



Pontificia Universidad
JAVERIANA
Cali

**IMPLEMENTACIÓN DE UN MODELO DE APRENDIZAJE PROFUNDO DE MÁQUINAS
PARA LA DETECCIÓN DE CÁNCER DE SENO**

JAIR ORLANDO ÁVILA GÓMEZ- 8975854
FABIAN ALONSO MALDONADO BENAVIDES- 8975652
LAURA YOHANNA BERMÚDEZ MURILLO- 8975538

*Proyecto Aplicado para optar al título de Magister en Ciencia de
Datos*

Director(a)

Dr. MANUEL GUILLERMO FORERO VARGAS

Codirector(a)

Mg. OMAR ANDRÉS CASTAÑO

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI
JULIO 17 DE 2024

FICHA RESUMEN

TÍTULO: IMPLEMENTACIÓN DE UN MODELO DE APRENDIZAJE PROFUNDO DE MAQUINAS PARA LA DETECCIÓN DE CÁNCER DE SENO

1. ÉNFASIS: APRENDIZAJE PROFUNDO
2. TIPO DE PROYECTO: INNOVACIÓN – INVESTIGACIÓN
3. ÁREA DE TRABAJO: INTELIGENCIA ARTIFICIAL APLICADA A LA MEDICINA.
4. ESTUDIANTE(S): JAIR AVILA GÓMEZ, FABIAN MALDONADO, LAURA BERMUDEZ
5. CORREO ELECTRÓNICO: jair1020@javerianacali.edu.co, fmaldonado@javerianacali.edu.co, laurabermudez@javerianacali.edu.co,
6. DIRECCIÓN Y TELÉFONO: Cra 29 # 28-49 SUR APTO 1107 Bogotá
7. DIRECTOR: DR. MANUEL FORERO
8. VINCULACIÓN DEL DIRECTOR: PROFESOR DE APOYO POST GRADOS, ACTUALMENTE DOCENTE UNIVERSIDAD DE IBAGUÉ DTO. DE ING. DE SISTEMAS
9. CORREO ELECTRÓNICO DEL DIRECTOR: manuel.forero@unibague.edu.co
10. CODIRECTOR: MG. OMAR ANDRÉS CASTAÑO
11. CORREO ELECTRÓNICO DEL CODIRECTOR: omar.castano@javerianacali.edu.co
12. PALABRAS CLAVE: Aprendizaje profundo, detección de nódulos, imágenes médicas, mamografías, cáncer de seno, identificación de cáncer, sistema de detección, inteligencia artificial en salud, diagnóstico por imágenes, procesamiento de imágenes médicas.
13. ODS QUE APLICA EL PROYECTO (Agenda 2030): ODS 3, Garantizar una vida saludable y promover el bienestar para todos en todas las edades.
14. FECHA DE INICIO: MAYO 2023
15. DURACIÓN ESTIMADA (En meses): 12 MESES
16. RESUMEN: El objetivo principal de este trabajo es implementar un modelo de aprendizaje profundo para la detección de cáncer de seno mediante el análisis de imágenes médicas obtenidas a través de mamografías. Se propone el desarrollo y entrenamiento de un sistema capaz de identificar nódulos sospechosos en estas imágenes. Para lograrlo, se utilizarán diversas arquitecturas convolucionales, como VGG16, VGG19, RESNET y RESNET50, así como Visión Transformer (VIT).

TABLA DE CONTENIDO

INTRODUCCIÓN	5
1. DEFINICIÓN DEL PROBLEMA	6
1.1. PLANTEAMIENTO DEL PROBLEMA	6
1.2. FORMULACIÓN DEL PROBLEMA.....	6
2. OBJETIVOS DEL PROYECTO.....	7
2.1. OBJETIVO GENERAL	7
2.2. OBJETIVOS ESPECÍFICOS	7
3. MARCO TEÓRICO Y ANTECEDENTES	8
3.1. MARCO TEÓRICO	8
3.1.1. Cáncer de mama.....	8
3.1.2. Red neuronal Artificiales (ANNS)	8
3.1.3. Arquitectura de las redes neuronales	9
3.1.4. Red neuronal recurrente (RNN):	10
3.1.5. Red neuronal Convolutiva (CNN):.....	11
3.1.6. Convolución.....	12
3.1.7. Maxpooling.....	13
3.1.8. Investigación y selección de arquitecturas (CNN).....	14
3.1.8.1. Arquitectura VGG16	15
3.1.8.2. Arquitectura VGG19	16
3.1.8.3. Arquitectura Res Net	17
3.1.8.4. Arquitectura Res Net -50.....	18
3.1.8.5. Arquitectura VIT	19
3.1.8.6. Arquitecta YOLO	21
3.1.9. Funciones de Activación en Redes Neuronales	22
3.1.9.1. Tipos Comunes de Funciones de Activación	23
Sigmoides (σ)	23
Tangente Hiperbólica (tanh)	23
3.1.9.2. Importancia de las Funciones de Activación	24
3.1.10. Imagenología:	25
3.1.11. Mamografía Digital	25

3.2.	ANTECEDENTES / TRABAJOS RELACIONADOS.....	26
4.	METODOLOGÍA.....	28
4.1.	SELECCIÓN DE DATOS:.....	29
4.1.1.	Búsqueda y recopilación de conjuntos de datos de mamografías médicas.	29
4.1.2.	Análisis de las características de los datos.....	32
4.1.3.	Obtención de permisos y aprobaciones éticas para el uso de datos médicos	36
4.2.	PRE- PROCESAMIENTO DE DATOS:.....	37
4.2.1.	Recursos Tecnológicos	37
4.2.2.	Normalización y estandarización de las imágenes.....	38
4.2.3.	División de datos en conjunto de entrenamiento, validación y prueba.....	41
4.2.4.	Revisión Para aumentar la variabilidad del conjunto de datos de entrenamiento	44
4.3.	DISEÑO DEL MODELO	46
4.3.1.	Preparación del dataset para la implementación de los modelos.....	47
4.3.2.	Diseño Modelo 1 VGG16	49
4.3.3.	Prueba Modelo 2 VGG19.....	52
4.3.4.	Prueba Modelo RESNET	55
4.3.5.	Prueba Modelo VTI.....	59
4.4.	Tablas de Resultados de las arquitecturas evaluadas.	61
4.4.1.	Tabla de Resultados gráficos de las arquitecturas evaluadas.....	64
4.4.2.	Análisis Comparativo.....	68
4.4.3.	Análisis del impacto de ajustes en el rendimiento de Modelos	69
4.4.4.	Interpretación y Recomendaciones	71
4.1.	Resumen de la Aplicación y Desempeño de Modelos de Aprendizaje Profundo	72
5.	CONCLUSIONES Y TRABAJOS FUTUROS	74
5.1.	CONCLUSIONES.....	74
5.2.	TRABAJOS FUTUROS	75
6	REFERENCIAS BIBLIOGRÁFICAS	76

LISTA DE FIGURAS Y TABLAS

Figura N° 1. Red Neuronal Artificial Monocapa	9
Figura N° 2. Red Neuronal Artificial Multicapa	10
Figura N° 3 . Combinaciones posibles en una Red Neuronal Recurrente	11
Figura N° 4 Red Neuronal Convolutiva	12
Figura N° 5. Operación de convolución en una CNN	12
Figura N° 6 Operación de Reducción - Pooling	13
Figura N° 7 . Metodología para desarrollo de la investigación- Creación propia	29
Tabla N° 8 Comparativa de las Diferentes bases de datos analizadas -creación Propia	33
Figura N° 9 VinDr-MammoVinDr-Mammo.....	34
Figura N° 10 Data análisis de Tipso de Split- Creación Propia	35
Figura N° 11 Tabla Data análisis de BI-RADS inicial- creación Propia	36
Figura N° 12 Acuerdo de uso de la fuente de datos	37
Figura N° 13 Imagen descargue dataset.....	37
Figura N° 14 Ejemplo práctico de recortes en Python y Recorte en programa visualización.....	40
Figura N° 15 VinDr-Mammo Ejemplo preparación recortes.....	41
Figura N° 16 VinDr-Mammo Ejemplo preparación recortes.....	42
Figura N° 17 VinDr-Mammo Ejemplos de recorres en general.....	44
Figura N° 18 VinDr-Mammo Ejemplo preparación recortes áreas de interés	44
Figura N° 19 VinDr-Mammo Ejemplo preparación recorte áreas de interés selección de hallazgo	45
Figura N° 20 Modelo VGG16Ajuste de parámetros	49
Figura N° 21 VGG16 Modificada.....	51
Figura N° 22 Modelo VGG19	53
Figura N° 23 Resnet Modificada.....	55
Figura N° 24 Modelo Resnet 50	57
Figura N° 25 Modelo VIT	59
Figura N° 26 Tabla de Resultados de las arquitecturas evaluadas- creación Propia.	61
Figura N° 27 Tabla de resultados datos de entrenamiento con o sin ajuste fino	62
Figura N° 28 Tabla de resultados datos de Validación con o sin ajuste fino.....	62
Figura N° 29 Tabla de métricas de evaluación de datos con etiqueta sin Hallazgo	63
Figura N° 30 Tabla de métricas de evaluación de datos con etiqueta Con Halla	63
Figura N° 31 Tabla Análisis Grafico Resumen de las arquitecturas.....	68
Figura N° 32 Tabla de resultados de la aplicación de los modelos- creación Propia.....	73

INTRODUCCIÓN

El cáncer de seno es una enfermedad caracterizada por la proliferación de células malignas en los tejidos mamarios. Afecta principalmente a las mujeres y se convierte en la segunda causa principal de muerte por cáncer en este grupo. Se estima que cada año se diagnostican más de 2 millones de casos nuevos de cáncer de seno en todo el mundo. La detección temprana es fundamental para mejorar las tasas de supervivencia y garantizar un tratamiento adecuado. Una de las herramientas más utilizadas para esta detección es la mamografía, una técnica que utiliza rayos X para producir imágenes del tejido mamario. Sin embargo, la interpretación de estas imágenes puede ser complicada debido a diversos factores, lo que requiere experiencia y habilidad por parte del experto [1].

El procesamiento de imágenes médicas y el desarrollo de algoritmos han cobrado relevancia en el ámbito científico. La inteligencia artificial y la visión por computadora han permitido analizar imágenes a gran escala y con alta velocidad. La implementación de modelos de redes neuronales convolucionales (CNN) para la detección temprana del cáncer de seno a partir de imágenes de mamografía se presenta como una solución efectiva para mejorar la precisión y eficiencia en el diagnóstico. Estos modelos de CNN tienen la capacidad de analizar grandes volúmenes de datos y aprender patrones complejos en las imágenes, lo que los convierte en herramientas ideales para identificar nódulos mamarios y tomar decisiones clínicas de manera eficaz.

En el presente trabajo se hace uso de diferentes modelos como VGG16, VGG19, RESNET, RESNET50 y VIT transformers a partir de los datos disponibles. Estas arquitecturas de aprendizaje moderno se comparan con técnicas clásicas de clasificación, empleando diferentes enfoques para encontrar la mejor estrategia. Asimismo, se identificaron debilidades, fortalezas y nuevas oportunidades que puedan ser implementadas en futuros modelos para tener un impacto significativo en la salud y bienestar de las mujeres.

1. DEFINICIÓN DEL PROBLEMA

1.1. PLANTEAMIENTO DEL PROBLEMA

El cáncer es la segunda causa de muerte en el mundo, representando cerca de 10 millones de defunciones cada año. Según cifras preliminares de la Cuenta de Alto Costo (CAC) previas al proceso de auditoría al 31 de diciembre de 2022, 491.048 personas han sido diagnosticadas con algún tipo de cáncer en el marco del aseguramiento en salud colombiano. Igualmente reporto que el cáncer de mama fue el más frecuente, seguido de cérvix y colorrectal en la población femenina [2].

El Cáncer de mama se puede diagnosticar a través de varios métodos, el segundo método más utilizado después del autoexamen de seno y que aun demuestra ser uno de los métodos de diagnóstico más efectivos es el uso de imágenes de rayos X, también conocidas como mamografías; sin embargo, la interpretación de estas imágenes puede llegar a ser compleja debido a diferentes factores como dificultad de identificación del nódulo y la densidad de las mamas [3]. En los últimos años, el aprendizaje de máquinas ha sido empleado para el desarrollo de modelos de identificación de presencia de quistes o nódulos cancerosos, sirviendo como herramienta al médico para un mejor diagnóstico [4].

La detección temprana del cáncer de seno es un factor clave para aumentar las probabilidades de supervivencia y mejorar los resultados del tratamiento. En este contexto, el uso de modelos de aprendizaje profundo para identificar nódulos en imágenes de mamografías se presenta como una alternativa relevante y crucial [5]. Sin embargo, este enfoque implica varios desafíos y esfuerzos, tanto en la obtención e interpretación de grandes conjuntos de datos de imágenes de mamografías como en el aprovechamiento de las infraestructuras de computación.

1.2. FORMULACIÓN DEL PROBLEMA

¿Cómo implementar un modelo de aprendizaje profundo de máquinas utilizando mamografías, para mejorar la precisión del diagnóstico en la detección de cáncer de seno?

¿Cuáles son los principales desafíos durante el proceso de recolección, limpieza y organización del dataset de mamografías para garantizar la calidad y usabilidad de la información en el proyecto?

¿Cómo se llevará a cabo el proceso de investigación, selección y entrenamiento del modelo de aprendizaje profundo, y qué estrategias se utilizarán para ajustar los hiperparámetros y optimizar la configuración del modelo con el objetivo de mejorar la detección de nódulos?

¿Cuáles serán las métricas y técnicas de evaluación a utilizar para la validación del rendimiento del modelo a implementar y cómo contribuyeron estas en los resultados para la optimización y mejoramiento del modelo?

2. OBJETIVOS DEL PROYECTO

2.1. OBJETIVO GENERAL

Implementar un modelo de aprendizaje profundo para la identificación de nódulos en imágenes de mamografías para la detección de cáncer de seno.

2.2. OBJETIVOS ESPECÍFICOS

- Recolectar y preparar un dataset de mamografías, realizando limpieza y organización de la información a utilizar en el proyecto.
- Investigar, seleccionar y entrenar el modelo de aprendizaje profundo utilizando el dataset preparado, ajustando los Hiperparámetros y optimizando su configuración para la detección de nódulos.
- Validar el rendimiento del modelo implementado, utilizando métricas y técnicas de evaluación.

3. MARCO TEÓRICO Y ANTECEDENTES

3.1. MARCO TEÓRICO

3.1.1. Cáncer de mama

El cáncer es el resultado de mutaciones, o cambios anómalos, en los genes que regulan el crecimiento de las células y las mantienen sanas. Los genes se encuentran en el núcleo de las células, el cual actúa como la “sala de control” de cada célula. Normalmente, las células del cuerpo se renuevan mediante un proceso específico llamado crecimiento celular: las células nuevas y sanas ocupan el lugar de las células viejas que mueren. Pero con el paso del tiempo, las mutaciones pueden “activar” ciertos genes y “desactivar” otros en una célula. La célula modificada adquiere la capacidad de dividirse sin ningún tipo de control u orden, por lo que produce más células iguales y genera un tumor, un tumor puede ser benigno (no es peligroso para la salud) o maligno (es potencialmente peligroso). Los tumores benignos no son considerados cancerosos: sus células tienen una apariencia casi normal, crecen lentamente y no invaden tejidos próximos ni se propagan hacia otras partes del cuerpo. Los tumores malignos son cancerosos. De no ser controladas, las células malignas pueden propagarse más allá del tumor original hacia otras partes del cuerpo.

El término “cáncer de mama” hace referencia a un tumor maligno que se ha desarrollado a partir de células mamarias. Generalmente, el cáncer de mama se origina en las células de los lobulillos, que son las glándulas productoras de leche, o en los conductos, que son las vías que transportan la leche desde los lobulillos hasta el pezón. Con menos frecuencia, el cáncer de mama puede originarse en los tejidos estromales, que incluyen a los tejidos conjuntivos grasos y fibrosos de la mama [6].

3.1.2. Red neuronal Artificiales (ANNS)

Las (ANNS) son un modelo de un algoritmo computacional inspirado en las redes neuronales biológicas. Se crearon como una herramienta que pueda aprender, generalizar y procesar automáticamente datos, con aplicaciones a tareas de clasificación y regresión. Dando su mayor aplicabilidad en el reconocimiento de patrones y por ende en la predicción de comportamientos[7].

Una neurona artificial se compone de una o múltiples entradas, un peso, un bias (sesgo), un sumador y una función de transferencia (o de activación), como se muestra en la figura 1. Normalmente una neurona, aún con múltiples entradas, no es suficiente, por lo tanto, las ANNS usan varias neuronas ubicadas en paralelo formando lo que se conoce como una capa [8].

3.1.3. Arquitectura de las redes neuronales

Redes neuronales monocapa y multicapa:

Una ANN puede poseer una capa o múltiples capas, la red neuronal monocapa se corresponde con la red neuronal más simple, está compuesta por una capa de neuronas que proyectan las entradas a una capa de neuronas de salida donde se realizan los diferentes cálculos.

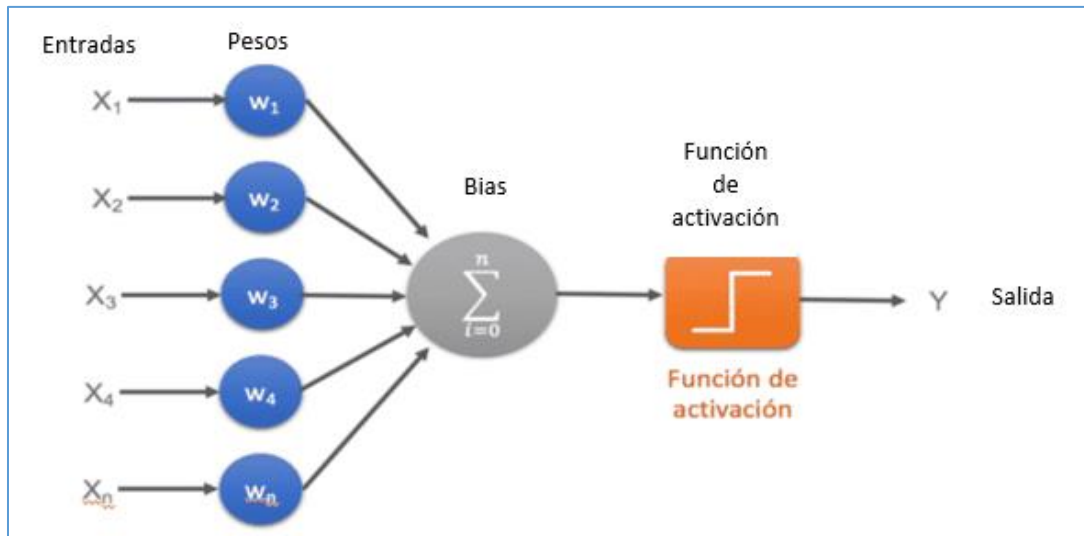


Figura N° 1. Red Neuronal Artificial Monocapa

Fuente : D. Calvo, "Clasificación de redes neuronales artificiales", Diego Calvo, <https://www.diegocalvo.es/clasificacion-de-redes-neuronales-artificiales/>, 2020.

La red neuronal multicapa es una generalización de la red neuronal monocapa, la diferencia reside en que mientras la red neuronal monocapa está compuesta por una capa de neuronas de entrada y una capa de neuronas de salida, esta dispone de un conjunto de capas intermedias (capas ocultas) entre la capa de entrada y la de salida, como se muestra en la figura 2.

Normalmente, todas las neuronas de una capa reciben señales de entrada desde otra capa anterior (la cual está más cerca a la entrada de la red), y envían señales de salida a una capa posterior (que está más cerca a la salida de la red). A estas conexiones se las denomina conexiones hacia adelante o feedforward. Sin embargo, en un gran número de estas redes también existe la posibilidad de conectar la salida de las neuronas de capas posteriores a la entrada de capas anteriores; a estas conexiones se las denomina conexiones hacia atrás o feedback. Estas dos posibilidades permiten distinguir entre dos tipos de redes con múltiples capas: las redes con

conexiones hacia adelante o redes feedforward, y las redes que disponen de conexiones tanto hacia adelante como hacia atrás o redes feedforward/feedback [10].

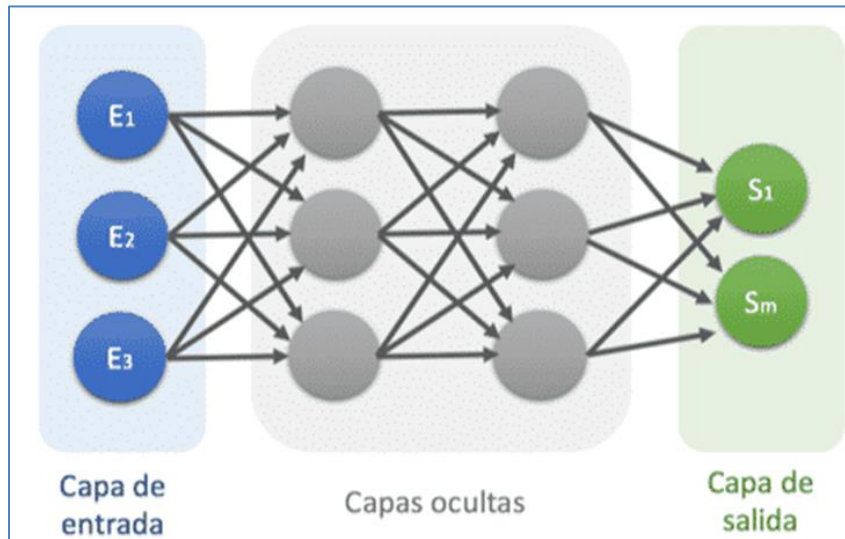


Figura N° 2. Red Neuronal Artificial Multicapa

Fuente : D. Calvo, “Clasificación de redes neuronales artificiales”, Diego Calvo, <https://www.diegocalvo.es/clasificacion-de-redes-neuronales-artificiales/>, 2020.

3.1.4. Red neuronal recurrente (RNN):

Las RNNs son una familia particular de Red de Neuronas Artificiales, caracterizadas por la existencia de conexiones de retroalimentación, esto es, ciclos o bucles dentro de la estructura de la red, mediante los cuales se pueden analizar secuencias de tiempo. De hecho, en el aspecto computacional, a menudo se realizan despliegues de la estructura, para obtener una versión de la red que dependa de una secuencia de entradas. Se puede mezclar parte de las entradas o de las salidas para obtener diferentes combinaciones, algunas de ellas se muestran en la figura 3. Por ejemplo, cabe enlazar uno a muchos para clasificar una secuencia de datos con una sola salida o para etiquetar el conjunto de sujetos presentes de una imagen[10].

