión para Xception sobre imágenes de mamografía.

5.5. Resultados Ensemble Learning

Finalmente, se procede a seleccionar el mejor modelo de cada arquitectura para armar un modelo de conjuntos. Se selecciona EfficientNetV2 pre-entrenado en S1, ResNet50 pre-entrenado en S10, y Xception *baseline*. Estos 3 modelos se combinan de acuerdo a las técnicas de *stacking* descritas en la sección anterior. Los resultados sobre el conjunto de *testing* pueden apreciarse en la siguiente Tabla:

	Loss	Accuracy	AUC	Precision	Recall
Promedio Simple	0.5617	0.7089	0.7914	0.7018	0.6954
Promedio Ponderado	0.5487	0.7168	0.8134	0.7125	0.7142

Tabla 5.10: Metricas obtenidas por los metamodelos de *ensemble learning* sobre el conjunto de *testing*.

Los pesos finales asignados por el metamodelo para el promedio ponderado se pueden ver en la Tabla 5.11

Modelo	Peso
EfficientNetV2 (S1)	0.1603
ResNet50 (S10)	0.2801
XCeption (Baseline)	0.5596

Tabla 5.11: Métricas obtenidas por los metamodelos de *ensemble learning* sobre el conjunto de *testing*.

A partir de esto, se observa que el metamodelo asigna un mayor peso a la arquitectura XCeption, que, precisamente, es la que obtiene los mejores resultados de forma individual. Además, los pesos asignados a ResNet50 y EfficientNetV2 no son despreciables, lo que sugiere que el metamodelo considera que la información que aportan no es redundante con la de XCeption. Es decir, contribuyen con información nueva, en línea con lo teorizado en el concepto de “panel de expertos”.

Capítulo 6

Conclusiones

En el presente capítulo se analizarán los objetivos e hipótesis definidos, con el objetivo de verificar su cumplimiento, y las razones, en caso de que corresponda, de su implementación.

6.1. Conclusiones generales

El trabajo realizado muestra que, si bien compleja, la tarea de clasificación de mamografías en etapas de avance del cáncer es posible haciendo uso de redes neuronales, como se pudo evidenciar en base al trabajo que se realizó sobre el dataset CBIS-DDSM. Más aún, se puede concluir que la tarea de identificar tipos de lesiones puede alcanzar una muy alta precisión, pudiendo distinguir sin mayor problema masas de calcificaciones. La verdadera dificultad de la tarea radica en poder asignar un estado de malignidad a estas. Si bien es posible alcanzar buenos resultados con la técnica propuesta en este trabajo, estos no son suficientes como para implementarse directamente en un sistema de salud. Esto nos lleva a concluir, además, que el poder separar las mamografías en estadios es muy poco probable que obtenga buenos resultados con el estado del arte actual, ya que corresponde a una tarea de clasificación con aún mayor granularidad que la presentada en este trabajo, pero se teoriza que sí es posible lograr un sistema que sea capaz de distinguir entre 3 posibles clases, siendo estas: “Sano”, “Benigno” y “Maligno”.

6.2. Conclusiones específicas

6.2.1. Sobre los Datasets

Vale la pena notar el hecho de que la mayoría de datasets disponibles de forma pública contienen imágenes muy antiguas, por lo que realizar un proceso de transferencia a una aplicación moderna requiere de constante vigilancia al momento de realizar la implementación. Además, se nota que ninguno

posee imágenes de mamografías para pacientes sanos, lo que puede afectar a un sistema que busca diferenciar entre tipos de pacientes. Este trabajo intentó aliviar esta problemática mediante el uso de parches extraídos de regiones en donde no había lesiones, pero se teoriza que podría implementarse una mejor solución contando con imágenes de pacientes sanos.

6.2.2. Sobre las Redes Neuronales

Resulta interesante también concluir sobre el comportamiento de cada una de las redes neuronales utilizadas en este trabajo. Las arquitecturas de EfficientNetV2, ResNet50 y Xception son capaces de aprender con muy buenos resultados a diferenciar entre los parches con lesiones, indicando que el principal problema de esta tarea radica en el utilizar imágenes de mamografías completas, en donde la lesión se encuentra en una zona bastante reducida de la imagen.

Sobre el uso de la técnica de pre-entrenar en parches, es posible concluir que para EfficientNetV2 y ResNet50 tuvo un impacto positivo, con especial notoriedad en ResNet50, obteniendo mejores resultados que los modelos “Baseline”. Esta diferencia en el impacto del pre-entrenamiento puede deberse a que EfficientNetV2 es una arquitectura más reciente que ResNet50. En esta misma línea, resulta especialmente interesante el caso de la red Xception, en donde el “Baseline” casi no presenta diferencias a los modelos pre-entrenados, llegando incluso a superarlos en algunos casos. Esto puede estar asociado al hecho de que Xception es una arquitectura creada con el propósito de trabajar con imágenes médicas, por lo que no sufre con los problemas asociados a este tipo de imágenes.