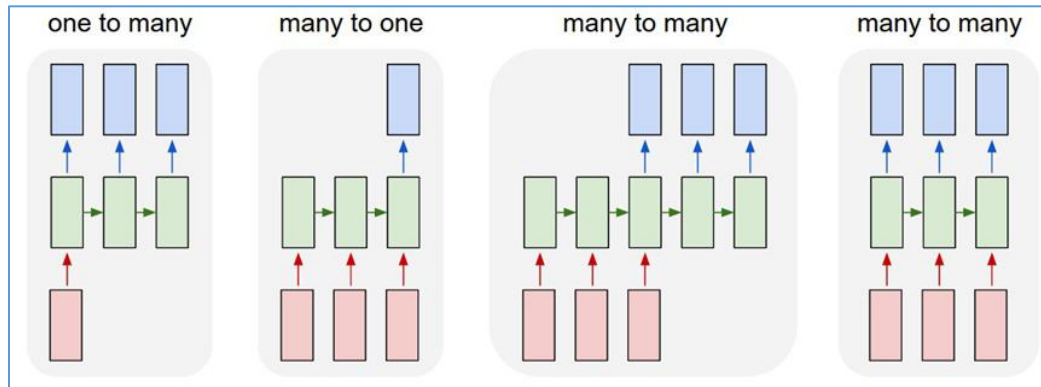


Figura N° 3 . Combinaciones posibles en una Red Neuronal Recurrente

Fuente : D. Calvo, "Clasificación de redes neuronales artificiales", Diego Calvo, <https://www.diegocalvo.es/clasificacion-de-redes-neuronales-artificiales/>, 2020.

3.1.5. Red neuronal Convolutiva (CNN):

La CNN es un tipo de Red Neuronal Artificial que procesa sus capas imitando al cortex visual del cerebro humano para identificar distintas características en las entradas. Para ello, la CNN contiene varias capas ocultas especializadas y con una jerarquía: esto significa que las primeras capas detectan propiedades o formas básicas y se van especializando hasta llegar a capas más profundas capaces de reconocer formas complejas como un rostro o una silueta, además, presenta invarianza a la traslación de los patrones a identificar [11].

Las redes neuronales convoluciones son muy potentes para todo lo que tiene que ver con el análisis de imágenes, debido a que son capaces de detectar características simples como por ejemplo detección de bordes, líneas, etc y componer en características más complejas hasta detectar lo que se busca [11].

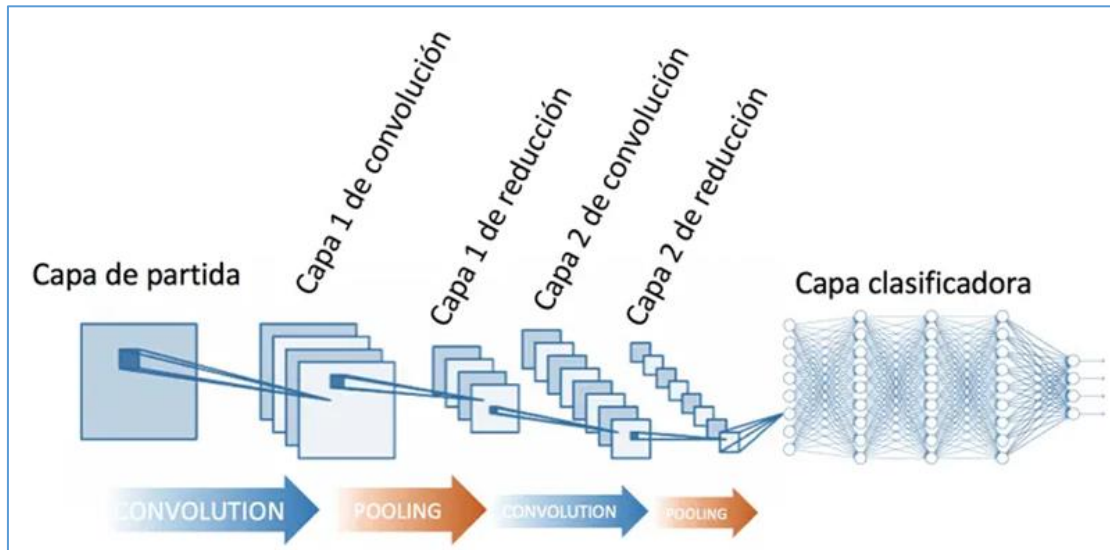


Figura N° 4 Red Neuronal Convocional

Fuente : D. Calvo, "Clasificación de redes neuronales artificiales", Diego Calvo, <https://www.diegocalvo.es/clasificacion-de-redes-neuronales-artificiales/>, 2020.

3.1.6. Convolución

En la convolución se realizan operaciones de productos y sumas entre la capa de partida y los n filtros (o kernel) que genera un mapa de características, como se muestra en la figura 5. Las características extraídas corresponden a cada posible ubicación del filtro en la imagen original. La ventaja es que el mismo filtro (= neurona) sirve para extraer la misma característica en cualquier parte de la entrada, con esto consigue reducir el número de conexiones y el número de parámetros a entrenar en comparación con una red multicapa de conexión total [12] .

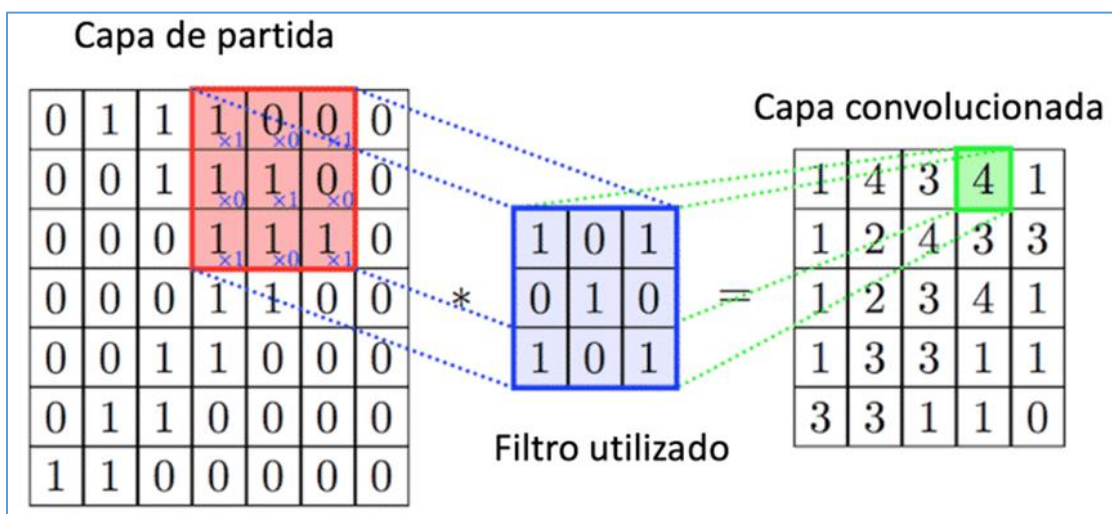


Figura N° 5. Operación de convolución en una CNN

Fuente : D. Calvo, "Clasificación de redes neuronales artificiales", Diego Calvo, <https://www.diegocalvo.es/clasificacion-de-redes-neuronales-artificiales/>, 2020.

Una vez terminado el proceso anterior, se reduce la cantidad de neuronas antes de realizar una nueva convolución, en la reducción se disminuye la cantidad de parámetros al quedarse con las características más comunes. Para reducir el tamaño de la próxima capa se realiza el proceso de subsampling en el que se reduce el tamaño de las imágenes filtradas, pero donde prevalecen las características más importantes que detectó cada filtro [11].

3.1.7. Maxpooling

Hay diversos tipos de subsampling, pero el más utilizado es el Max-Pooling. Si suponemos que se realiza una Max-pooling de tamaño 2x2 se recorrerán cada una de las imágenes de características obtenidas anteriormente de izquierda a derecha, arriba-abajo, pero en lugar de a un solo pixel se toman de 2x2 – 2 de alto por 2 de ancho – y se preserva el valor más elevado de esos 4 pixeles, de ahí el termino Max. En este caso, usando 2x2 la imagen resultante se reduce a la mitad y por ende pasamos a tener la mitad de las neuronas, como se muestra en la figura 6. El descenso es considerable y teóricamente almacenan la información más importante para detectar las características deseadas

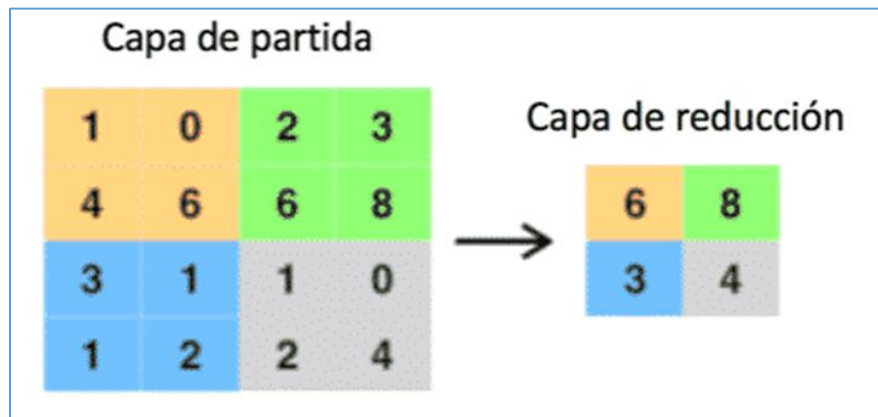


Figura N° 6 Operación de Reducción - Pooling

Fuente : D. Calvo, "Clasificación de redes neuronales artificiales", Diego Calvo, <https://www.diegocalvo.es/clasificacion-de-redes-neuronales-artificiales/>, 2020.

3.1.8. Investigación y selección de arquitecturas (CNN)

Las redes neuronales convolucionales han existido desde la década de 1980, pero su popularidad y avances significativos surgieron más tarde. Un hito importante fue el desarrollo de LeNet-5 por Yann LeCun en 1998, que fue una de las primeras arquitecturas de CNN utilizadas con éxito para el reconocimiento de dígitos escritos a mano.

Sin embargo, el gran auge de las CNN ocurrió a partir de 2012, cuando Alex Krizhevsky, Ilya Sutskever y Geoffrey Hinton desarrollaron AlexNet, una CNN profunda que ganó el concurso ImageNet Large Scale Visual Recognition Challenge (ILSVRC) con un margen significativo, mostrando la eficacia de las CNN en tareas de clasificación de imágenes a gran escala.

A pesar de su éxito, el ILSVRC finalizó en 2017 debido a la saturación del conjunto de datos ImageNet y la dificultad para seguir mejorando los resultados. Sin embargo, su legado perdura y continúa siendo una referencia importante en el campo de la visión por computadora.

Desde entonces, las CNN han sido la base de muchos avances en visión por computadora, incluyendo reconocimiento facial, detección de objetos, segmentación semántica, entre otros.

Características:

1. **Convoluciones:** Las CNN utilizan capas convolucionales para detectar características específicas en las imágenes. Estas capas aplican filtros que convolucionan la imagen original para extraer características como bordes, texturas, y patrones.
2. **Pooling:** Después de las capas de convolución, se suelen aplicar capas de "pooling" para reducir la dimensionalidad de las características extraídas, preservando las más relevantes.
3. **Arquitectura en capas:** Las CNN suelen tener una arquitectura en capas, con capas convolucionales intercaladas con capas de activación, pooling y, finalmente, capas completamente conectadas que generan la salida final.
4. **Aprendizaje jerárquico de características:** Las CNN son capaces de aprender características de manera jerárquica, donde las primeras capas detectan características simples como bordes, y las capas más profundas combinan estas características simples para detectar patrones complejos.

Campo de uso:

Las CNNs han demostrado un gran éxito en una amplia gama de tareas de visión por computadora, incluyendo:

- **Reconocimiento de imágenes:** Clasificar imágenes en diferentes categorías, como gatos, perros, muebles o paisajes.
- **Detección de objetos:** Localizar y delimitar objetos presentes en una imagen, indicando también su categoría.
- **Segmentación de imágenes:** Dividir una imagen en diferentes regiones o segmentos, correspondientes a distintos objetos o categorías.
- **Análisis de imágenes:** Extraer información relevante de una imagen, como la presencia de rostros, emociones o texto.
- **Generación de imágenes:** Crear nuevas imágenes a partir de ejemplos existentes o descripciones textuales.

Otras aplicaciones:

Además de la visión por computadora, las CNNs también se han aplicado en otros campos como:

- **Procesamiento del lenguaje natural:** Analizar y comprender texto, como la clasificación de documentos o la traducción automática.
- **Bioinformática:** Analizar datos biológicos, como la predicción de estructuras moleculares o la identificación de enfermedades.
- **Robótica:** Controlar robots y vehículos autónomos para navegar y manipular objetos.

3.1.8.1. Arquitectura VGG16

La VGG16 es una arquitectura específica de red neuronal convolucional (CNN) propuesta por Karen Simonyan y Andrew Zisserman en 2014 [12].

Características Principales:

- **Profundidad:** La VGG16 es una red neuronal muy profunda, con un total de 16 capas de procesamiento (de ahí su nombre). Estas capas consisten en capas convolucionales, capas de submuestreo (pooling) y algunas capas totalmente conectadas al final para la clasificación.
- **Estructura repetitiva:** La arquitectura VGG16 se caracteriza por tener bloques repetitivos de capas convolucionales y de submuestreo. En lugar de utilizar capas convolucionales con tamaños de kernel muy grandes, la VGG16 utiliza filtros de convolución más pequeños (3x3) en todas sus capas, seguidos de una capa de activación ReLU. Esto permite aprender características más detalladas y facilita la interpretación de la arquitectura.

- **Capas de submuestreo:** Después de algunas capas convolucionales, la VGG16 utiliza capas de submuestreo (pooling) para reducir la dimensionalidad de las características y aumentar la invariancia a la translación de las mismas. En la VGG16, se utilizan capas de submuestreo máximo (max pooling) con ventanas de 2x2 y un paso (stride) de 2.
- **Capas totalmente conectadas:** Al final de la red, después de las capas convolucionales y de submuestreo, se utilizan capas totalmente conectadas para la clasificación. Estas capas toman las características extraídas por las capas anteriores y las utilizan para clasificar la imagen en diferentes categorías.
- **Pre-entrenamiento y transferencia de aprendizaje:** Debido a su éxito en tareas de clasificación de imágenes, la VGG16 se ha utilizado ampliamente como una red pre-entrenada para transferir aprendizaje a otros conjuntos de datos y tareas. Esto significa que la red se entrena inicialmente en un conjunto de datos grande (como ImageNet) y luego se ajusta o fine-tunea para una tarea específica con un conjunto de datos más pequeño.

3.1.8.2. Arquitectura VGG19

la VGG19 es una variante de la arquitectura VGG16 que fue propuesta por Karen Simonyan y Andrew Zisserman en 2014. Es similar a la VGG16 en muchos aspectos, pero con una estructura más profunda, ya que incluye un total de 19 capas de procesamiento [12].

Características Principales:

- **Mayor profundidad:** La VGG19 tiene una estructura más profunda que la VGG16, con cuatro capas convolucionales adicionales. Estas capas adicionales permiten que la red aprenda representaciones más complejas y abstractas de las imágenes, lo que puede conducir a un rendimiento mejorado en ciertas tareas.
- **Estructura similar:** Al igual que la VGG16, la VGG19 se caracteriza por tener bloques repetitivos de capas convolucionales y de submuestreo. Utiliza filtros de convolución de tamaño pequeño (3x3) en todas sus capas convolucionales, seguidos de funciones de activación ReLU, lo que facilita la interpretación de la arquitectura.
- **Capas de submuestreo:** La VGG19 utiliza capas de submuestreo máximo (max pooling) después de algunas de sus capas convolucionales para reducir la dimensionalidad de las características y aumentar la invariancia a la translación de las mismas, similar a la VGG16.

- **Capas totalmente conectadas:** Al final de la red, después de las capas convolucionales y de submuestreo, se utilizan capas totalmente conectadas para la clasificación, al igual que en la VGG16.
- **Preentrenamiento y transferencia de aprendizaje:** Al igual que la VGG16, la VGG19 se puede utilizar como una red reentrenada para transferir aprendizaje a otros conjuntos de datos y tareas. Esto significa que la red se entrena inicialmente en un conjunto de datos grande (como ImageNet) y luego se ajusta o fine-tunea para una tarea específica con un conjunto de datos más pequeño.

Campo de Acción Y problemas experimentados Arquitectura VGG16-VGG19

- Reconocimiento de imágenes
- Detección de objetos
- Segmentación de imágenes

Problemas experimentados:

- Profundidad: Son redes muy profundas, lo que puede generar problemas de sobreajuste y alta complejidad computacional.
- Falta de mecanismos de atención: No cuentan con mecanismos de atención que les permitan enfocarse en las partes más relevantes de la imagen.
- Baja eficiencia: Su rendimiento computacional puede ser inferior a arquitecturas más modernas.

3.1.8.3. Arquitectura Res Net

es una arquitectura de red neuronal profunda que fue presentada por Kaiming He, et al. en 2015. ResNet se diseñó para abordar el problema del desvanecimiento del gradiente que puede surgir en redes neuronales muy profundas. El desvanecimiento del gradiente puede dificultar el entrenamiento de redes neuronales profundas, ya que las actualizaciones de los pesos se vuelven muy pequeñas a medida que se propagan hacia atrás a través de muchas capas, lo que puede llevar a que el modelo no converja o converja muy lentamente.

La característica principal de ResNet es el uso de conexiones residuales, que permiten que la información original fluya directamente a través de la red sin ser transformada, lo que facilita el entrenamiento de redes profundas. Estas conexiones residuales se implementan mediante

bloques residuales, que consisten en una o más capas convolucionales seguidas de una conexión directa (shortcut connection) que agrega la entrada original a la salida del bloque [13].

Principales características:

- **Conexiones residuales:** En lugar de tratar de aprender las funciones de mapeo directamente, ResNet aprende las diferencias (residuos) entre las entradas y las salidas. Esto significa que cada bloque de la red solo necesita aprender las transformaciones residuales, lo que facilita el entrenamiento de redes profundas.
- **Arquitectura en cascada:** ResNet utiliza una arquitectura en cascada donde las capas se organizan en bloques, y cada bloque contiene varias capas convolucionales seguidas de una conexión residual. Este diseño permite la construcción de redes neuronales muy profundas con miles de capas.
- **Variantes:** Además de la arquitectura original de ResNet, existen varias variantes, como ResNet-50, ResNet-101, ResNet-152, que se diferencian en el número y la configuración de los bloques residuales. Estas variantes se utilizan en diferentes aplicaciones según la complejidad del problema y los recursos disponibles.

3.1.8.4. Arquitectura Res Net -50

Es una variante específica de la arquitectura ResNet que consta de 50 capas en total. Fue introducida por Kaiming He, et al., como parte de la familia de modelos ResNet en su artículo "Deep Residual Learning for Image Recognition" en 2015.

Principales características:

- **Conexiones residuales:** Al igual que otras arquitecturas ResNet, ResNet50 utiliza conexiones residuales para facilitar el entrenamiento de redes neuronales profundas. Estas conexiones residuales permiten que la información original fluya directamente a través de la red sin ser transformada, lo que ayuda a mitigar el problema del desvanecimiento del gradiente y facilita el entrenamiento de redes profundas.
- **Bloques residuales:** ResNet50 está compuesto por bloques residuales que contienen varias capas convolucionales seguidas de una conexión residual. Estos bloques residuales se repiten varias veces a lo largo de la red para construir una arquitectura profunda.
- **Arquitectura en cascada:** ResNet50 sigue una arquitectura en cascada donde las capas se organizan en bloques, y cada bloque contiene múltiples capas convolucionales

seguidas de una conexión residual. La arquitectura en cascada permite construir redes neuronales muy profundas con miles de capas, lo que ha demostrado ser beneficioso para tareas de visión por computadora.

- **Capas de agrupamiento (pooling):** ResNet50 también incluye capas de agrupamiento (pooling) para reducir la dimensionalidad de las características y aumentar la invariancia a la translación de las mismas. Estas capas de agrupamiento ayudan a mejorar la eficiencia computacional y a evitar el sobreajuste.
- **Pre-entrenamiento y transferencia de aprendizaje:** Al igual que otras arquitecturas ResNet, ResNet50 se puede utilizar como una red pre-entrenada para transferir aprendizaje a otras tareas o conjuntos de datos específicos.
- Esto significa que el modelo se puede entrenar inicialmente en un conjunto de datos grande (como ImageNet) y luego ajustar o fine-tunear para una tarea específica con un conjunto de datos más pequeño.

Campo de acción Y problemas experimentados Arquitectura ResNet-ResNet50

- Reconocimiento de imágenes
- Detección de objetos
- Segmentación de imágenes
- Traducción automática
- Procesamiento del lenguaje natural

Problemas experimentados:

- Profundidad: Aunque son más eficientes que VGG, siguen siendo redes relativamente profundas, lo que puede generar problemas de sobreajuste.
- Falta de escalabilidad: Su rendimiento puede disminuir al aumentar el tamaño del conjunto de datos o la complejidad de la tarea.

3.1.8.5. Arquitectura VIT

Una red VIT (Transformador de Imagen Vision Transformer, por sus siglas en inglés) es una arquitectura de red neuronal que utiliza la estructura de transformador, originalmente diseñada para tareas de procesamiento de lenguaje natural (NLP), y la adapta para el procesamiento de

imágenes. Este enfoque fue propuesto en el artículo "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" por Dosovitskiy et al., en 2020 [14].

Principales características:

- Estructura de transformador: Al igual que en los modelos de lenguaje natural, las redes VIT están compuestas por bloques de atención, donde cada bloque calcula las relaciones entre todas las posiciones de una secuencia de entrada. En el caso de VIT, la entrada no es una secuencia de palabras, sino una cuadrícula de píxeles de una imagen.
- Parches de la imagen: Para convertir una imagen en una secuencia de entrada que pueda ser procesada por un transformador, la imagen se divide en pequeños parches o recortes de tamaño fijo. Cada parche se representa como un vector y se aplanan para formar una secuencia que se pasa al modelo de transformador.
- Embeddings de posición: Al igual que en los modelos de lenguaje natural, las redes VIT utilizan embeddings de posición para codificar la posición de cada parche en la secuencia de entrada. Esto permite al modelo capturar la información espacial de la imagen.
- Encabezado clasificador: Después de procesar la secuencia de parches con bloques de atención, las redes VIT utilizan un encabezado clasificador para realizar la tarea de clasificación. Este encabezado puede ser una capa totalmente conectada que toma la representación final de la secuencia como entrada y predice las etiquetas de clase correspondientes.
- Pre-entrenamiento y fine-tuning: Al igual que otros modelos de aprendizaje profundo, las redes VIT se pueden pre-entrenar en grandes conjuntos de datos no etiquetados y luego ajustar o fine-tunear en conjuntos de datos específicos para tareas de clasificación de imágenes.

Campo de acción Y problemas experimentados Arquitectura VIT

- Reconocimiento de imágenes
- Detección de objetos
- Segmentación de imágenes
- Traducción automática
- Procesamiento del lenguaje natural

Problemas experimentados:

- **Requerimientos computacionales altos:** Su entrenamiento y uso pueden ser computacionalmente costosos, especialmente en comparación con arquitecturas basadas en convoluciones.
- **Dificultad para comprender relaciones espaciales:** VIT puede tener dificultades para comprender relaciones espaciales complejas en las imágenes.
- **Falta de investigación:** VIT es una arquitectura relativamente nueva, por lo que aún se necesitan más investigaciones para comprender sus limitaciones y optimizar su rendimiento.

3.1.8.6. Arquitecta YOLO

Es una familia de algoritmos de detección de objetos que utiliza una sola red neuronal para predecir múltiples bounding boxes y sus correspondientes clases de objetos dentro de una imagen. A diferencia de otros métodos que aplican el algoritmo de detección a varias partes de la imagen de forma independiente, YOLO ve toda la imagen a la vez, lo que lo hace extremadamente rápido [15].

Principales características de YOLO

- **Velocidad:** YOLO es conocido por su capacidad de procesar imágenes en tiempo real, alcanzando velocidades de hasta 45 cuadros por segundo en versiones más recientes como YOLOv4 y YOLOv5.
- **Detección en una sola pasada:** YOLO divide la imagen de entrada en una cuadrícula $S \times SS \times S$ y, para cada celda de la cuadrícula, predice B bounding boxes y las probabilidades de clase para esos boxes.
- **Arquitectura de red neuronal:** Utiliza una sola red neuronal convolucional (CNN) que pasa la imagen completa y predice bounding boxes y las probabilidades de clase simultáneamente.
- **Precisión:** Aunque es muy rápida, la precisión de YOLO también es competitiva, especialmente en las versiones más recientes como YOLOv4 y YOLOv5, que han mejorado en gran medida la precisión y el recall.
- **Facilidad de uso:** Implementaciones y librerías en frameworks como TensorFlow, Keras y PyTorch facilitan el uso y la integración de YOLO en proyectos.

Campo de acción Y problemas experimentados YOLO

- **Seguridad y vigilancia:** Detección de objetos sospechosos, monitoreo de áreas restringidas.
- **Automoción:** Detección de peatones, vehículos y señales de tráfico en vehículos autónomos.
- **Medicina:** Análisis de imágenes médicas, como la detección de anomalías en mamografías o resonancias magnéticas.
- **Agricultura:** Monitoreo de cultivos, detección de plagas y enfermedades en plantas.
- **Retail y comercio:** Análisis de comportamientos de compra, detección de productos en anaqueles.

Problemas experimentados

- **Detección de objetos pequeños:** YOLO puede tener dificultades para detectar objetos muy pequeños dentro de una imagen, ya que el tamaño de la cuadrícula puede no ser lo suficientemente fino para capturarlos.
- **Generalización:** Como con cualquier modelo de aprendizaje profundo, YOLO puede sobreajustarse a los datos de entrenamiento y no generalizar bien a nuevas imágenes que no se parecen a las de su conjunto de entrenamiento.
- **Precisión frente a velocidad:** Aunque YOLO es rápido, algunas versiones más antiguas pueden sacrificar precisión por velocidad. Las versiones más nuevas han mejorado en este aspecto, pero aún pueden no ser tan precisas como métodos más lentos y complejos.
- **Ambigüedad en bounding boxes:** YOLO puede predecir múltiples bounding boxes para el mismo objeto, lo que puede requerir técnicas adicionales como Non-Maximum Suppression (NMS) para filtrar las detecciones redundantes.
- **Problemas con la detección en condiciones adversas:** La precisión de YOLO puede disminuir en condiciones de poca iluminación, imágenes con mucho ruido o cuando los objetos están parcialmente ocultos.

3.1.9. Funciones de Activación en Redes Neuronales

Las funciones de activación son componentes cruciales en las redes neuronales convolucionales (CNN), ya que introducen no linealidades en el modelo, permitiendo que la red aprenda y represente relaciones complejas en los datos. Sin estas funciones, las redes neuronales serían simplemente combinaciones lineales de las entradas, limitando su capacidad para resolver problemas complejos [16].

3.1.9.1. Tipos Comunes de Funciones de Activación

Sigmoide (σ)

- **Fórmula:**

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

- **Características:** Su salida está en el rango (0, 1), lo que la hace útil para problemas de clasificación binaria. Sin embargo, puede sufrir de problemas de Desvanecimiento del gradiente.

Tangente Hiperbólica (tanh)

- **Fórmula:**

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

- **Características:** Su salida está en el rango (-1, 1). Es similar a la sigmoide pero centrada en cero, lo que puede ayudar a que la red converja más rápidamente.

ReLU (Rectified Linear Unit)

- **Fórmula:**

$$\text{ReLU}(x) = \max(0, x)$$

- **Características:** Es la función de activación más utilizada en redes profundas debido a su simplicidad y eficiencia. Ayuda a mitigar el problema del desvanecimiento del gradiente, pero puede sufrir de "neuronas muertas" si muchas unidades se activan a cero.

Leaky ReLU

- **Fórmula:**

$$\text{Leaky ReLU}(x) = \begin{cases} x & \text{si } x > 0 \\ \alpha x & \text{si } x \leq 0 \end{cases}$$

Características: Similar a ReLU, pero permite un pequeño gradiente cuando la unidad está inactiva, lo que puede ayudar a evitar el problema de las neuronas muertas.

Softmax

- **Fórmula:**

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

- **Características:** Convierte un vector de valores en una distribución de probabilidad. Es comúnmente utilizada en la capa de salida para problemas de clasificación multiclase.

3.1.9.2. Importancia de las Funciones de Activación

Las funciones de activación permiten que las redes neuronales aprendan patrones no lineales, lo cual es esencial para tareas complejas como la detección de nódulos en mamografías. Al seleccionar y ajustar adecuadamente estas funciones, se puede mejorar significativamente el rendimiento del modelo [16].

Aplicación en Modelos Específicos

- **VGG16 y VGG19:** Utilizan principalmente ReLU en sus capas ocultas, lo que les permite aprender características complejas de las imágenes.

- **ResNet y ResNet50:** También utilizan ReLU, pero su arquitectura incluye conexiones residuales que ayudan a mitigar el problema del desvanecimiento del gradiente en redes muy profundas.
- **ViT (Vision Transformer):** Aunque es una arquitectura basada en transformadores, también puede beneficiarse de funciones de activación no lineales para mejorar su capacidad de aprendizaje.

3.1.10. Imagenología:

La imagenología es el conjunto de técnicas y procedimientos que permiten obtener imágenes biológicas o médicas del cuerpo humano con el propósito de examinar y diagnosticar enfermedades. La imagenología biológica, o bioimaging en inglés, tiene como objetivo la detección de enfermedades infecciosas, cáncer y condiciones inflamatorias a partir del estudio de las células (citología) y del tejido (histopatología). La imagenología médica tiene como objetivo el estudio de los órganos humanos a partir de imágenes obtenidas por diferentes métodos, como la resonancia magnética (magnetic resonance imaging, MRI), la tomografía computarizada (computed tomography, CT), el ultrasonido, la tomografía por emisión de positrones (positron emission tomography, PET), los rayos X 2D y 3D, entre otros [8].

Los recientes avances en la tecnología han permitido el desarrollo de una amplia gama de equipos y sistemas de adquisición de imágenes de alta resolución. La interpretación, hecha por médicos, está sujeta al error humano y a las diferencias de diagnóstico entre intérpretes. Es por esto que la interpretación de imágenes biológicas y médicas se constituye como un área de investigación potencial donde nuevas metodologías como las ANNs tienen cabida, permitiendo asistir el dictamen médico y contribuir a un diagnóstico más objetivo y preciso. Esta área de investigación es conocida como diagnóstico asistido por computador (computer-aided diagnosis, CAD). La principal arquitectura de ANN usada en aplicaciones de CAD son las CNNs debido a su excelente desempeño en tareas de visión computarizada, y reconocimiento y clasificación de patrones en imágenes [17].

3.1.11. Mamografía Digital

La mamografía es una técnica de radiología basada en los distintos grados de absorción de los rayos X por los diferentes componentes en los tejidos blandos de la mama. A pesar de que la dosis media absorbida por la glándula mamaria en cada placa es una dosis muy baja (1,0-1,5 mGy), cuyo riesgo se considera mínimo, la exposición a radiaciones ionizantes en dosis acumulativas y potencialmente carcinógenas es una preocupación constante.

Una mamografía digital utiliza la misma tecnología de rayos X que las mamografías convencionales, pero en lugar de usar película, se utilizan detectores de estado sólido para registrar el patrón de rayos X que pasa a través de la mama. Estos detectores convierten los rayos X que pasan a través de ellos en señales electrónicas que se envían a una computadora. Luego, la computadora convierte estas señales electrónicas en imágenes que se pueden mostrar en un monitor y también almacenarse para su uso posterior. Varias ventajas del uso de la mamografía digital sobre la mamografía de película incluyen: la capacidad de manipular el contraste de la imagen para una mejor claridad, la capacidad de utilizar la detección asistida por computadora de anomalías y la capacidad de transmitir fácilmente archivos digitales a otros expertos para una segunda opinión. Además, las mamografías digitales pueden disminuir la necesidad de las retomas, que son comunes con la mamografía de película debido a técnicas de exposición incorrectas o problemas con el desarrollo de la película. Como resultado, la mamografía digital puede conducir a una menor exposición a los rayos X. Hasta la fecha, no hay evidencia de que la mamografía digital sea mejor que la mamografía de película para reducir el riesgo de una mujer de morir de cáncer de mama, sin embargo, la detección digital puede ser más precisa para encontrar cánceres en mujeres más jóvenes o mujeres con senos densos [18].

3.2. ANTECEDENTES / TRABAJOS RELACIONADOS

El cáncer de mama es una de las principales causas de muerte en mujeres en todo el mundo y también representa el 19,7% de todos los casos de cáncer en mujeres, se presenta en una edad promedio de 60 años en Colombia identificándose con mayor incidencia en las regiones de la Costa Atlántica y la región Andina. La detección temprana y el diagnóstico preciso de esta enfermedad son fundamentales para mejorar las posibilidades de curación y la calidad de vida de las pacientes.

La mamografía es una técnica que permite examinar el tejido mamario y detectar posibles anomalías que podrían ser signos de cáncer. Sin embargo, la interpretación de las imágenes no es una tarea sencilla, sino que implica una gran carga de trabajo, repetición y subjetividad por parte del radiólogo, lo que puede aumentar el riesgo de error humano. Por eso, es necesario buscar soluciones que faciliten y mejoren el diagnóstico de las lesiones mamarias.

El uso de técnicas de inteligencia artificial (IA) en la detección y diagnóstico del cáncer de mama ha ganado considerable atención en la última década. Diversos estudios han explorado y demostrado la eficacia de distintas arquitecturas de modelos de aprendizaje profundo para el análisis de imágenes mamográficas, con el objetivo de mejorar la precisión y eficiencia en la detección de cáncer de mama.

Uno de los estudios destacados es el realizado por la Universidad de Medicina China de Anhui en 2021, titulado "Supervised breast cancer prediction using integrated dimensionality reduction CNN"[19]. Este estudio empleó una técnica supervisada de reducción de dimensionalidad integrada

con redes neuronales convolucionales (CNN) para la predicción de cáncer de mama, utilizando 3000 imágenes mamográficas. Los resultados mostraron una precisión del 82.5% en el test y del 84.3% en la validación, destacando que la reducción de dimensionalidad puede mejorar la eficiencia del modelo sin incrementar su complejidad. Además, el estudio se centró en medir la precisión, sensibilidad y especificidad del modelo, utilizando imágenes en formato DICOM, convertidas posteriormente a PNG para su procesamiento. Fortalezas: mejora la eficiencia del modelo con reducción de dimensionalidad. Debilidades: necesidad de datos etiquetados y riesgo de sobreajuste.

Otro estudio significativo es el llevado a cabo por la Universidad de Sharjah en 2021, denominado "Breast cancer detection using artificial intelligence techniques [20]. Este estudio empleó una combinación de múltiples técnicas de IA, incluidas CNN, máquinas de vectores de soporte (SVM), redes neuronales profundas (DNN), transformers y métodos de aumento de datos, para la detección de cáncer de mama en 2800 imágenes mamográficas. Los resultados indicaron una precisión del 85% en el test y del 86.2% en la validación, demostrando que la integración de diversas técnicas de IA puede mejorar la precisión en la detección de cáncer de mama. Este estudio también utilizó imágenes en formato DICOM y resaltó la importancia de desarrollar técnicas que mejoren la interpretabilidad de los modelos para su uso clínico. Fortalezas: integración de diversas técnicas mejora la precisión. Debilidades: complejidad en la interpretación de los modelos y necesidad de grandes cantidades de datos.

En Italia para la vigencia 2023, titulado "Convolutional Networks and Transformers for Mammography Classification"[21], exploró la combinación de redes neuronales convolucionales (CNN) y transformers para la clasificación de mamografías, utilizando 3500 imágenes mamográficas. Los resultados mostraron una precisión del 89.5% en el test y del 90% en la validación, indicando que los transformers pueden mejorar significativamente la precisión en la clasificación de imágenes mamográficas cuando se combinan con CNN. Este estudio subrayó la necesidad de optimizar los hiperparámetros de los modelos para mejorar su rendimiento y utilizó principalmente imágenes en formato DICOM. Fortalezas: alta precisión y optimización de modelos. Debilidades: alta demanda de recursos computacionales y complejidad en el ajuste de hiperparámetros.

En Japón, un estudio titulado "Quality control system for mammographic breast positioning using deep learning 2023"[22], realizado en 2023 por la Universidad de Niigata, la Universidad de Ciencias Médicas de Hiroshima y el Hospital de Cáncer de Osaka, desarrolló un sistema de control de calidad para el posicionamiento mamográfico utilizando técnicas de deep learning. Utilizando 1631 vistas mediolaterales oblicuas mamográficas, el estudio mostró mejoras en la calidad del posicionamiento mamográfico, con precisiones de 78.36% y 72.78% para IMF y Nipple Profile respectivamente en el test, y 58.07% y 36.86% en la validación. Este estudio destaca la efectividad de los sistemas de control de calidad basados en deep learning para mejorar la precisión del posicionamiento mamográfico [23]. Además, este estudio identificó la integración de sistemas de control de calidad como un área clave para futuros desarrollos en la detección de cáncer de mama. Fortalezas: mejora la calidad del posicionamiento mamográfico. Debilidades: variabilidad en la precisión y necesidad de optimización continua.

Finalmente, un estudio realizado en 2023 por la Universidad de Tecnología de Información y Gestión de Olsztyn en Polonia, titulado "Automatic detection of breast masses using deep learning with YOLO approach"[24], aplicó el enfoque YOLO (You Only Look Once) para la detección automática de masas mamarias utilizando las bases de datos Vindr-Mammo y MIAS. Con una precisión del 80% en la validación interna y del 72% en la prueba externa, este estudio demostró que el uso de YOLOv5 es eficiente para la detección de nódulos mamarios, especialmente cuando se aplican técnicas de transferencia de aprendizaje y aumento de datos. Este estudio también resaltó la necesidad de aumentar la diversidad de los datos y mejorar la generalización de los modelos para su aplicación clínica. Fortalezas: eficiencia en la detección de nódulos con YOLOv5. Debilidades: necesidad de mejorar la generalización y diversidad de los datos.

Estos antecedentes muestran que las redes neuronales convolucionales (CNN) son una técnica prometedora para la detección y clasificación de cáncer de seno a partir de imágenes de mamografía. También se identifican grandes retos y limitaciones, como la escasez y heterogeneidad de los datos públicos, la metodología de la toma de las mamografías, calidad de las imágenes, el acceso a historias clínicas, entendimiento de los resultados, la falta de estandarización y validación de los métodos.

4. METODOLOGÍA

En el desarrollo de este proyecto, se plantea la metodología para la implementación de un modelo de aprendizaje profundo destinado a la identificación de nódulos en imágenes de mamografías, con el propósito de detectar el cáncer de seno. Esta metodología se enmarca en un proyecto de tipo práctico y experimental.

Se considera práctico debido a la implementación de un modelo real para abordar una problemática concreta: la identificación de nódulos en mamografías, contribuyendo así a la detección temprana del cáncer de seno. Por otro lado, se cataloga como experimental, ya que implica la realización de pruebas y experimentos durante el proceso de implementación. Esto incluye el entrenamiento, ajuste y validación del modelo de aprendizaje profundo, con la generación y evaluación de datos, así como la realización de ajustes en función de los resultados obtenidos para mejorar el rendimiento del modelo.

La metodología propuesta se estructura en varias etapas, destacando la selección de la base de datos (dataset) como una de las principales. Este paso resulta fundamental, dado que la calidad y cantidad de los datos de entrada influirán significativamente en el desempeño del modelo.



Figura N° 7. Metodología para desarrollo de la investigación- Creación propia

4.1. SELECCIÓN DE DATOS:

4.1.1. Búsqueda y recopilación de conjuntos de datos de mamografías médicas.

Durante el proceso de investigación, se han evaluado varias bases de datos públicas que contienen imágenes de mamografías. A continuación, presentamos un resumen de las bases de datos identificadas, sus características y limitaciones:

Base de Datos MIAS (The mini-MIAS Database 1994) [24].

- **Proveedor:** Instituto Nacional de Física e Ingeniería Nuclear del Reino Unido.
- **Contenido:** 322 imágenes digitales de mamografías, que representan a 161 pacientes, incluyendo mamografías izquierda y derecha. También se registra 208 imágenes normales, 63 imágenes con anomalías benignas y 51 imágenes con anomalías malignas
- **Tamaño de la data :** 2.3 GB.
- **Resolución:** 1024 x 1024 píxeles en formato PGM, con 8 bits por píxel.
- **Información Adicional:** Etiquetas que describen el tejido de fondo, la clase y la severidad de la anomalía, 2 imágenes por paciente, coordenadas y radio del centro de la anomalía. Las anomalías se clasifican en siete categorías: calcificación, masa bien definida/circunscrita, masa específica, masa mal definida, distorsión arquitectónica, asimetría y normal. Las imágenes con áreas sospechosas tienen información a nivel de píxel sobre la ubicación y el tipo de regiones sospechosas.

- **Limitaciones:** La base de datos es antigua, tiene baja resolución y un número limitado de imágenes.

Base de Datos DDSM (DDSM 1999) [25].

- **Proveedor:** Complex Adaptive Systems Laboratory, Instituto Nacional de Física e Ingeniería Nuclear del Reino Unido.
Contenido: 10,239 imágenes de mamografías correspondientes a 6.775 estudios de 1.566 pacientes. Presenta 1.500 casos malignos, 2.000 normales y 3275 Nódulos no cancerígenos ni malignos.
- **Tamaño:** 163 GB.
- **Resolución:** 1024 x 1024 píxeles en formato DICOM, con 8 bits por píxel.
- **Información Adicional:** Contiene casos normales, benignos y malignos con información patológica verificada. También se incluyen segmentaciones y cajas delimitadoras actualizadas de las regiones de interés (ROI) para las anomalías que indican una posición general de las lesiones, y el diagnóstico patológico.
- **Limitaciones:** La base de datos es un subconjunto de la base de datos DDSM, por lo que no contiene todos los casos y tipos de anomalías que se encuentran en la DDSM. Esto puede limitar la generalización y la comparación de los resultados obtenidos con esta base de datos. La base de datos contiene imágenes digitalizadas de películas mamográficas, que pueden tener una calidad y una resolución inferiores a las imágenes digitales directas. La base de datos proporciona el diagnóstico patológico solo para los datos de entrenamiento, pero no para los datos de prueba.

Base de Datos INbreast (2012) [26].

- **Proveedor:** Hospital Universitario en Porto, Portugal.
- **Contenido:** 115 casos con 410 imágenes, que incluyen mujeres con ambas mamas afectadas y pacientes mastectomizadas. Tiene 48 Casos malignos, 46 normales y 21 nódulos no son cancerosas ni malignas.
- **Tamaño:** 3,7 GB.
- **Resolución:** Variable, dependiendo del tamaño del pecho. El rango va desde 2816 x 2040 hasta 3328 x 2560 píxeles en formato DICOM, con 14 bits por píxel.
- **Información Adicional:** Etiquetas que indican el tipo y la severidad de la anomalía, las coordenadas del centro y el radio aproximado de la anomalía. Las anomalías se clasifican en cuatro categorías: calcificación, masa, asimetría y distorsión. Las imágenes con anomalías tienen anotaciones ROI que indican una posición general de las anomalías, pero no una segmentación precisa para ellas. La base de datos también proporciona

información clínica adicional sobre los pacientes, como la edad, el historial familiar y los resultados del examen físico.

- **Limitaciones:** Requiere software específico para procesar imágenes en formato DICOM y la el bajo volumen de registros. La base de datos tiene un número limitado de casos e imágenes en comparación con otras bases de datos más grandes. La base de datos tiene una distribución desequilibrada entre las clases de anomalías, siendo las masas las más frecuentes y las asimetrías y las distorsiones las menos frecuentes. La base de datos no proporciona información patológica verificada para los casos de anomalías.

Base de Datos CBIS-DDSM Curated Breast Imaging Subset of DDSM (2017) [27].

- **Proveedor:** Instituto Nacional de Estándares y Tecnología (NIST), EE.UU.
- **Contenido:** 753 imágenes de mamografías, correspondientes a 660 casos de pacientes. Presenta 310 casos benignos y 350 casos malignos según información de patología verificada
- **Resolución:** Alta resolución espacial (70 micras por píxel), alta resolución de contraste (12 bits por píxel).
- **Tamaño:** 43 GB.

- **Información Adicional:** Etiquetas que indican el tipo y la severidad de la anomalía, las coordenadas del centro y el radio aproximado de la anomalía. Las anomalías se clasifican en cuatro categorías: calcificación, masa, asimetría y distorsión arquitectónica. Las imágenes con anomalías tienen segmentaciones ROI y cuadros delimitadores actualizados de las regiones de interés, que proporcionan una segmentación precisa para las anomalías.

La base de datos CBIS-DDSM es una mejora de la base de datos DDSM, que es una base de datos ampliamente utilizada y reconocida en el campo del diagnóstico asistido por ordenador para el cáncer de mama. La base de datos CBIS-DDSM contiene imágenes descomprimidas y convertidas al formato DICOM, contiene anotaciones más precisas y actualizadas que la base de datos DDSM, lo que facilita el desarrollo y la evaluación de sistemas de apoyo a la decisión para la detección y el diagnóstico del cáncer de mama.

- **Limitaciones:** La base de datos CBIS-DDSM es un subconjunto de la base de datos DDSM, por lo que no contiene todos los casos y tipos de anomalías que se encuentran en la DDSM. Esto puede limitar la generalización y la comparación de los resultados obtenidos con esta base de datos. La base de datos CBIS-DDSM contiene imágenes digitalizadas de películas mamográficas, que pueden tener una calidad y una resolución inferiores a las imágenes digitales directas.