6.3. Trabajo Futuro

Para trabajo futuro, se postula el utilizar la arquitectura de Vision Transformers en esta aplicación. Este tipo de modelos ha demostrado obtener muy buenos resultados sobre ciertas tareas, y resulta interesante que la forma en que trabajan con imágenes incluya un proceso de trabajar sobre parches de la imagen (similar al trabajo realizado en esta memoria) lo que podría llevar a buenos resultados. La única limitante de este tipo de arquitecturas suele ser la cantidad de datos requeridos para poder entrenarlos, lo que, en el contexto de imágenes médicas, puede ser un problema ya que suelen ser difíciles de adquirir.

Además, sería útil explorar modelos de detección de objetos sobre las imágenes de mamografías, acercando la solución a una herramienta de diagnóstico asistido al mostrar la ubicación de la posible lesión a los médicos.

Bibliografía

- [1] Aksac A., Demetrick D.J., and Ozyer T. et al. Brecahad: a dataset for breast cancer histopathological annotation and diagnosis. *BMC Res Notes*, 12, 2019.
- [2] Abien Fred M. Agarap. On breast cancer detection: An application of machine learning algorithms on the wisconsin diagnostic dataset. In *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing*, Icmisc '18, page 5–9, New York, NY, USA, 2018. Association for Computing Machinery.
- [3] Richard J. Bleicher, Karen Ruth, Elin R. Sigurdson, J. Robert Beck, Eric Ross, Yu-Ning Wong, Sameer A. Patel, Marcia Boraas, Eric I. Chang, Neal S. Topham, and Brian L. Egleston. Time to surgery and breast cancer survival in the united states. *JAMA Oncology*, 2(3):330–339, 03 2016.
- [4] Dar-Ren Chen, Ruey-Feng Chang, Wen-Jia Kuo, Ming-Chun Chen, and Y.u-Len Huang. Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks. *Ultrasound in Medicine & Biology*, 28(10):1301–1310, 2002.
- [5] François Chollet. Xception: Deep learning with depthwise separable convolutions, 2017.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- [7] Murat Karabatak and M. Cevdet Ince. An expert system for detection of breast cancer based on association rules and neural network. *Expert Systems with Applications*, 36(2, Part 2):3465–3469, 2009.
- [8] Shen L., Margolies L.R., and Rothstein J.H. et al. Deep learning to improve breast cancer detection on screening mammography. *Sci Rep*, 9, 2019.
- [9] Arnold M, Morgan E, and Rumgay H et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040, 2022.

- [10] M. M. Mehdy, P. Y. Ng, E. F. Shair, N. I. Md Saleh, and C. Gomes. Artificial neural networks in image processing for early detection of breast cancer. *Computational and Mathematical Methods in Medicine*, 2017, 2017.
- [11] Nada Mobark, Safwat Hamad, and S. Z. Rida. Coronet: Deep neural network-based end-to-end training for breast cancer diagnosis. *Applied Sciences*, 12(14), 2022.
- [12] Mohammad R. Mohebian, Hamid R. Marateb, Marjan Mansourian, Miguel Angel Mañanas, and Fariborz Mokarian. A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (hpbc) using optimized ensemble learning. *Computational and Structural Biotechnology Journal*, 15:75–85, 2017.
- [13] Udit Pratap and Saurabh Chhabra. Breast cancer prediction using different machine learning algorithms. In *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, pages 451–454, 2021.
- [14] Sawyer-Lee R., Gimenez F., Hoogi A., and Rubin D. Curated breast imaging subset of digital database for screening mammography (cbis-ddsm) (version 1) [data set], 2016.
- [15] Satish Saini and Ritu Vijay. Mammogram analysis using feed-forward back propagation and cascade-forward back propagation artificial neural network. In *2015 Fifth International Conference on Communication Systems and Network Technologies*, pages 1177–1180, 2015.
- [16] Mingxing Tan and Quoc V. Le. Efficientnetv2: Smaller models and faster training, 2021.
- [17] Adam Yala, Constance Lehman, Tal Schuster, Tally Portnoi, and Regina Barzilay. A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology*, 292(1):60–66, 2019. Pmid: 31063083.
- [18] Guodong Zhang, Peiyu Yan, Hong Zhao, and Xin Zhang. A computer aided diagnosis system in mammography using artificial neural networks. In *2008 International Conference on BioMedical Engineering and Informatics*, volume 2, pages 823–826, 2008.