Proporciona el diagnóstico patológico solo para los datos de entrenamiento, pero no para los datos de prueba y tiene una estructura con múltiples identificadores de paciente, lo que puede causar confusión.

Base de Datos VinDr-Mammo (2022)

- **Proveedor:** VinBrain, Vietnam [28].
- **Contenido:** 5,000 exámenes de mamografía de un total de 20.000 imágenes, cada uno con cuatro vistas estándar. se divide en dos conjuntos: el conjunto de entrenamiento y el conjunto de prueba. El conjunto de entrenamiento contiene 3.000 exámenes (anotaciones detalladas, mientras que el conjunto de prueba contiene 2.000 exámenes (8.000 imágenes) sin anotaciones. De los 4.000 exámenes del conjunto de entrenamiento. De los 1.000 exámenes del conjunto de prueba, se desconoce el número exacto de casos benignos y malignos
- **Resolución:** Variable, dependiendo del tamaño del pecho. El rango va desde 2816 x 3584 hasta 3328 x 4096 píxeles en formato DICOM, con 16 bits por píxel.
- **Tamaño:** GB.
- **Información Adicional:** Etiquetas que indican el tipo y la severidad de la lesión, las coordenadas del centro y el radio aproximado de la lesión. Las lesiones se clasifican en cuatro categorías: calcificación, masa, asimetría y distorsión arquitectónica. Las imágenes con lesiones tienen anotaciones ROI que indican una posición general de las lesiones, pero no una segmentación precisa para ellas.
También proporciona información clínica adicional sobre los pacientes, como la edad, los resultados del examen físico y el BI-RADS.
- **Limitaciones:** La base de datos es nueva y aún no ha sido ampliamente utilizada o evaluada por la comunidad científica. La base de datos tiene una distribución desequilibrada entre las clases de lesiones, siendo las calcificaciones y las masas las más frecuentes y las asimetrías y las distorsiones arquitectónicas las menos frecuentes [29].

4.1.2. Análisis de las características de los datos.

Se desarrollo una priorización de las diferentes bases de datos desarrollando una calificación del contenido de cada estudio:

Se toma la decisión de apoyar el estudio basado en la base VinDr-Mammo y el CBIS-DDSM tienen datos o coordenadas de entorno del hallazgo. Estos datos son útiles para localizar y segmentar el hallazgo en la imagen, así como para medir su tamaño y forma.

Continuación se realizó una calificación de los estudios:

Estudio	MIAS	DDSM	INbreast	CBIS-DDSM	VinDr-Mammo
Relevancia	3	4	5	4	5
Actualidad	1	2	4	5	5
Representatividad	2	3	4	5	5
Disponibilidad y acceso	5	4	3	5	5
Calidad de las imágenes	2	3	5	5	5
Cantidad de registros	2	5	2	3	5
Coordenadas y área de hallazgo	0	0	0	5	5
Calificación total	15	21	23	32	35

Tabla N° 8 Comparativa de las Diferentes bases de datos analizadas -creación Propia

Se define utilizar la base de datos VinDr-Mammo puesto que tiene mayor calificación y adicionalmente ofrece datos con coordenadas de entorno del hallazgo. Estos datos son útiles para localizar y segmentar el hallazgo en la imagen, así como para medir su tamaño y forma, al igual también tiene un volumen de imágenes considerable para su análisis

La base de datos VinDr-Mammo contiene 5,000 exámenes de mamografía, cada uno con cuatro vistas estándar (craniocaudal y mediolateral de cada seno) ejemplo:

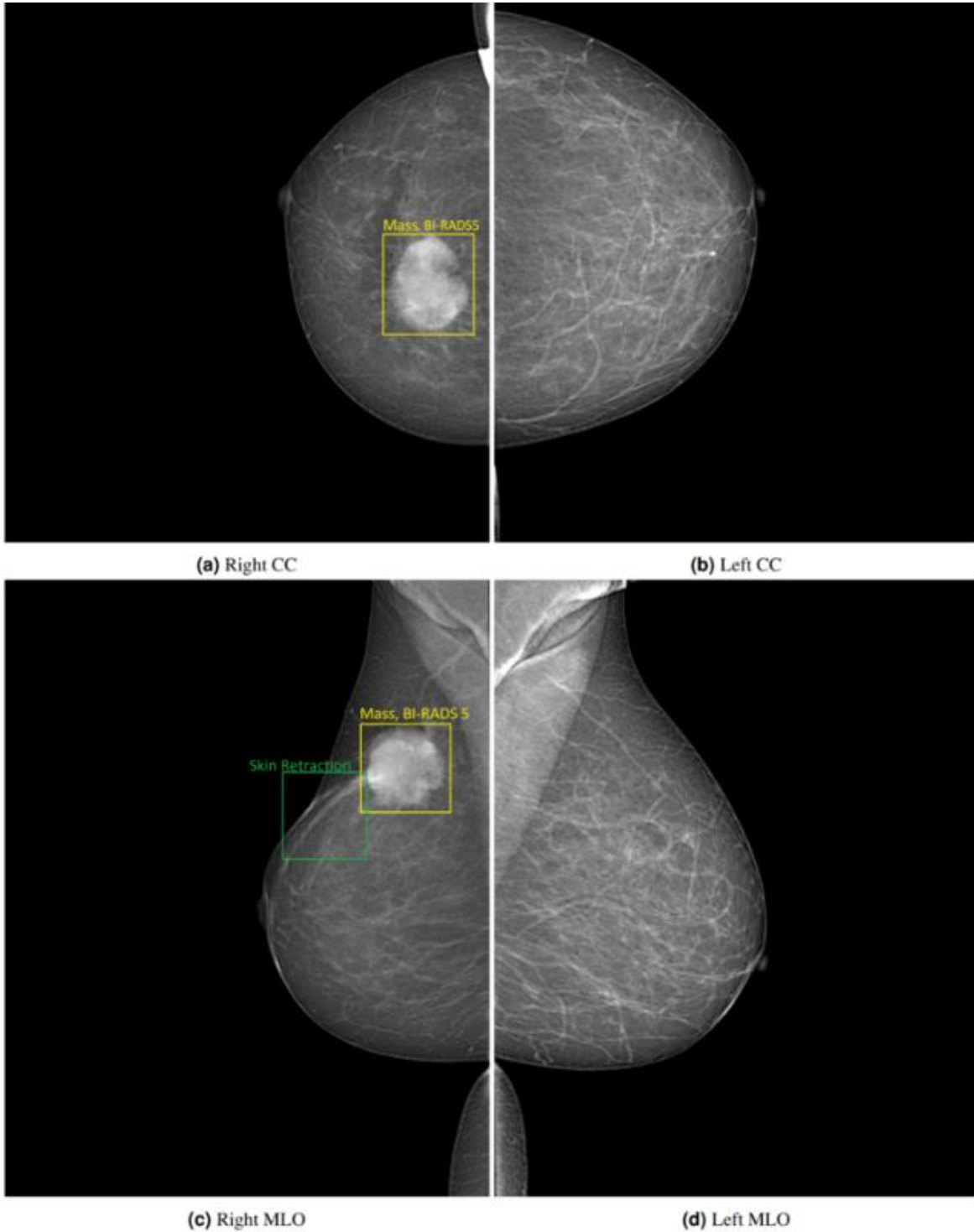


Figura N° 9 VinDr-MammoVinDr-Mammo

En la imagen de ejemplo se aprecia que el examen de mamografía de muestra con la mama derecha evaluada un BI-RADS 5, densidad B, y la mama izquierda con BI-RADS 1, densidad B. CC denota craneocaudal y MLO denota oblicuo medio lateral [30].

Esta base de datos está dividida en:

Tipo de Split	Cantidad de estudios	Cantidad de imágenes	Cantidad de imágenes con nódulos
test	1.000	4.000	452
training	4.000	16.000	1.796
Total general	5.000	20.000	2.245

Figura N° 10 Data análisis de Tipso de Split- Creación Propia

Se identifica que de los 5000 estudios 1000 son de prueba y 4000 de entrenamiento con un total de 20.000 imágenes (4.000 Prueba y 16.000 Entrenamiento) la cantidad total de nódulos por imagen encontrados son 2.245.

Los exámenes están anotados con información relevante, como la evaluación BI-RADS a nivel de mama y la densidad mamaria.

Dentro de la descripción se identifica la evaluación del nivel del seno y los hallazgos según el sistema BI-RADS (Breast Imaging Reporting and Data System), que es un sistema estandarizado para clasificar las anomalías mamarias. El BI-RADS tiene una escala prevista del 0 al 6, donde:

- 0 significa que se necesita una evaluación adicional,
- 1 significa que el seno es normal,
- 2 significa que hay hallazgos benignos,
- 3 significa que hay hallazgos probablemente benignos,
- 4 significa que hay hallazgos sospechosos,
- 5 significa que hay hallazgos altamente sospechosos
- 6 significa que hay hallazgos malignos confirmados.

En la base de datos se identifican los siguientes resultados:

Breast_birads	Cantidad de estudios	Cantidad de imágenes	Cantidad de imágenes con nódulos
BI-RADS 1	4188	13406	1
BI-RADS 2	1867	4676	13
BI-RADS 3	450	930	845
BI-RADS 4	373	762	972
BI-RADS 5	113	226	426

Total, general	5000	20000	2245
-----------------------	-------------	--------------	-------------

Figura N° 11 Tabla Data análisis de BI-RADS inicial- creación Propia

Las anotaciones también incluyen la clasificación inicial de breast-BI-RADS obteniendo solo resultado de finding-BI-RADS 3,4,5.

Se identifican anotaciones de la densidad del seno según el sistema BI-RADS, que es una medida de la cantidad de tejido fibroglandular en relación con el tejido graso. El sistema BI-RADS tiene una escala del A al D, donde A significa que el seno es casi todo graso, B significa que hay tejido fibroglandular disperso, C significa que el seno es heterogéneamente denso y D significa que el seno es extremadamente denso.

La base de datos también proporciona información de la edad del paciente y las anotaciones indican la categoría, las coordenadas del centro y el radio aproximado de la lesión. Las lesiones se clasifican en cuatro categorías principales: calcificación, masa, asimetría y distorsión arquitectónica.

4.1.3. Obtención de permisos y aprobaciones éticas para el uso de datos médicos

La base de datos fue descargada de la ruta <https://vindr.ai/datasets/mammo>, en la cual nos relaciona el conjunto de datos VinDr-Mammo y nos redirecciona a Physionet:

Se realiza firma de acuerdo de uso de datos:

Acuerdo de uso de datos de signos: VinDr-Mammo: un conjunto de datos de referencia a gran escala para la detección y el diagnóstico asistidos por computadora en mamografía digital de campo completo v1.0.0

Firme el siguiente acuerdo de uso de datos para acceder a los archivos en [VinDr-Mammo: un conjunto de datos de referencia a gran escala para la detección y el diagnóstico asistidos por computadora en mamografía digital de campo completo v1. 0.0.](#)

Acuerdo de uso de datos de salud restringidos de PhysioNet 1.5.0

Si se me concede acceso a la base de datos:

1. No intentaré identificar a ningún individuo o institución a la que se haga referencia en los datos restringidos de PhysioNet.
2. Ejerceré todo cuidado razonable y prudente para evitar la divulgación de la identidad de cualquier individuo o institución a la que se haga referencia en los datos restringidos de PhysioNet en cualquier publicación u otra comunicación.
3. No compartiré el acceso a los datos restringidos de PhysioNet con nadie más.
4. Ejerceré todo cuidado razonable y prudente para mantener la seguridad física y electrónica de los datos restringidos de PhysioNet.
5. Si encuentro información dentro de los datos restringidos de PhysioNet que creo que podría permitir la identificación de cualquier individuo o institución, informaré la ubicación de esta información de inmediato por correo electrónico a PHI-report@physionet.org, citando la ubicación de la información específica en cuestión.
6. Utilizaré los datos con el único fin de uso legal en investigación científica y ningún otro.
7. Acepto contribuir con el código asociado a las publicaciones que surjan de estos datos en un repositorio abierto a la comunidad de investigación.
8. Este acuerdo puede ser rescindido por cualquiera de las partes en cualquier momento, pero mis obligaciones con respecto a los datos de PhysioNet continuarán después de la rescisión.

Estoy de acuerdo

No estoy de acuerdo

Acceso concedido - VinDr-Mammo: un conjunto de datos de referencia a gran escala para la detección y el diagnóstico asistidos por ordenador en mamografía digital de campo completo v1.0.0

Gracias por firmar el acuerdo de uso de datos para VinDr-Mammo: un conjunto de datos de referencia a gran escala para la detección y el diagnóstico asistidos por computadora en mamografía digital de campo completo v1.0.0- /span>

Figura N° 12 Acuerdo de uso de la fuente de datos

Files
Total uncompressed size: 337.8 GB.

Access the files

- Download the ZIP file (50.7 GB)
- Download the files using your terminal: `wget -r -N -c -np --user fma1donado2703 --ask-password https://physionet.org/files/vindr-mammo/1.0.0/`

Folder Navigation: <base>

Name	Size	Modified
images		
LICENSE.txt	2.3 KB	2022-03-21
SHA256SUMS.txt	2.7 MB	2022-03-22
breast-level_annotations.csv	2.7 MB	2021-12-08
finding_annotations.csv	3.3 MB	2021-12-08
metadata.csv	4.2 MB	2021-12-08

Figura N° 13 Imagen descargue dataset

<https://www.physionet.org/content/vindr-mammo/1.0.0/>

Se concluye que el Dataset es óptimo y proporciona información adecuada para llevar a cabo el estudio planificado.

4.2. PRE- PROCESAMIENTO DE DATOS:

4.2.1. Recursos Tecnológicos

Este proyecto aplica redes neuronales para analizar mamografías e identificar posibles casos de cáncer de mama. Para llevar a cabo este análisis, se utilizó Google Colab como plataforma inicial para el procesamiento de imágenes. Se optó por la licencia Colab Pro+, la cual requirió un pago de 400 unidades de computación adicionales, elevando el total a 500 unidades por mes. Es importante destacar que estas unidades de computación tienen una vigencia de 90 días y fue necesaria realizar su compra en dos ocasiones, Colab Pro+ ofrece diversas ventajas, como acceso a unidades de procesamiento gráfico (GPU) más rápidas, la posibilidad de cambiar a GPUs premium más potentes, y la capacidad de ejecutar cuadernos en segundo plano, lo que permitió que los procesamientos continuaran durante un máximo de 24 horas incluso después de cerrar el navegador. Sin embargo, se determinó que estas capacidades no eran suficientes para las necesidades del proyecto.

Para abordar esta limitación, se adquirió un equipo con 32 GB de RAM y una tarjeta gráfica Nvidia 3060. Esto permitió probar modelos más complejos y ejecutarlos durante períodos extendidos, incluso durante las noches, para aprovechar al máximo los recursos y reducir los tiempos de ejecución.

4.2.2. Normalización y estandarización de las imágenes

Se ha realizado cargue de las imágenes en Python para identificar sus características y la resolución, lo cual es un factor determinante para el adecuado resultado del estudio, encontrando que de acuerdo al peso de la imagen y la resolución de las imágenes es variable, dependiendo del tamaño del pecho y la toma, pero el formato DICOM tiene un valor adecuado de calidad de la imagen en las muestras analizadas.

De acuerdo con la información de las coordenadas del xmin, ymin, xmax, ymax de la lesión se generó del total de las imágenes con identificación de nódulos una carpeta con recortes de las 2.245 lesiones las cuales relacionamos en el siguiente código:

Python

```
import numpy as np
import cv2
import pydicom as dicom
import os
import csv

def CargarImagen(xmin,xmax,ymin,ymax,ruta,key):
    xmin_r = 0 if round(xmin) < 0 else round(xmin) #
    Coordenada X mínima
    xmax_r = 0 if round(xmax) < 0 else round(xmax) #
    Coordenada X máxima
    ymax_r = 0 if round(ymax) < 0 else round(ymax) #
    Coordenada Y máxima
    ymin_r = 0 if round(ymin) < 0 else round(ymin) #
    Coordenada Y mínima
    rutaAbrir = ruta

    ds = dicom.dcmread(rutaAbrir)
    imagen = ds.pixel_array
    imagen_recortada = imagen[ymin_r:ymax_r,
xmin_r:xmax_r]

    imagen = cv2.normalize(imagen_recortada, None, 0,
255, cv2.NORM_MINMAX)

imagen = imagen.astype(np.uint8)
ds.PixelData = imagen_recortada.tobytes()
ds.Rows, ds.Columns = imagen_recortada.shape
ds.save_as(f'mask\{key}.dicom')

with open ('finding_annotacions.csv', mode='r') as
finding:

    lector_csv = csv.DictReader(finding, delimiter=';')
    # count=0

    for i, fila in enumerate(lector_csv):
        nameImg= fila['image_id']
        nameDir= fila['study_id']
        ruta= os.getcwd() +
f"\{nameDir}\{nameImg}.dicom"
        if (not os.path.exists(ruta)):continue

        if (fila['finding_categories'] != "[No Finding]"):
            xmin= float(fila['xmin'])
```

```
xmax= float(fila['xmax'])  
ymin= float(fila['ymin'])  
ymax= float(fila['ymax'])  
key= fila['KEYS']  
    CargarImagen(xmin,xmax,ymin,ymax,ruta,key)
```

Posterior al desarrollo anterior se realiza una comparativa de las máscaras obtenidas en el código de Python con el recorte de la imagen directamente en el programa (Imagen) del cual se confirma en la muestra analizadas de 20 imágenes la presencia de una adecuada precisión de la información Presentada en la base de datos, Ejemplo:

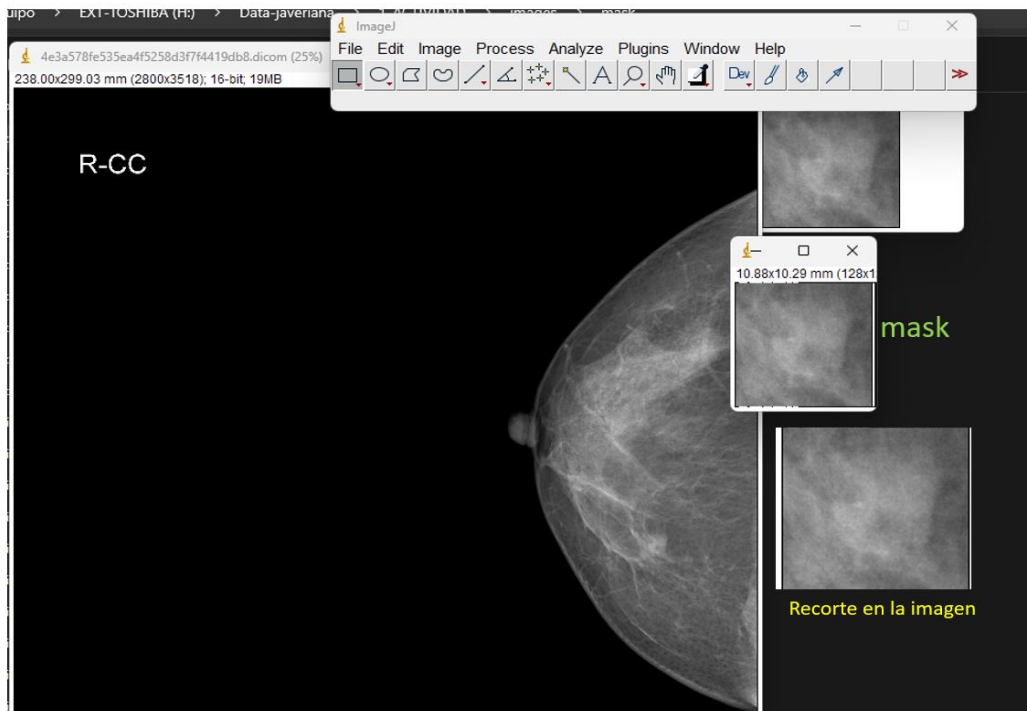


Figura N° 14 Ejemplo práctico de recortes en Python y Recorte en programa visualización.

En este estudio, analiza imágenes médicas en el formato DICOM, ampliamente utilizado como estándar para el intercambio de información radiológica específicamente para mamografías. Utilizamos la biblioteca pydicom para la lectura y manipulación de estas imágenes. Posteriormente, fue necesario llevar a cabo la conversión de las imágenes al formato PNG para su utilización como entrada en una red neuronal.

El formato DICOM es esencial en el ámbito médico, ya que asegura la compatibilidad entre diferentes equipos y sistemas, permitiendo una comunicación eficiente y segura en entornos clínicos. Esta interoperabilidad facilita una gestión más eficiente y segura de los datos médicos, crucial para diagnósticos precisos y la planificación terapéutica.

La elección del formato PNG se basó en varias consideraciones ya que este formato ofrece ventajas como su compatibilidad con diversas plataformas, programas y librerías, además de su capacidad para comprimir imágenes sin pérdida de calidad, resultando en una reducción en el espacio de almacenamiento y tiempos de carga más eficientes. A pesar de estas ventajas, es crucial tener en cuenta algunas desventajas potenciales, como la posible alteración del rango

dinámico de las imágenes, lo que podría afectar el contraste y la luminosidad, así como la introducción de ruido. Por lo tanto, se ha prestado especial atención a evaluar el impacto de esta conversión en el rendimiento de la red neuronal.

Es relevante señalar que durante la transformación las imágenes no experimentaron cambios significativos en resolución o nitidez, lo que las hace óptimas para su procesamiento en la red neuronal. Es importante destacar que, debido a las dimensiones variables de cada imagen, se ha llevado a cabo una cuidadosa evaluación para garantizar la coherencia en el procesamiento y la interpretación de los resultados.

Se llevó a cabo la preparación del conjunto de mamografías destinado a su implementación en el entrenamiento de la red neuronal. Este proceso implicó disponer de imágenes para estandarizar de resolución a 224x224 píxeles. fue necesario realizar recortes de las imágenes de manera que todas mantuvieran la misma resolución.

4.2.3. División de datos en conjunto de entrenamiento, validación y prueba

A partir de cada imagen completa, se identificaron aproximadamente 210 recortes potenciales, considerando la significativa variabilidad presente en estas tomas. Sin embargo, con el fin de optimizar la calidad de los datos, se descartaron aquellos recortes que exhibían valores elevados de negros. Esta medida garantiza que el conjunto de datos resultante sea coherente, homogéneo y apto para el entrenamiento de la red neuronal minimizando la incorporación de áreas no relevantes para la tarea específica de detección de cáncer de seno.

A continuación, en la figura14, se identifica un ejemplo del desarrollo de los recortes para las imágenes sin hallazgos:

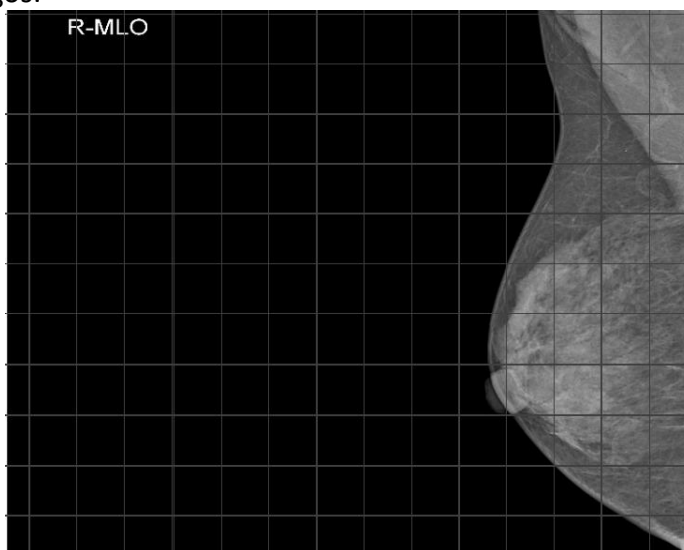


Figura N° 15 VinDr-Mammo Ejemplo preparación recortes

Durante el proceso de la generación de los recortes vinculados a las áreas de interés o hallazgo, definidas en el archivo CSV adjunto, se tuvo en cuenta el punto central del hallazgo lo cual nos lleva a una recalificación de los recortes, considerando especialmente que un hallazgo puede abarcar varios recortes. Esta particularidad impone una demanda adicional al modelo, ya que este debe comprender correctamente la distribución del hallazgo a lo largo de varios fragmentos. Inicialmente, se desarrolló redimensionar el hallazgo, esta estrategia generó alto consumo de procesamiento, pero no resultó óptima, como se evidenció en los resultados.

La necesidad de ajustar la base de datos de entrenamiento se hizo evidente durante las pruebas, ya que se identificaron aspectos que requerían una atención más detallada. Este ajuste se llevó a cabo con el propósito de obtener métricas más realistas y alineadas con la complejidad del problema. La adaptación de la base de datos permitió al modelo comprender de manera adecuada la variabilidad de los hallazgos y mejorar interpretación en la clasificación de las imágenes y detección de áreas de interés en mamografías.

A continuación, se presenta la forma como se preparan las imágenes para el ingreso de la información en la red neuronal de las zonas de interés o que presentan hallazgos.

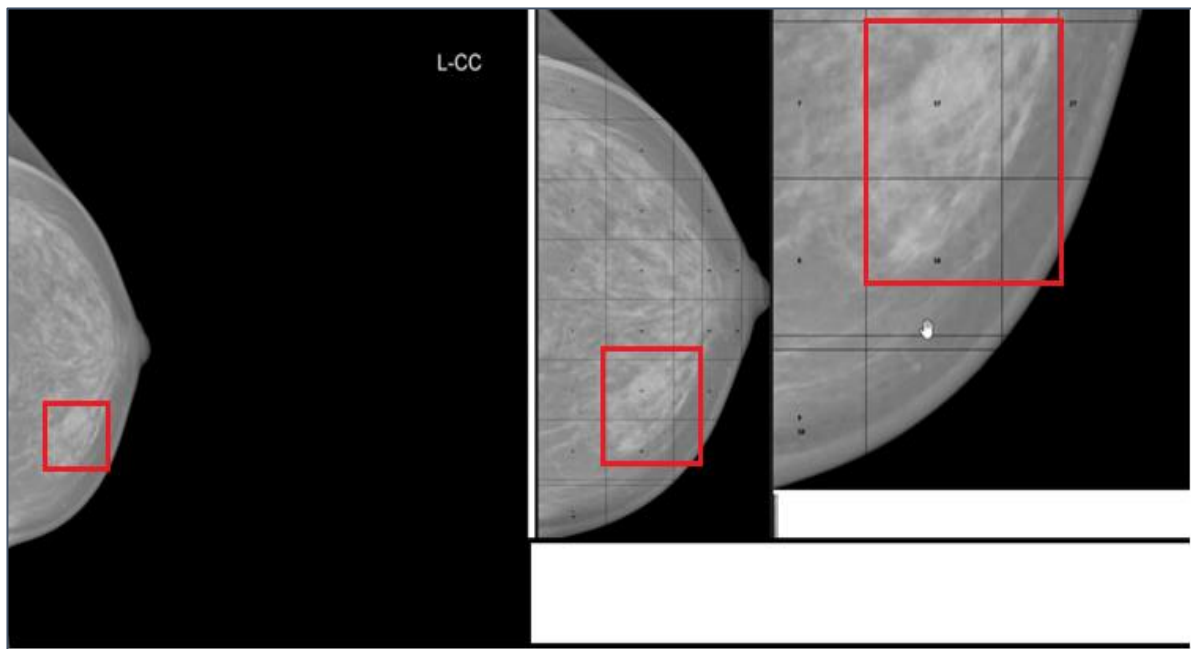
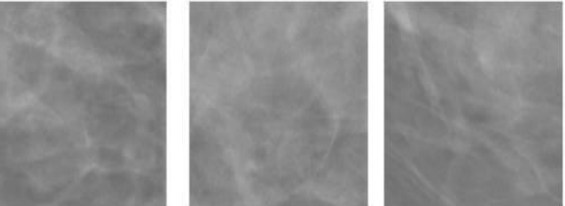
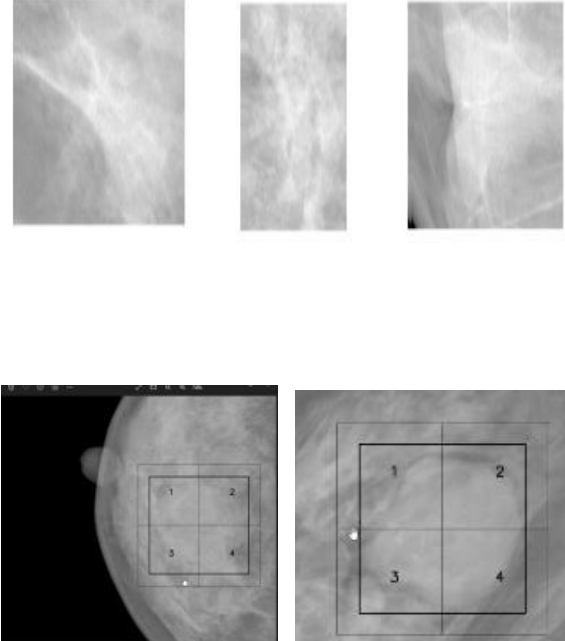


Figura N° 16 VinDr-Mammo Ejemplo preparación recortes

Para el desarrollo de las bases de entrenamiento y de prueba se reorganiza 3 carpetas que contienen:

TIPO IMAGEN	EJEMPLO	OBSERVACIÓN
<p>1-Base recortes con Sin hallazgos</p> <p>ENTRENAMIENTO</p>	<p>224 x 224 px</p> 	<p>Esta base contiene recortes de imágenes para los cuales se tiene certeza de que no presentan hallazgos contribuyendo al modelo a su capacidad para distinguir áreas sin anomalías.</p>
<p>2. Base de Recortes Con hallazgos</p> <p>ENTRENAMIENTO</p>	<p>217 x 180 px 196 x 267 px 224 x 224 px</p> 	<p>En esta base se incluyen demarcaciones de zonas de interés que contienen hallazgos. Estas pueden presentar dimensiones diversas, y es esencial ajustar la marcación para asegurar que las zonas de interés se capturen de manera apropiada en recortes de 224x224 píxeles. Durante las pruebas del modelo, se implementaron dos enfoques:</p> <ul style="list-style-type: none"> • Redimensionar la imagen seleccionado la zona de interés demarcada en el archivo. • Recortar la imagen completa centrada en el hallazgo y teniendo claro que este quedara distribuido en varios recortes. <p>Estas estrategias buscan optimizar la representación de los hallazgos en los recortes, contribuyendo a mejorar la capacidad del modelo para identificar de manera efectiva áreas de interés.</p>

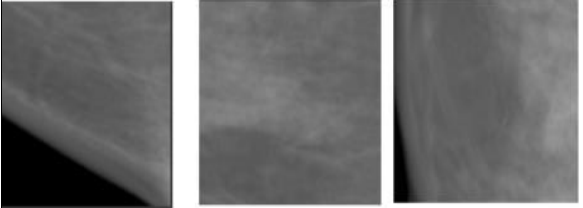
<p>3. Base de recortes para realizar validación Test</p> <p>EVALUACIÓN Y PRUEBA</p>	<p style="text-align: center;">224 x 224 px</p> 	<p>Estos recortes, marcados con o sin hallazgos, se utilizan específicamente para evaluar y validar el rendimiento del modelo. Forman parte de la fase de prueba del modelo, contribuyendo a medir su precisión en la identificación de hallazgos.</p>
--	---	--

Figura N° 17 VinDr-Mammo Ejemplos de recortes en general

4.2.4. Revisión Para aumentar la variabilidad del conjunto de datos de entrenamiento

Se han enfrentado diversas dificultades en la gestión de las bases de entrenamiento, lo que ha impulsado la exploración de nuevas estrategias para el manejo de recortes centrados en los hallazgos.

Inicialmente se dio un manejo basado en las coordenadas reportadas en el archivo adjunto de la base de imágenes como se identifica en la siguiente ilustración:

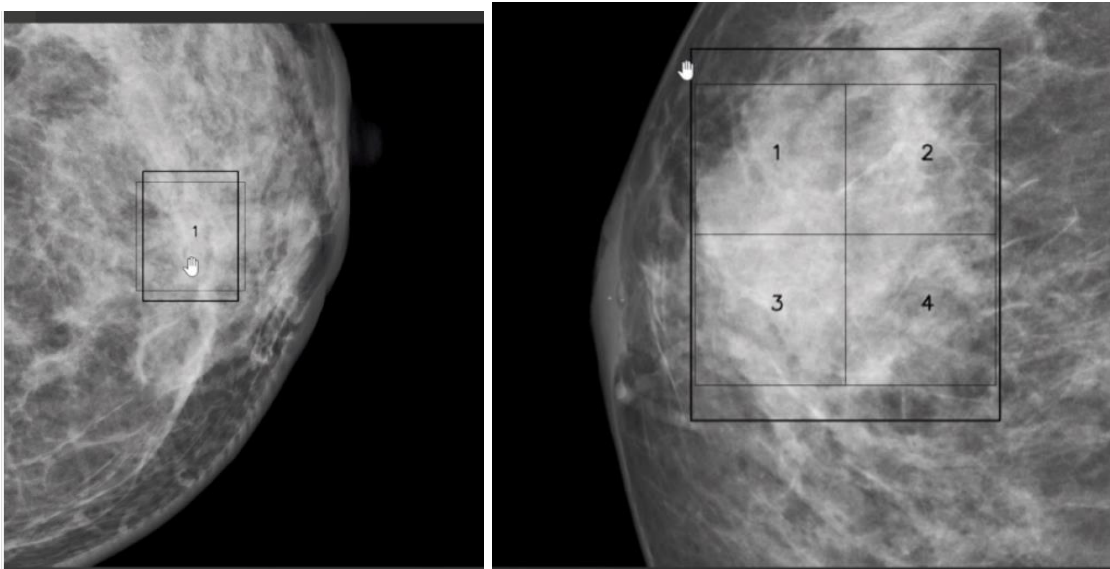


Figura N° 18 VinDr-Mammo Ejemplo preparación recortes áreas de interés

A partir de las coordenadas, se construye el conjunto de datos de entrenamiento con hallazgos, sin embargo, se enfrenta a diversos problemas de dimensionalidad. Para abordar esta situación y con la necesidad de establecer las dimensiones en 224x224, Se generan dificultades al implementar el modelo para comprender los hallazgos, lo que impacta negativamente en el desarrollo de las pruebas.

Ante estos inconvenientes, se decide cambiar la estrategia y generar la base de entrenamiento con hallazgos realizando recortes homogéneos en toda la imagen de la siguiente manera:

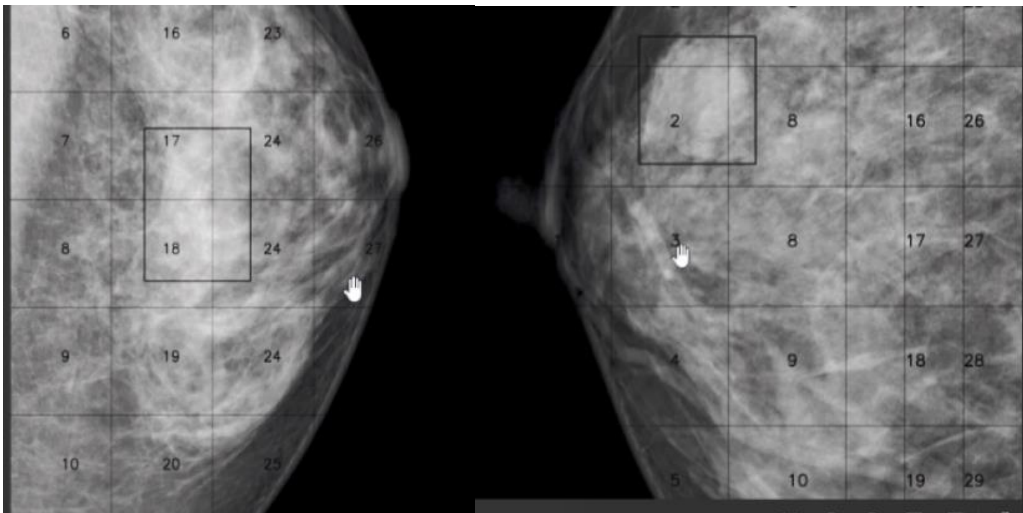


Figura N° 19 VinDr-Mammo Ejemplo preparación recorte áreas de interés selección de hallazgo

Se realizan recortes Homogéneos en el formato 224x224 y se identifica las coordenadas de los hallazgos seleccionando los recortes que tienen un 70 % de la zona de interés o del Hallazgo. Como se identifica en la imagen N°18 para nuestro data de hallazgos se tomarían los recortes 18 y18; que en este ejemplo tienen más del 70 % y para la imagen N°18 tomaríamos únicamente el recorte 2.

Como parte de este proceso exploratorio, se ha planteado la idea de entrenar o poner a prueba el modelo con otras bases de mamografías. Para analizar la capacidad del modelo en la interpretación de esta información, proporcionando así una perspectiva más sólida y adaptable en la detección de hallazgos

4.3. DISEÑO DEL MODELO

Después de completar la preparación de los datos, y la investigación de las arquitecturas, es crucial definir la arquitectura a implementar. Se han seleccionado varias arquitecturas de redes neuronales convolucionales (CNN) que se implementarán y compararán en términos de rendimiento y eficacia. Estas arquitecturas son:

- **VGG16 y VGG19:** Son redes neuronales profundas que procesan imágenes en capas sucesivas, capturando características complejas a medida que avanzan en la red.
- **ResNet y ResNet50:** Introducen conexiones especiales que ayudan a evitar la pérdida de información en redes muy profundas, permitiendo un mejor aprendizaje de características importantes.
- **ViT (Vision Transformer):** Utiliza un enfoque basado en la atención para procesar imágenes, dividiéndolas en parches y examinando relaciones entre ellos para reconocer patrones.

El procesamiento de imágenes utilizando TensorFlow es una parte crucial de este proyecto, ya que implica la construcción y entrenamiento de modelos de aprendizaje profundo para identificar posibles casos de cáncer de mama en mamografías. Para lograr esto de manera eficiente y efectiva, se requiere un entorno de trabajo adecuado que proporcione las herramientas y recursos necesarios para el desarrollo y la experimentación.

El establecimiento de un entorno de trabajo Linux integrado con una GPU para el entrenamiento de modelos de redes neuronales convolucionales (CNN) mediante TensorFlow y Keras implicó un proceso complejo y técnico que demandó una meticulosa selección de la distribución de Linux, una instalación precisa de los controladores de gráficos Nvidia compatibles con la versión específica de TensorFlow utilizada y el kernel de Linux instalado.

Posteriormente, se afrontaron desafíos adicionales relacionados con la configuración de TensorFlow para optimizar el aprovechamiento de la GPU, incluyendo la instalación de la versión correcta de TensorFlow, la configuración de variables de entorno y las opciones de compilación adecuadas para habilitar la aceleración por GPU.

A pesar de los numerosos obstáculos y errores encontrados durante el proceso, se logró establecer un entorno de trabajo funcional, aunque no se alcanzó el máximo rendimiento posible. Sin embargo, esta configuración permitió finalmente el aprovechamiento de la GPU para el entrenamiento de un modelo CNN utilizando TensorFlow y Keras.

4.3.1. Preparación del dataset para la implementación de los modelos

Se disponen para el entrenamiento del modelo 4 fuentes de datos:

- **Imágenes de Hallazgos:** Esta fuente de datos proporciona patrones que representan máscaras o áreas de interés en las imágenes de mamografías. Estas máscaras resaltan regiones específicas del tejido mamario que pueden contener anomalías o cáncer de seno.
- **Imágenes sin hallazgos:** Contiene recortes de las imágenes de mamografías sin hallazgos.
- **Imágenes de TEST:** Esta fuente de datos se utiliza para evaluar el rendimiento del modelo una vez que ha sido entrenado. Contiene imágenes de mamografías que el modelo no ha visto previamente, lo que permite una evaluación imparcial de su capacidad para generalizar a nuevos datos.
- **labelas:** Es un archivo en CSV que contiene la data que indica si cada imagen contiene o no cáncer de seno. Estas etiquetas son esenciales para el entrenamiento supervisado del modelo y para evaluar su rendimiento durante el entrenamiento y la evaluación.

Se aplicaron funciones esenciales para cargar y preprocesar imágenes en un proyecto de aprendizaje automático con Tensor Flow. La función `load_and_preprocess_from_path_label` carga imágenes desde una ruta de archivo y sus etiquetas correspondientes, realizando operaciones de decodificación, normalización y conversión de escala de grises a RGB para preparar los datos de entrada del modelo. Por otro lado, la función `load_and_preprocess_from_test` se encarga de cargar y preprocesar imágenes de prueba, buscando las etiquetas asociadas en un DataFrame y utilizando la función anterior para realizar el procesamiento. Ambas funciones aseguran la consistencia y coherencia en el manejo de datos de imágenes, lo que es crucial para el éxito del entrenamiento y evaluación del modelo de aprendizaje automático.

El enfoque en la normalización de píxeles, la conversión de escala de grises a RGB y la asociación de etiquetas con imágenes de prueba destaca la importancia de un preprocesamiento riguroso de los datos de imagen para obtener resultados precisos del modelo. Además, el modularidad de estas funciones permite su fácil integración en este proyecto de aprendizaje automático.

Adicionalmente en la preparación de los conjuntos de datos de imágenes se definieron algunos parámetros importantes, como `AUTOTUNE`, que permite a TensorFlow optimizar la carga de

datos, y `BATCH_SIZE`, que determina el tamaño de los lotes de datos utilizados durante el entrenamiento.

Luego, se crea un Dataset para las imágenes de máscaras y otro para las imágenes de cortes, utilizando las rutas de archivo definidas anteriormente (Imágenes con hallazgos y las imágenes sin hallazgos). A continuación, se procede a asignar etiquetas a cada conjunto de datos: las imágenes con hallazgos se etiquetan como 1, mientras que las imágenes sin hallazgos se etiquetan como 0, se carga y preprocesa cada imagen en paralelo utilizando las funciones (`load_and_preprocess_from_path_label`) (`load_and_preprocess_from_test`), Finalmente, se mezclan los conjuntos de datos. Luego, se procede a crear lotes de datos para el entrenamiento y la prueba, utilizando el tamaño de lote definido anteriormente (`BATCH_SIZE`) y se almacenan en caché para su uso posterior en el entrenamiento del modelo.

Se realiza definición callbacks utilizados durante el entrenamiento de un modelo de aprendizaje automático. El callback `EarlyStopping` se configura para monitorear la métrica de precisión en el conjunto de validación (`val_accuracy`). Si no se observa mejora en esta métrica después de 8 épocas consecutivas, el entrenamiento se detendrá prematuramente. Además, este callback restaurará automáticamente los pesos del modelo a los mejores obtenidos durante el entrenamiento si se produce detención anticipada. Por otro lado, el callback `LearningRateScheduler` se utiliza para programar dinámicamente la tasa de aprendizaje durante el entrenamiento. En este caso, la función `scheduler` ajusta la tasa de aprendizaje exponencialmente después de las primeras 10 épocas, lo que puede ayudar a mejorar la convergencia del modelo y evitar el estancamiento del entrenamiento en mesetas. Estos callbacks son herramientas valiosas para controlar el proceso de entrenamiento y optimizar el rendimiento del modelo de manera eficiente.

Este proceso asegura que los datos estén preparados de manera adecuada y eficiente para su uso en el entrenamiento del modelo, incluyendo la asignación de etiquetas, el preprocesamiento de imágenes y la optimización del rendimiento mediante el uso de paralelismo y la caché.

Para cada medición e implementación de cada arquitectura se va realizar 2 modelos uno con entrenamiento ajustado y posterior se realizara un (fine-tuning) lo que implica adaptar un modelo se modificando algunas capas finales o se descongelan ciertas capas para permitir que el modelo se ajuste a los nuevos datos, conservando las características generales aprendidas

previamente. Esto permite aprovechar el conocimiento previo del modelo preentrenado y adaptarlo para obtener un mejor rendimiento en la tarea deseada

4.3.2. Diseño Modelo 1 VGG16

En la implementación del modelo VGG16, se generaron dos arquitecturas distintas. La primera se basa en la arquitectura VGG16 preentrenada, a la que se le realizaron modificaciones en los parámetros para optimizar su rendimiento en la tarea específica de detección de cáncer de mama en mamografías. La segunda arquitectura corresponde a una versión personalizada de VGG16, diseñada específicamente para adaptarse mejor a las características de nuestro conjunto de datos.

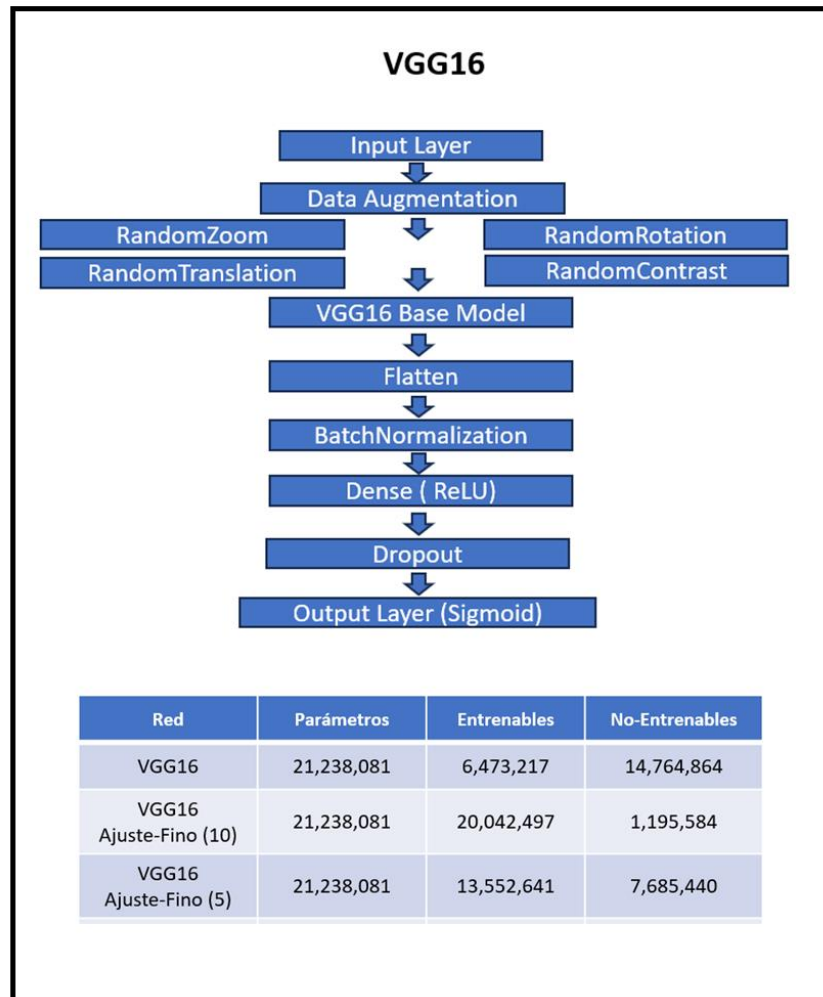


Figura N° 20 Modelo VGG16Ajuste de parámetros

El modelo VGG16 elegido se basa en un conocimiento preentrenado, aprovechando su capacidad para reconocer características generales en imágenes. El modelo está diseñado para realizar una clasificación binaria en mamografías, detectando la presencia o ausencia de cáncer de seno. La arquitectura del modelo se describe a continuación:

1. **Capa de Entrada (Input Layer):** Recibe las imágenes de mamografías.
2. **Aumento de Datos (Data Augmentation):**
 - **RandomZoom:** Realiza zoom aleatorio en las imágenes.
 - **RandomRotation:** Aplica rotaciones aleatorias.
 - **RandomTranslation:** Traduce las imágenes aleatoriamente.
 - **RandomContrast:** Ajusta el contraste de las imágenes aleatoriamente.
3. **Modelo Base VGG16 (VGG16 Base Model):** Red preentrenada en ImageNet.
4. **Aplanar (Flatten):** Convierte las características en un vector de una dimensión.
5. **Normalización por Lotes (BatchNormalization):** Normaliza las salidas de la capa anterior.
6. **Capa Densa (Dense):** Con activación ReLU.
7. **Dropout:** Para prevenir sobreajuste, desactiva aleatoriamente algunas neuronas durante el entrenamiento.
8. **Capa de Salida (Output Layer):** Con activación Sigmoid para la clasificación binaria

Su aplicación a imágenes médicas de alta resolución presenta desafíos en la interpretación. La arquitectura original de VGG16 incluye varias capas convolucionales, capas de agrupación máxima y normalización, permitiendo una extracción eficiente de características.

Para adaptar el modelo a las necesidades del proyecto, se añadieron capas de aplanamiento, normalización y capas densas, culminando en una capa de salida con activación sigmoide para la clasificación binaria. El modelo se compila utilizando una función de pérdida de entropía cruzada binaria y métricas de precisión, lo que optimiza el proceso de entrenamiento.

El proceso de ajuste fino fue crucial para mejorar el rendimiento del modelo. Se llevaron a cabo varios experimentos con diferentes configuraciones de capas descongeladas. En una configuración, se descongelaron 10 capas del modelo base, lo que permitió ajustar 6,473,217 parámetros entrenables de un total de 21,238,081, mientras que 14,764,864 permanecieron no entrenables. En otra variante, se mantuvieron la mayoría de las capas congeladas, dejando solo 1,195,584 parámetros ajustables. Esta estrategia de ajuste fino busca optimizar la capacidad del modelo para aprender características específicas del conjunto de datos.

La descongelación de capas también se realizó en configuraciones más limitadas, como en el caso de permitir el entrenamiento de solo 5 capas. En esta configuración, se ajustaron 7,685,440 parámetros, lo que permitió un enfoque más controlado para la adaptación del modelo

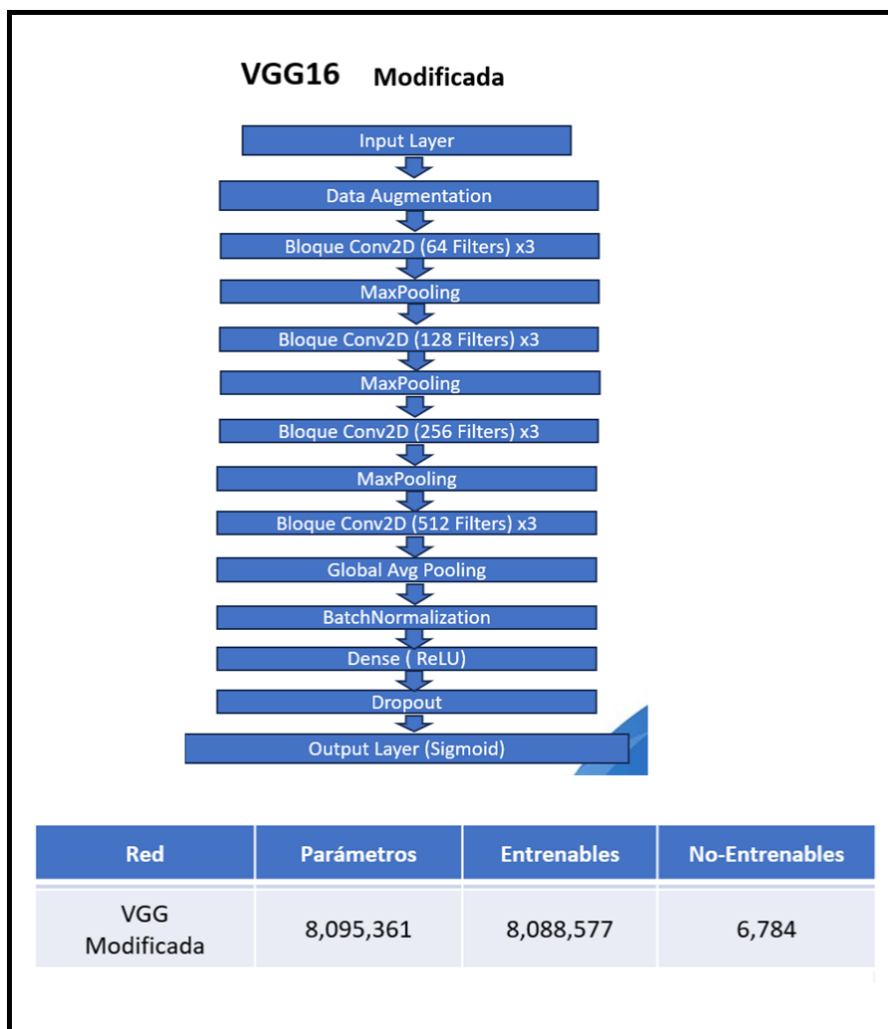


Figura N° 21 VGG16 Modificada

La versión modificada de VGG16, más eficiente, Comienza con una capa de entrada para recibir las imágenes y aplica aumento de datos para mejorar la generalización. Luego, pasa por varios bloques de capas convolucionales con filtros incrementales (64, 128, 256 y 512), cada uno seguido de una capa de MaxPooling para reducir la dimensionalidad. Después de extraer características a diferentes niveles de abstracción, una capa de Global Average Pooling

comprime estas características. La normalización por lotes se aplica para estabilizar el entrenamiento. Finalmente, el modelo incluye una capa densa con activación ReLU, seguida de una capa de dropout para prevenir el sobreajuste, y culmina con una capa de salida sigmoide para predecir la probabilidad de hallazgos en las imágenes.

cuenta con aproximadamente 8 millones de parámetros. Esta configuración optimiza la cantidad de parámetros entrenables y presenta una capa densa con 512 neuronas, mejorando la capacidad del modelo para captar características relevantes en las imágenes de mamografías. Esta arquitectura reducida facilita un entrenamiento más rápido y menos susceptible al sobreajuste.

4.3.3. Prueba Modelo 2 VGG19

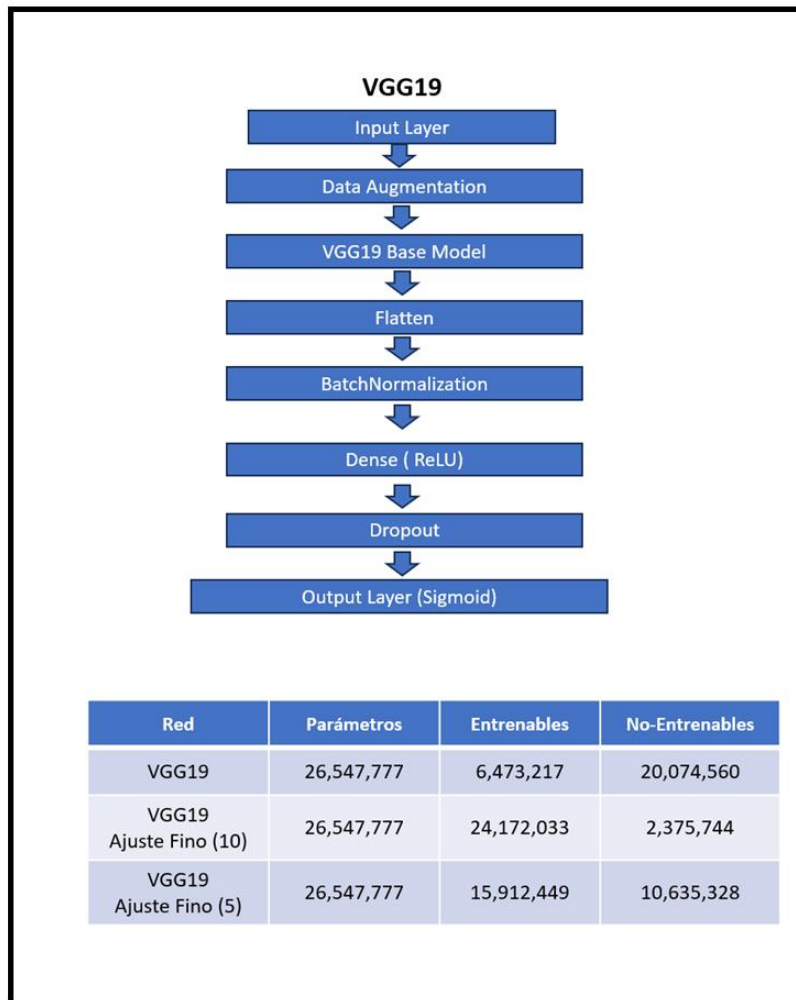


Figura N° 22 Modelo VGG19

En este proyecto, se implementaron tres variantes del modelo VGG19 para la clasificación binaria de imágenes de mamografías en la detección de cáncer de mama. Cada modelo se configuró con diferentes parámetros y enfoques de entrenamiento, optimizando su rendimiento en función de las necesidades específicas del conjunto de datos.

Aquí está una descripción de cada capa y su función:

1. **Input Layer (Capa de Entrada):** La capa de entrada acepta las imágenes de mamografías. Estas imágenes pasan a través de todas las capas subsiguientes para el procesamiento y clasificación.
2. **Data Augmentation (Aumento de Datos):** Esta etapa aplica técnicas de aumento de datos como rotaciones, cambios de escala, recortes, etc., para generar variaciones de las imágenes de entrada. Esto ayuda a aumentar la diversidad del conjunto de datos de entrenamiento y mejora la capacidad de generalización del modelo.
3. **VGG19 Base Model (Modelo Base VGG19):** Esta es la parte principal del modelo VGG19, una red neuronal convolucional profunda preentrenada en un gran conjunto de datos como ImageNet. La VGG19 contiene 19 capas, incluyendo 16 capas convolucionales y 3 capas completamente conectadas. En este modelo, probablemente se utilice solo la parte convolucional del VGG19 para extraer características de alto nivel de las imágenes de mamografías.
4. **Flatten (Aplanar):** Esta capa convierte las características multidimensionales extraídas por el modelo VGG19 en un vector unidimensional. Esto es necesario para conectar las capas convolucionales con las capas completamente conectadas.
5. **BatchNormalization (Normalización por Lotes):** Esta capa normaliza las activaciones de la capa anterior en mini-lotes, lo que acelera el entrenamiento y proporciona cierta regularización, mejorando así la estabilidad y rendimiento del modelo.
6. **Dense (ReLU) (Capa Densa con ReLU):** Esta es una capa completamente conectada (también conocida como capa densa) que utiliza la función de activación ReLU (Rectified Linear Unit). La función ReLU introduce no linealidades en el modelo y ayuda a capturar relaciones complejas en los datos.
7. **Dropout:** Esta capa apaga aleatoriamente una fracción de las neuronas durante el entrenamiento, lo que ayuda a prevenir el sobreajuste. La tasa de dropout especifica la fracción de neuronas que se apagará.
8. **Output Layer (Sigmoid) (Capa de Salida con Sigmoide):** La capa de salida utiliza la función de activación Sigmoid, que es ideal para tareas de clasificación binaria como la identificación

de hallazgos en mamografías. La función sigmoide produce una probabilidad entre 0 y 1, indicando la presencia o ausencia de hallazgos.

Modelo VGG19 Base: Este modelo de clasificación binaria utilizando la arquitectura VGG19 preentrenada como base. Se congela la base del modelo para evitar la actualización de sus pesos durante el entrenamiento. Luego se agregan capas de aplanamiento, normalización por lotes, densas con activación ReLU y abandono para regularización, seguidas de una capa densa de salida con activación sigmoide. El modelo se compila con el optimizador Adam y la función de pérdida de entropía cruzada binaria. La estructura básica incluye aproximadamente 26.5 millones de parámetros, con Entrenables: 24.4% y No Entrenables: 75.6%.

Durante la primera fase de entrenamiento, todas las capas del modelo base se congelan para evitar la actualización de sus pesos. Se añaden capas personalizadas, que incluyen una capa de aplanamiento, normalización por lotes y una capa densa con 512 neuronas. La capa de salida utiliza una activación sigmoide, adecuada para la clasificación binaria.

Modelo VGG19 con Ajuste Fino (10 Capas Descongeladas) Este modelo se basa en la misma arquitectura que el modelo base, pero se aplica el ajuste fino para mejorar el rendimiento. En este caso el Porcentaje de Parámetros entrenables es de 90.8% y No Entrenables de 9.2%.

Las capas seleccionadas para el ajuste fino permiten que el modelo aprenda características más específicas de las mamografías. Este modelo también se recompila con una tasa de aprendizaje más baja, optimizando así el proceso de entrenamiento sin perder las características generales aprendidas.

Modelo VGG19 con Ajuste Fino (5 Capas Descongeladas)

Este modelo, similar al anterior, también utiliza la arquitectura VGG19, pero se descongelan solo 5 capas del modelo base. En este caso entrenables 59.9% y No Entrenables 40.1% . La estrategia de descongelar menos capas permite un enfoque más conservador en el ajuste fino, ideal para evitar el sobreajuste en un conjunto de datos limitado. Este modelo también incluye capas densas de 512 neuronas y utiliza la activación sigmoide en la capa de salida.

4.3.4. Prueba Modelo RESNET

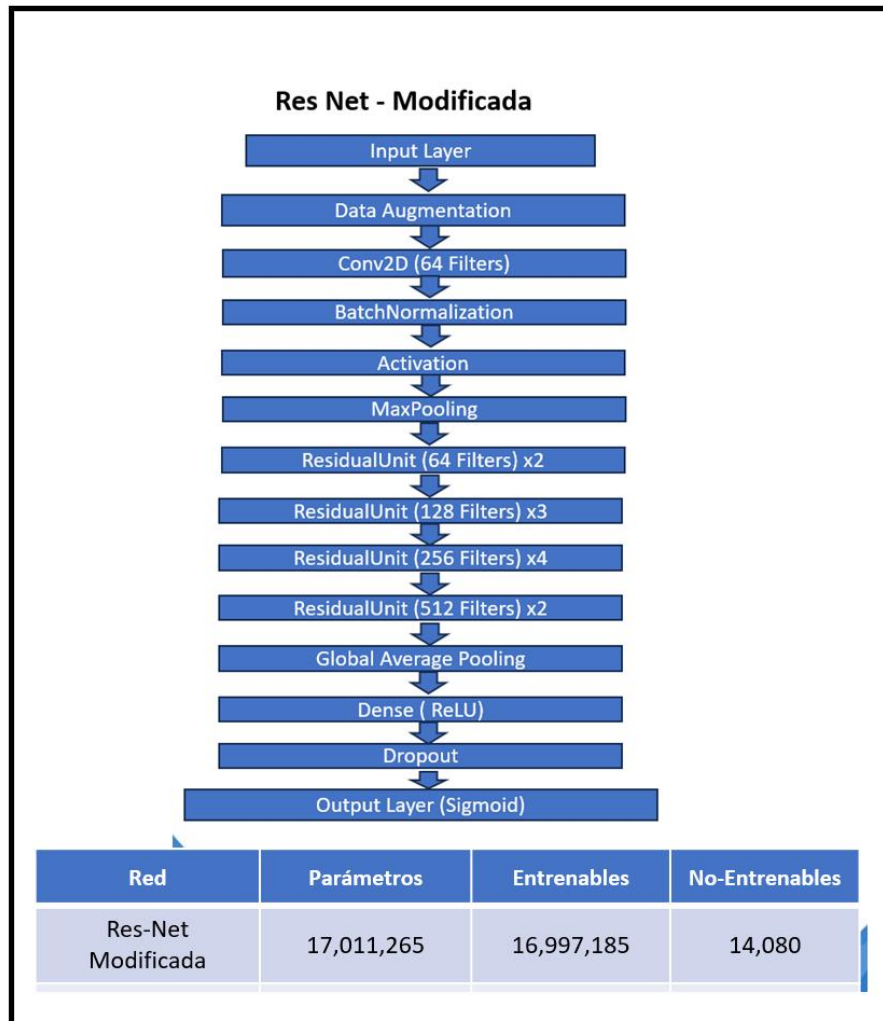


Figura N° 23 Resnet Modificada

El modelo ResNet modificado se implementó teniendo en cuenta varios factores cruciales para la detección de cáncer de seno a partir de mamografías. La estructura del modelo es la siguiente:

1. **Capa de Entrada (Input Layer):** Recibe las imágenes de mamografías.
2. **Aumento de Datos (Data Augmentation):** Aplica técnicas de aumento de datos como rotaciones y traslaciones para incrementar la variabilidad del conjunto de datos y mejorar la capacidad de generalización del modelo.
3. **Capa Convolutiva (Conv2D) con 64 Filtros:** Extrae características locales de la imagen.

4. **Normalización por Lotes (BatchNormalization):** Normaliza las activaciones de la capa anterior para estabilizar y acelerar el entrenamiento.
5. **Activación:** Introduce una función de activación no lineal (generalmente ReLU) para introducir no linealidad en el modelo.
6. **MaxPooling:** Reduce la dimensionalidad espacial de las características extraídas, manteniendo las características más relevantes.
7. **Bloques Residuales (Residual Units):** Facilitan el entrenamiento de redes profundas permitiendo la propagación de gradientes a través de atajos (skip connections). Se utilizan unidades residuales con diferentes números de filtros (64, 128, 256 y 512 filtros).
8. **Global Average Pooling:** Reduce cada mapa de características a un solo valor promedio, ayudando a prevenir el sobreajuste.
9. **Capa Densa (Dense) con Activación ReLU:** Introduce una capa totalmente conectada con función de activación ReLU para combinar las características extraídas.
10. **Dropout:** Aplica la técnica de dropout para evitar el sobreajuste, apagando aleatoriamente un porcentaje de las neuronas durante el entrenamiento.
11. **Capa de Salida (Output Layer) con Activación Sigmoid:** Proporciona la salida final del modelo. Dado que este es un problema de clasificación binaria (detección de cáncer de seno), se usa la función de activación Sigmoid para producir una probabilidad entre 0 y 1.

Configuración Adicional:

- **Scheduler de Tasa de Aprendizaje (LR Scheduler):** Sí, se utiliza un scheduler para ajustar dinámicamente la tasa de aprendizaje durante el entrenamiento.
- **Capas Densas (Dense Layers):** Se utiliza una capa densa de 1024 neuronas.
- **Unfreezing Layers:** No se especifican capas descongeladas.
- **Fine Tuning:** No se realiza afinamiento fino (fine-tuning).

Parámetros del Modelo:

- **Total de Parámetros:** 17,011,265
- **Parámetros Entrenables:** 16,997,185
- **Parámetros No Entrenables:** 14,080

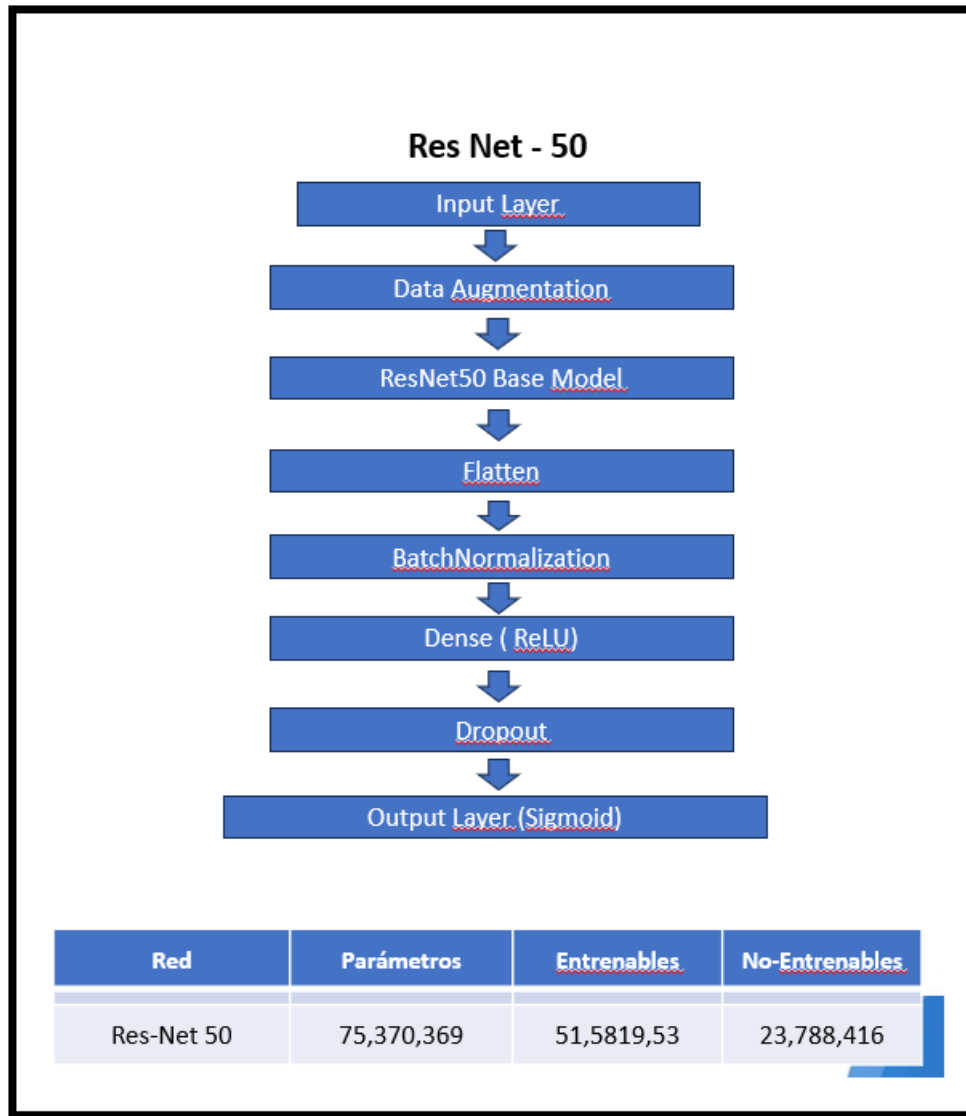


Figura N° 24 Modelo Resnet 50

El modelo ResNet50 se implementó teniendo en cuenta la estructura y robustez del modelo base ResNet50, optimizado para la detección de cáncer de seno en mamografías. La estructura del modelo es la siguiente:

1. **Capa de Entrada (Input Layer):** Recibe las imágenes de mamografías.
2. **Aumento de Datos (Data Augmentation):** Aplica técnicas de aumento de datos como rotaciones y traslaciones para incrementar la variabilidad del conjunto de datos y mejorar la capacidad de generalización del modelo.

3. **Modelo Base ResNet50 (ResNet50 Base Model):** Utiliza la arquitectura preentrenada ResNet50 para extraer características de las imágenes. La ResNet50 consta de 50 capas, incluyendo múltiples capas convolucionales y bloques residuales.
4. **Aplanamiento (Flatten):** Convierte los mapas de características multidimensionales en un vector unidimensional.
5. **Normalización por Lotes (BatchNormalization):** Normaliza las activaciones de las capas previas para estabilizar y acelerar el proceso de entrenamiento.
6. **Capa Densa (Dense) con Activación ReLU:** Introduce una capa totalmente conectada con función de activación ReLU para combinar las características extraídas.
7. **Dropout:** Aplica la técnica de dropout para evitar el sobreajuste, apagando aleatoriamente un porcentaje de las neuronas durante el entrenamiento.
8. **Capa de Salida (Output Layer) con Activación Sigmoid:** Proporciona la salida final del modelo. Dado que este es un problema de clasificación binaria (detección de cáncer de seno), se usa la función de activación Sigmoid para producir una probabilidad entre 0 y 1.

Configuración Adicional:

- **Scheduler de Tasa de Aprendizaje (LR Scheduler):** Sí, se utiliza un scheduler para ajustar dinámicamente la tasa de aprendizaje durante el entrenamiento.
- **Capas Densas (Dense Layers):** Se utiliza una capa densa de 512 neuronas.
- **Unfreezing Layers:** No se especifican capas descongeladas.
- **Fine Tuning:** No se realiza afinamiento fino (fine-tuning).

Parámetros del Modelo:

- **Total de Parámetros:** 75,370,369
- **Parámetros Entrenables:** 51,581,953
- **Parámetros No Entrenables:** 23,788,416

El uso del modelo ResNet-50 en la detección de cáncer de seno a partir de mamografías proporciona una estructura robusta y preentrenada que, junto con las técnicas de aumento de datos y normalización, puede ofrecer un desempeño efectivo en la clasificación binaria. Las configuraciones específicas de parámetros entrenables y no entrenables permiten ajustar el modelo para optimizar su rendimiento y reducir el riesgo de sobreajuste.

4.3.5. Prueba Modelo VTI

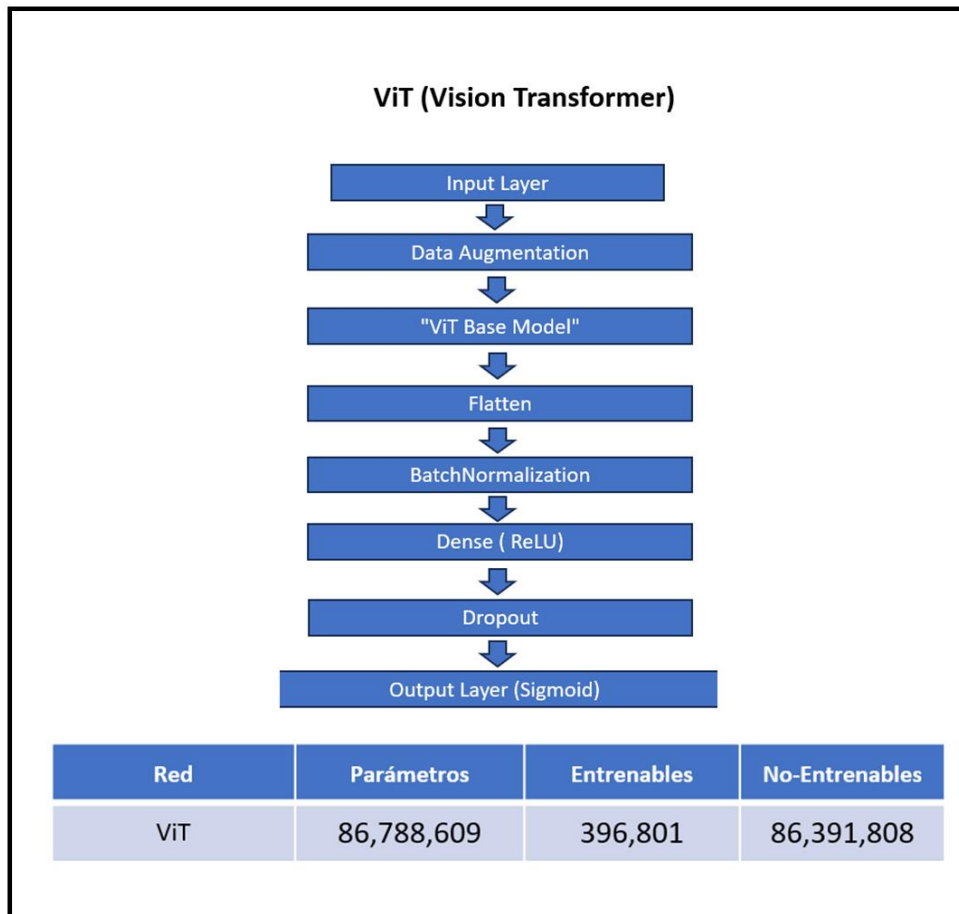


Figura N° 25 Modelo ViT

El modelo ViT (Vision Transformer) fue aplicado en el proyecto fue descargado de Hugging Face, se basa en una arquitectura de transformadores originalmente diseñada para el procesamiento de lenguaje natural, adaptada para el análisis de imágenes. El ViT divide cada imagen en pequeños parches y trata cada parche permitiendo así capturar relaciones complejas dentro de la imagen .

Primero, se carga el modelo ViT preentrenado y se congela su entrenamiento para mantener sus pesos fijos con 512 neuronas . Luego, se definen las capas de entrada para los valores de píxeles de las imágenes, seguidas de capas de aumento de datos para mejorar la generalización del modelo. Se aplica el modelo base ViT a las imágenes de entrada, seguido de una serie de capas para procesar

la salida del modelo base. Estas capas incluyen aplanamiento, normalización por lotes, una capa densa con regularización L2, normalización por lotes adicional y una capa de abandono para prevenir el sobreajuste. Finalmente, se construye y compila el modelo completo, especificando el optimizador, la función de pérdida (entropía cruzada binaria) y las métricas (precisión binaria).

El modelo Visión por Transformadores (ViT) preentrenado para la clasificación de imágenes y lo adapta para una tarea específica. Primero, se carga el modelo ViT preentrenado y se congela su entrenamiento para mantener sus pesos fijos. Luego, se definen las capas de entrada para los valores de píxeles de las imágenes, seguidas de capas de aumento de datos para mejorar la generalización del modelo. Se aplica el modelo base ViT a las imágenes de entrada, seguido de una serie de capas para procesar la salida del modelo base.

Estas capas incluyen aplanamiento, normalización por lotes, una capa densa con regularización L2, normalización por lotes adicional y una capa de abandono para prevenir el sobreajuste. Finalmente, se construye y compila el modelo completo, especificando el optimizador, la función de pérdida (entropía cruzada binaria) y las métricas (precisión binaria).

4.4. Tablas de Resultados de las arquitecturas evaluadas.

Mode	LR Scheduler	Dense layers	Unfreezing layers	Fine Tuning	Accuracy Train	Accuracy Val	Label	Recall 1/0	Precisión 1/0	f1-score 1/0
VGG16 V1	No	256	5	No	0.7585	0.6995				
				Yes	0.7941	0.6899	1 0	0.78, 0.62	0.63, 0.76	0.70, 0.68
VGG16 V2	No	256	10	No	0.7445	0.6980				
				Yes	0.8028	0.6804	1 0	0.71 0.66	0.64 0.73	0.67 0.69
VGG16 V3	No	512	-	No	0.7747	0.6816				
				Yes	0.7747	0.6816	1 0	0.78 0.58	0.61 0.76	0.68 0.66
VGG19 V1	No	128	5	No	0.7169	0.6887				
				Yes	0.7696	0.6896	1 0	0.76 0.63	0.63 0.76	0.69 0.69
VGG19 V2	No	256	10	No	0.7307	0.6884				
				Yes	0.7828	0.6742	1 0	0.84 0.53	0.60 0.80	0.70 0.64
VGG19 V3	No	512	10	No	0.7539	0.6862				
				Yes	0.8108	0.6813	1 0	0.77 0.61	0.62 0.76	0.67 0.69
Rest Net	Yes	1024	-	No	0.8084	0.6504				
				Yes	0.8084	0.6504	1 0	0.75, 0.60	0.65 0.70	0.69 0.65
Rest Net 50	Yes	512	-	No	0.7356	0.6025				
				Yes	0.7356	0.6025	1 0	0.81 0.42	0.54 0.73	0.65 0.53
ViT	Yes	512	-	No	0.8297	0.7254				
				Yes	0.8297	0.7254	1 0	0.71 0.74	0.70 0.75	0.70 0.75

Figura N° 26 Tabla de Resultados de las arquitecturas evaluadas- creación Propia.

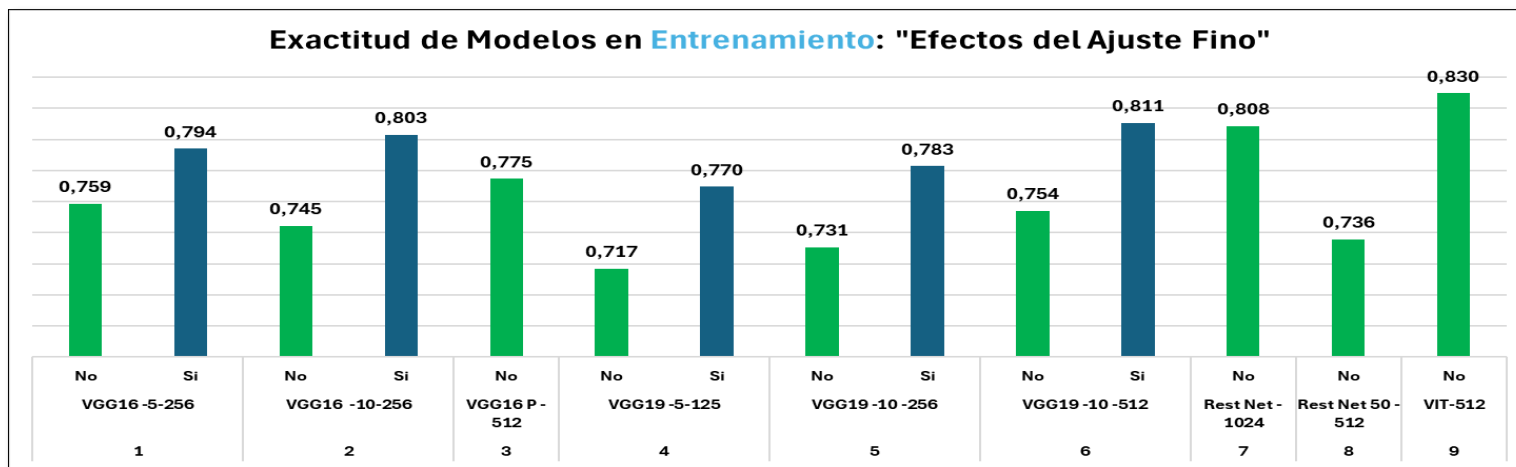


Figura N° 27 Tabla de resultados datos de entrenamiento con o sin ajuste fino .

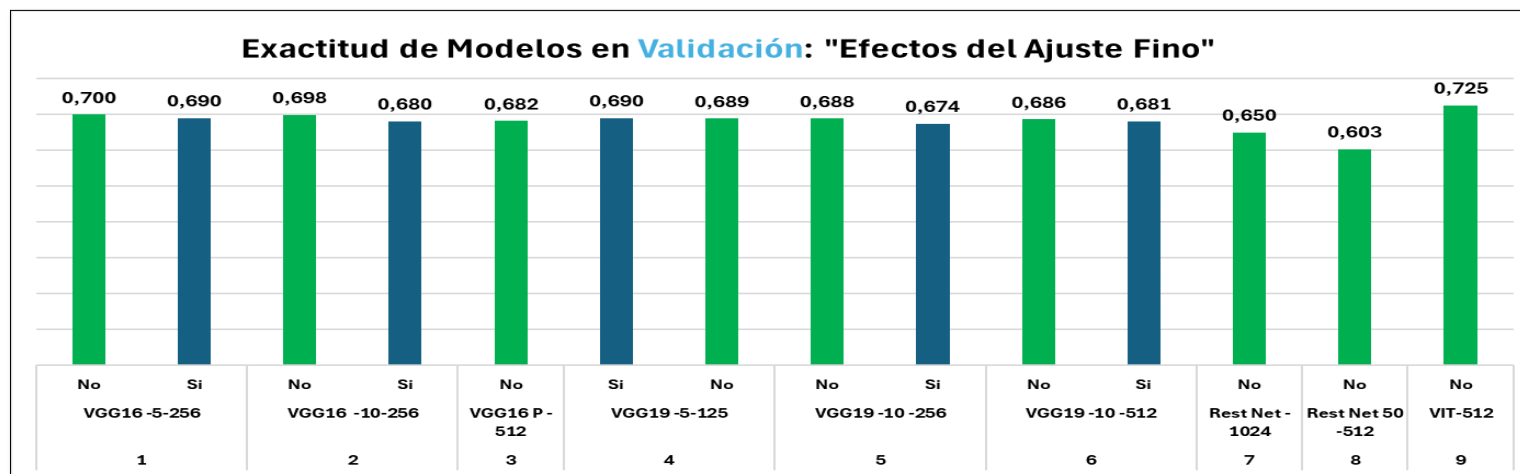


Figura N° 28 Tabla de resultados datos de Validación con o sin ajuste fino.

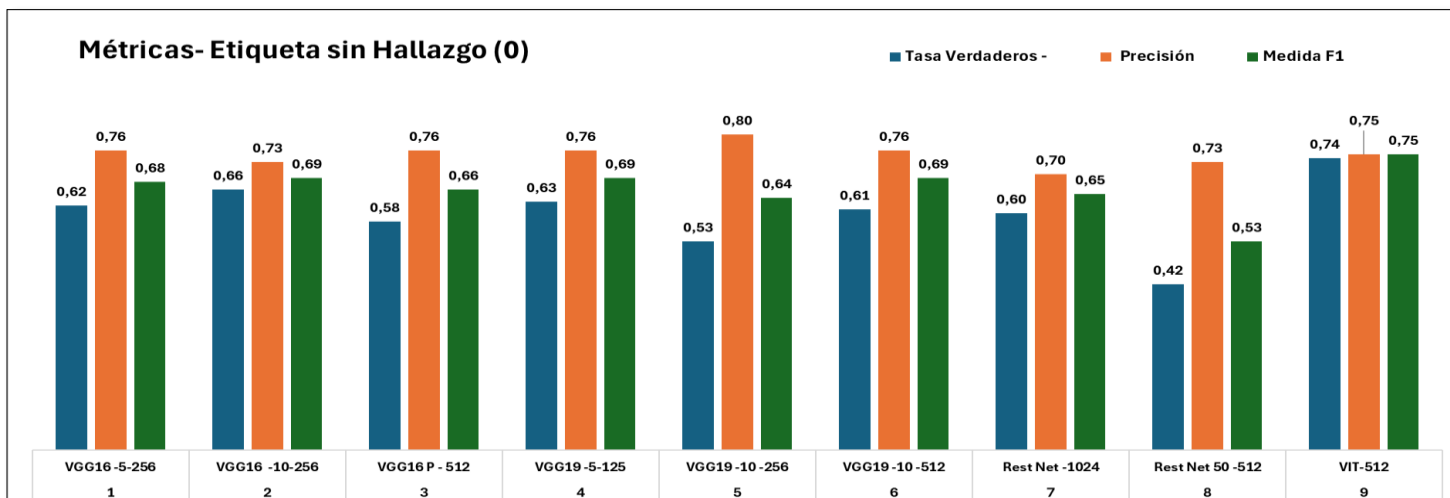


Figura N° 29 Tabla de métricas de evaluación de datos con etiqueta sin Hallazgo .

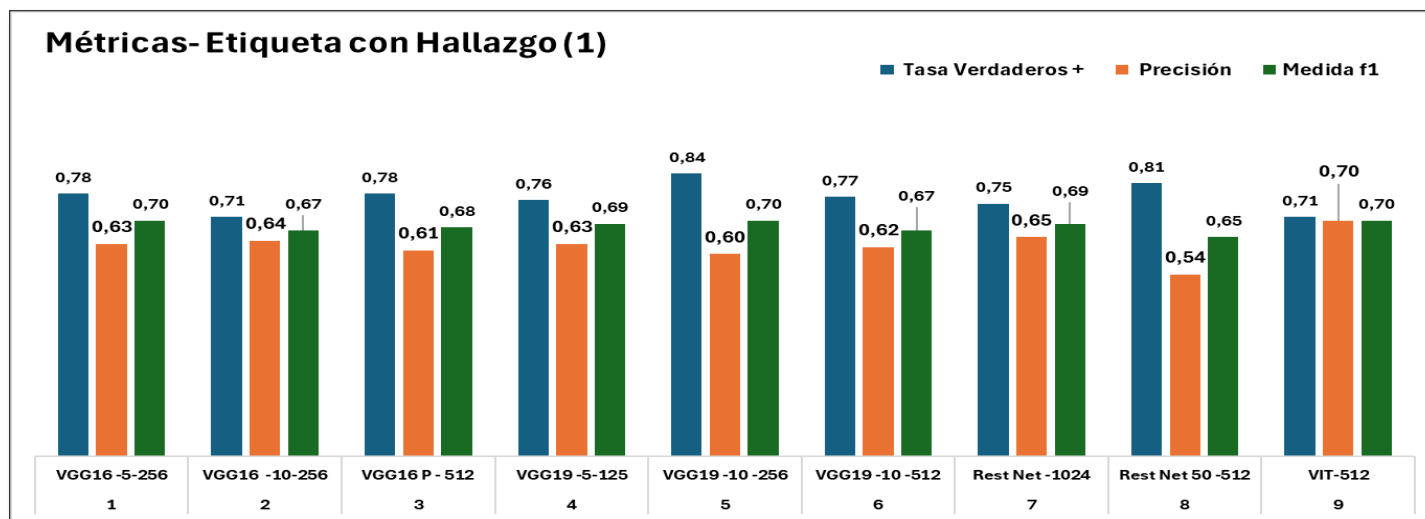
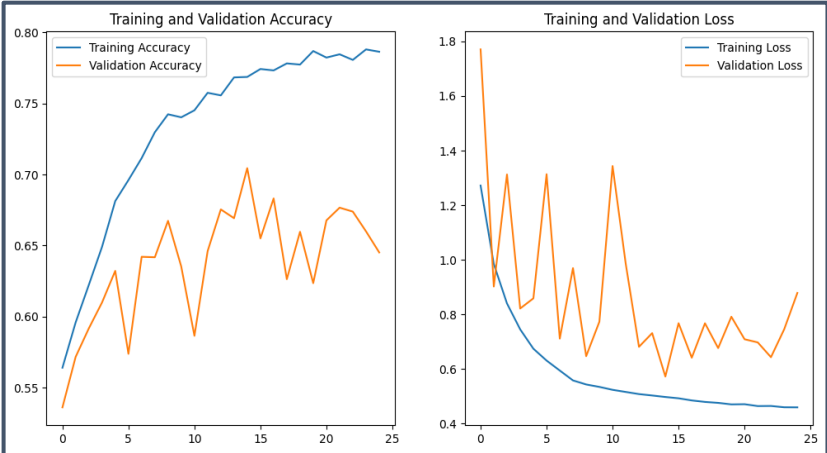


Figura N° 30 Tabla de métricas de evaluación de datos con etiqueta Con Halla

4.4.1. Tabla de Resultados gráficos de las arquitecturas evaluadas.

Modelos	Graficas de Entrenamiento
<p data-bbox="181 415 594 445">VGG16 C</p> <p data-bbox="181 453 594 1297">La exactitud de entrenamiento aumenta constantemente a medida que avanza el entrenamiento, alcanzando un valor máximo de aproximadamente 0.78, sin embargo, la exactitud y la pérdida de validación está oscilando de manera errática y muestra una mala generalización de los datos. Hay un claro sobreajuste, como lo indican la gran brecha entre la precisión de entrenamiento y la de validación, así como la alta variabilidad en la pérdida de validación. El modelo parece estar memorizando los datos de entrenamiento en lugar de aprender características generales que se aplican a nuevos datos.</p> <p data-bbox="181 1348 594 1407">Tiempo de ejecución :1 hora y 30 minutos</p>	 <p>The figure contains two line graphs side-by-side. The left graph, titled 'Training and Validation Accuracy', plots accuracy from 0 to 25. The blue line (Training Accuracy) starts at approximately 0.57 and rises steadily to about 0.78. The orange line (Validation Accuracy) starts at approximately 0.53 and fluctuates between 0.58 and 0.71. The right graph, titled 'Training and Validation Loss', plots loss from 0 to 25. The blue line (Training Loss) starts at approximately 1.3 and decreases to about 0.45. The orange line (Validation Loss) starts at approximately 1.75 and fluctuates between 0.6 and 1.35.</p>

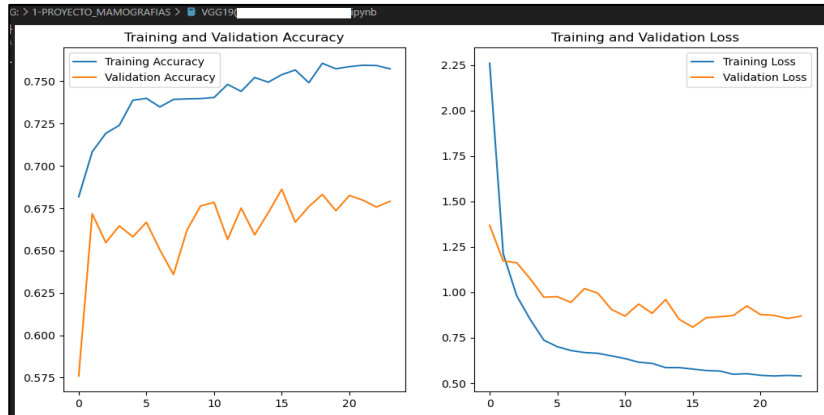
VGG19 -UF10 v2

Estas gráficas de entrenamiento muestran una tendencia creciente en la curva de entrenamiento, lo que indica que el modelo está aprendiendo. Igualmente, se puede observar que la curva de validación también aumenta, aunque con una pendiente menor.

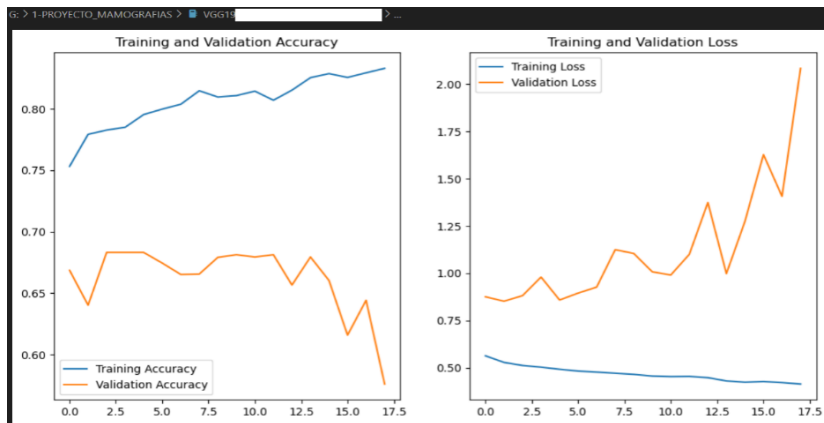
Sin embargo, en la etapa de fine-tuning, observamos cómo las gráficas comienzan a divergir drásticamente, mostrando un evidente sobreajuste (overfitting) del modelo y una notable pérdida en la exactitud del conjunto de evaluación.

Tiempo de ejecución: 40 minutos

Modelos



fine-tuning



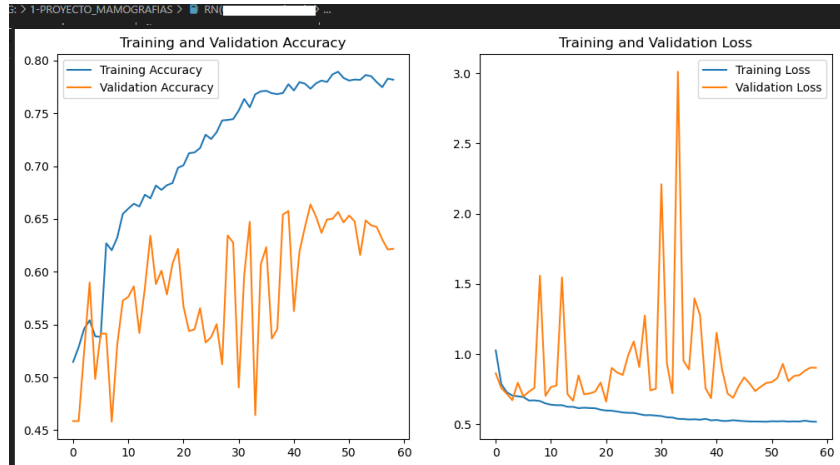
Graficas de Entrenamiento

RESNET:

La precisión de entrenamiento (línea azul) muestra un incremento constante y se estabiliza alrededor del 80% después de unos 50 epochs. Esto indica que el modelo está aprendiendo bien de los datos de entrenamiento.

Sin embargo, la precisión de validación presenta grandes fluctuaciones, fluctúa alrededor del 60% con muchos picos y valles a lo largo de los epochs, lo que indica la necesidad de realizar cambios en los hiperparámetros para mejorar el rendimiento en el conjunto de validación.

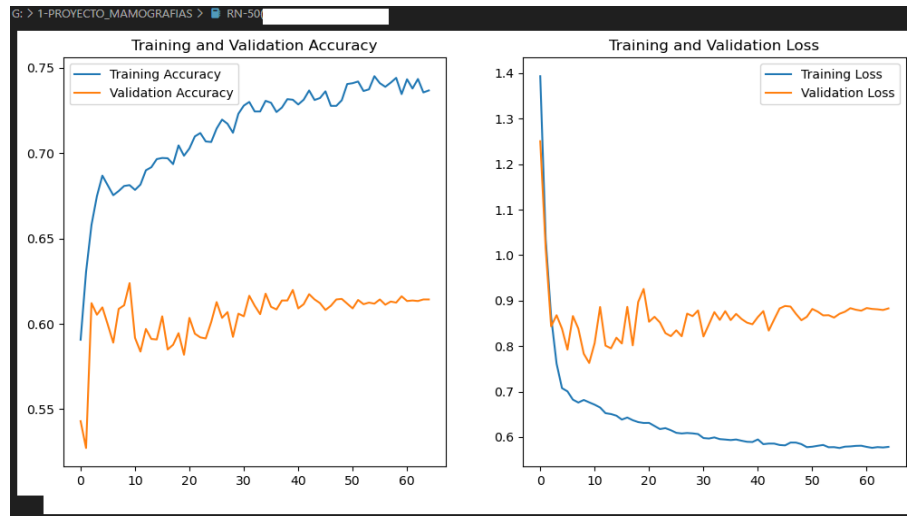
Tiempo de ejecución: 43 minutos



RESNET 50:

La precisión de entrenamiento aumenta gradualmente, mientras que la precisión de validación fluctúa sin una tendencia clara. De manera similar, las curvas de pérdida muestran que la pérdida de validación no presenta una tendencia definida, mientras que la pérdida de entrenamiento sigue disminuyendo. Hay un cierto grado de sobreajuste, como lo indica la brecha entre la precisión de entrenamiento y la de validación, así como la diferencia en las curvas de pérdida.

Tiempo de ejecución: 3 horas y 43 minutos



Modelos

Gráficas de Entrenamiento

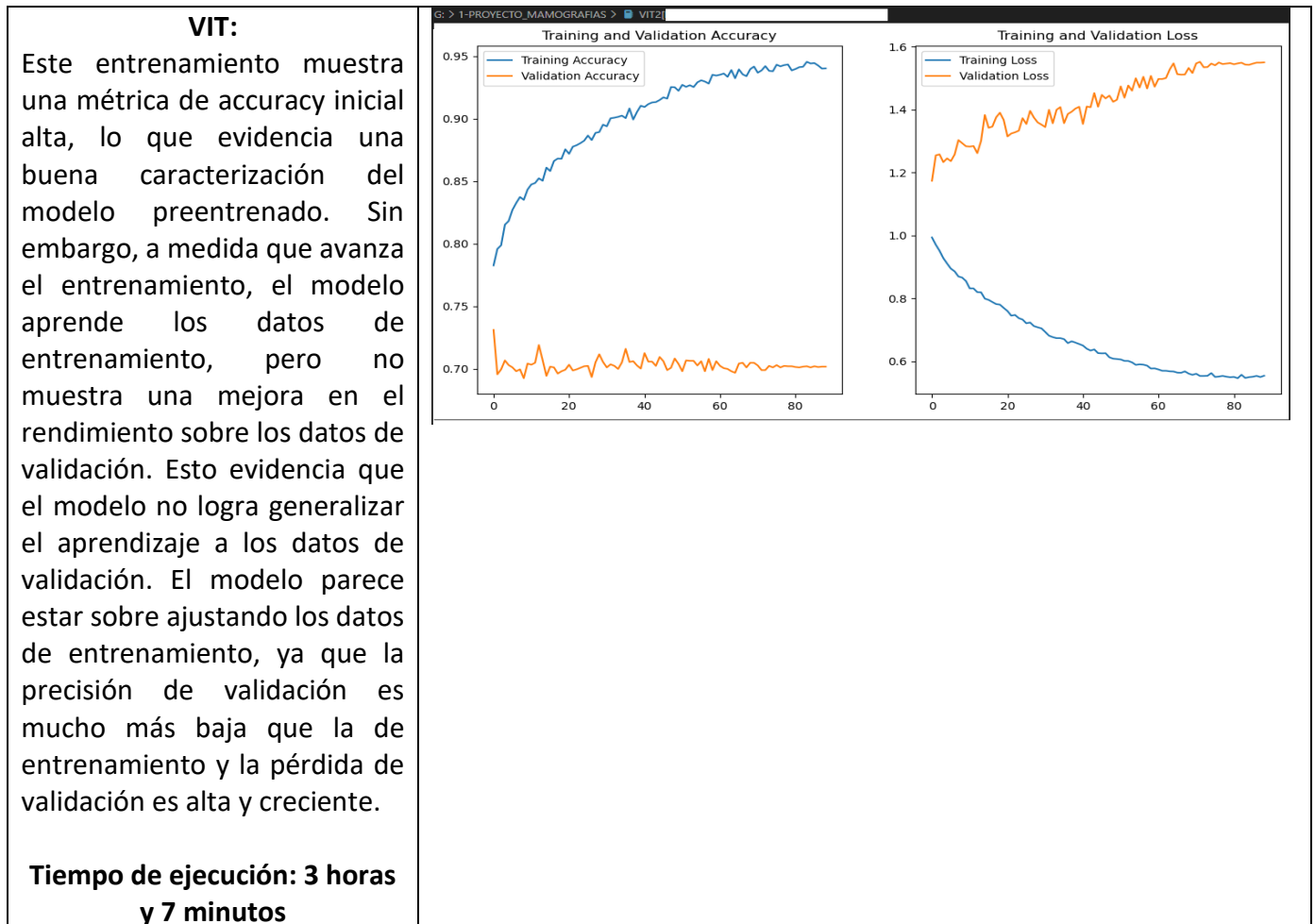


Figura N° 31 Tabla Análisis Grafico Resumen de las arquitecturas

4.4.2. Análisis Comparativo

A continuación analizamos las principales métricas de evaluación de rendimiento para cada uno de los modelos entrenados y su análisis comparativo entre sí.

Exactitud (Accuracy)

- VIT muestra la mejor exactitud en el conjunto de entrenamiento (0.8297) y validación (0.7254).
- VGG16 UL 5 también presenta una buena precisión en el entrenamiento (0.7941) y validación (0.6899).
- VGG19 UF5 muestra una exactitud decente tanto en el entrenamiento (0.7696) como en la validación (0.6896).

- ResNet PR y ResNet-50 tienen menor precisión en la validación (0.6504 y 0.6025 respectivamente).

Recall

- VGG19 UF5 muestra un buen equilibrio en recall (1: 0.76, 0: 0.63).
- VGG19 -UF10 v1 tiene un recall muy alto para la clase 1 (0.84) pero bajo para la clase 0 (0.53).
- ResNet-50 presenta el mejor recall para la clase 0 (0.81), aunque es más bajo para la clase 1 (0.42).

Precisión

- VGG16 UL 5 y VGG16 - UL 10 mantienen una precisión equilibrada en clase 1 con (0.64) y (0.66) respectivamente.
- VGG19 -UF10 v1 tiene buena precisión para la clase 0 (0.80) pero menor para la clase 1 (0.60).

F1-score

- VIT tiene el F1-score más equilibrado (1: 0.70, 0: 0.75).
- VGG19 UF5 y VGG16 también tienen buenos F1-scores con 0.70 para la clase 1.
- VGG19 -UF10 v1 muestra un F1-score decente (1: 0.70, 0: 0.64).

4.4.3. Análisis del impacto de ajustes en el rendimiento de Modelos

A continuación, analizamos los principales ajustes realizados (incluyendo el tamaño del lote ("Batch Size"), el aprendizaje ("LR Scheduler"), la deserción ("Dropout"), la cantidad de capas densas ("Dense"), el descongelamiento de capas ("Unfreezing"), y el ajuste fino ("Fine"). para cada tipo de modelo y como estos influyen o no en los resultados de rendimiento para cada uno de los modelos entrenados.

El objetivo de este análisis es comprender cómo estos ajustes afectan el rendimiento de cada modelo en términos de precisión, recall, precisión y F1-score.

Análisis por Modelo

Modelo VGG16:

- **UL 5 vs. UL 10:**
 - Al comparar UL 5 con UL 10, se observa una mejora notable en la precisión (0.7585 vs. 0.7941) y en el F1-score (0.78 vs. 0.82) con el ajuste fino activado.
 - El recall y la precisión se mantienen relativamente consistentes.
 - Esto sugiere que aumentar el número de capas descongeladas (UL 10) permite al modelo extraer mejor las características relevantes de los datos, lo que conduce a predicciones más precisas.
- **UL 5 vs. C:**
 - Al comparar UL 5 con C, se observa una mejora significativa en la precisión (0.7585 vs. 0.7747) y en el F1-score (0.78 vs. 0.8) con el ajuste fino activado.
 - El recall y la precisión se mantienen consistentes.
 - Esto indica que la arquitectura de la red neuronal (C) es menos efectiva que la arquitectura base VGG16 (UL 5) para la detección de nódulos.

Modelo VGG19:

- **UF5 vs. UF10 v1:**
 - Al comparar UF5 con UF10 v1, se observa una mejora marginal en la precisión (0.7169 vs. 0.7307) y en el F1-score (0.7 vs. 0.71) con el ajuste fino activado.
 - El recall y la precisión se mantienen relativamente consistentes.
 - Esto sugiere que aumentar el número de capas descongeladas (UF10 v1) tiene un impacto menor en el rendimiento de este modelo.
- **UF5 vs. UF10 v2:**
 - Al comparar UF5 con UF10 v2, se observa una mejora marginal en la precisión (0.7169 vs. 0.7539) y en el F1-score (0.7 vs. 0.71) con el ajuste fino activado.
 - El recall y la precisión se mantienen relativamente consistentes.
 - Similar al caso anterior, aumentar el número de capas descongeladas (UF10 v2) tiene un impacto menor en el rendimiento de este modelo.

Modelo Rest Net:

- **PR vs. 50:**
 - Al comparar PR con 50, se observa una leve mejora en la precisión (0.8084 vs. 0.7356) y en el F1-score (0.75 vs. 0.69) con el ajuste fino activado.
 - El recall y la precisión se mantienen consistentes.
 - Esto sugiere que la arquitectura ResNet (PR) tiene un rendimiento comparable a la arquitectura VGG16 (50) para la detección del cáncer de mama.

4.4.4. Interpretación y Recomendaciones

- VIT sigue siendo el modelo más prometedor, con los mejores valores de precisión y F1-score, lo que indica un buen equilibrio entre precisión y recall.
- Los resultados sugieren que el ajuste fino tiene un impacto positivo en el rendimiento de la mayoría de los modelos evaluados, especialmente en términos de precisión y F1-score.
- En el caso del modelo VGG16, aumentar el número de capas descongeladas (UL 10) también contribuye a mejorar el rendimiento.
- El modelo ResNet (PR) presenta un rendimiento comparable al modelo VGG16 (50).
- RN y RN-50 muestran buen rendimiento en precisión durante el entrenamiento, pero tienen margen de mejora en validación. Ajustar el learning rate y explorar más técnicas de regularización podría ayudar.
- Continuar ajustando hiperparámetros, como el learning rate y el dropout, y explorar diferentes configuraciones de capas densas y regularizadores de kernel para optimizar aún más el rendimiento de los modelos.
- Basándonos en los resultados obtenidos en el estudio publicado por Quiñones [28], donde la red YOLO fue entrenada y validada utilizando la misma base de datos que nosotros, Vindr-Mammo, realizamos una comparación exhaustiva.

En dicho estudio, el mejor modelo se obtuvo con YOLOv5x, alcanzando una sensibilidad máxima del 80% en la validación interna y del 72% en una prueba externa con la base de datos de "The Mammographic Image Analysis Society" (MIAS).

- En contraste, nuestros experimentos, también basados en la base de datos Vindr-Mammo, si bien no contaron con pruebas externas ni la misma capacidad computacional que el estudio de Quiñones, demostraron resultados comparables, si bien no se realizaron pruebas externas, estos resultados sugieren que YOLOv5x, particularmente en su versión más avanzada, representa una opción prometedora para futuras investigaciones en la detección de cáncer de seno utilizando la base de datos Vindr-Mammo.

- En general podemos concluir que, a pesar de las limitaciones de nuestro estudio, las redes neuronales evaluadas brindan información valiosa sobre el comportamiento de estas arquitecturas con esta base de datos, lo que puede servir como base para futuras investigaciones con mayor capacidad computacional y pruebas externas.

4.1. Resumen de la Aplicación y Desempeño de Modelos de Aprendizaje Profundo

Esta tabla proporciona una visión general de la aplicación y desempeño de diversos modelos de aprendizaje profundo utilizados en el proyecto, destacando las observaciones clave, los beneficios y los desafíos asociados con cada modelo.

Modelo	Interpretación	Recomendaciones
VIT (Visión Transformer)	VIT muestra la mayor exactitud tanto en el entrenamiento como en la validación, lo que sugiere que la arquitectura Transformer es muy efectiva para la tarea de clasificación de mamografías. Su equilibrio en el F1-score también destaca su capacidad para manejar ambas clases de manera eficiente.	Se debería considerar el uso de VIT para aplicaciones donde la precisión y el balance entre clases son críticos. Además, investigar más a fondo el ajuste fino de este modelo podría potencialmente mejorar aún más su rendimiento. El entrenamiento y la inferencia con modelos VIT requieren un manejo cuidadoso de los recursos de GPU y RAM debido a sus altos requerimientos de memoria y capacidad de procesamiento. Implementar estrategias de optimización y ajustar el tamaño del modelo son pasos cruciales para abordar estos desafíos y aprovechar al máximo el poder de VIT.
VGG16 y VGG19	Ambos modelos ofrecen un buen rendimiento general, con VGG19 mostrando un mejor recall para la clase 1 en ciertas configuraciones. Los ajustes finos y la variación en los hiperparámetros resultan en	VGG16 y VGG19 son opciones sólidas para la clasificación de mamografías. Se recomienda seguir optimizando estos modelos, especialmente explorando diferentes configuraciones de hiperparámetros y técnicas de aumento de datos para maximizar su potencial.

Modelo	Interpretación	Recomendaciones
	mejoras notables, aunque moderadas.	
ResNet y ResNet-50	Los modelos ResNet tienen un rendimiento decente, pero no sobresalen en comparación con VGG y VIT. La falta de ajuste fino puede haber limitado su rendimiento.	Explorar el ajuste fino y la optimización de hiperparámetros para los modelos ResNet podría mejorar su rendimiento. También, evaluar variantes de ResNet podría ofrecer mejores resultados.

Figura N° 32 Tabla de resultados de la aplicación de los modelos- creación Propia

5. CONCLUSIONES Y TRABAJOS FUTUROS

5.1. CONCLUSIONES

En términos de las arquitecturas evaluadas, los modelos VGG19, VGG16 y VIT han mostrado resultados prometedores en cuanto a precisión en el conjunto de validación. La precisión y las métricas de recuperación (recall) son fundamentales en la detección de casos de cáncer de seno, y estos modelos exhiben valores razonables en estas métricas. Además, el ajuste fino parece ser una estrategia eficaz para mejorar el rendimiento de los modelos, y los hiperparámetros como la tasa de aprendizaje, el tamaño del lote (batch size) y la regularización desempeñan un papel crucial en su desempeño. La exploración de diversas configuraciones de hiperparámetros sugiere un enfoque sistemático para encontrar la configuración óptima que maximice la eficacia de los modelos en la tarea de detección de cáncer de seno.

En el tiempo disponible para el proyecto, se observó la importancia de realizar un análisis detallado de los modelos para comprender mejor los factores que influyen en su rendimiento, como la elección de la arquitectura y los hiperparámetros adecuados. Este proceso resultó ser uno de los más desafiantes, pero también ofreció valiosas lecciones que podrán aplicarse en futuras investigaciones y proyectos similares.

Además, se reconoce la necesidad de llevar a cabo una búsqueda sistemática de hiperparámetros para encontrar la combinación óptima que maximice el rendimiento del modelo en la detección de cáncer de seno. Aunque se adaptaron los plazos y los cronogramas para ajustarse a las circunstancias, se reconoce que una mayor disponibilidad de recursos podría haber permitido una exploración más exhaustiva y, potencialmente, mejores resultados.

Es importante destacar que, durante el desarrollo del proyecto, se enfrentaron desafíos significativos debido a las limitaciones de recursos. A pesar de contar únicamente con recursos propios de máquina y la necesidad de pagar por un espacio adicional de trabajo en Google Colab, se logró avanzar en la investigación. Sin embargo, esta limitación pudo haber influido en la capacidad para obtener resultados aún más sólidos y precisos.

En resumen, si bien se lograron avances significativos en la investigación, queda claro que hay margen para mejoras adicionales. Se sugiere seguir investigando y explorando nuevas arquitecturas, técnicas y ajustes de hiperparámetros, con el objetivo de maximizar la eficacia del sistema en la detección de cáncer de seno y contribuir así a la salud pública.

5.2. TRABAJOS FUTUROS

Es vital seguir avanzando en la investigación y mejora del sistema mediante la exploración de nuevas arquitecturas, técnicas de aprendizaje profundo y conjuntos de datos más amplios y diversos. Esta búsqueda constante nos permitirá mantenernos a la vanguardia de los avances en el campo y asegurarnos de que nuestro sistema esté aprovechando plenamente las últimas innovaciones para mejorar la detección de cáncer de seno.

Además, es esencial que la teoría se complemente con la práctica. Por lo tanto, sugerimos encarecidamente que en futuras investigaciones se realice una evaluación clínica exhaustiva de los modelos mejorados. Esta evaluación en entornos del mundo real será fundamental para validar la eficacia de los modelos y su capacidad para contribuir de manera efectiva a la detección temprana y precisa de casos de cáncer de seno.

Asimismo, proponemos el desarrollo de una interfaz de usuario intuitiva para el sistema de detección de nódulos de cáncer de seno. Esto facilitará su uso práctico por parte de profesionales médicos en entornos clínicos. Creemos firmemente que apoyar y facilitar el acceso a desarrollos tecnológicos como estos es crucial para lograr un impacto significativo en la salud pública de las personas. La implementación práctica en entornos clínicos es fundamental para llevar estas innovaciones desde el laboratorio hasta la práctica clínica y hacer una diferencia tangible en la vida de los pacientes.

6 REFERENCIAS BIBLIOGRÁFICAS

- [1] American Cancer Society. “Breast Cancer Facts & Figures 2021-2022.” Breast Cancer Facts & Figures 2022-2024
- [2] c. d. a. c. colombia, «cuenta de alto costo colombia.org,» 20 01 2023. [En línea]. Available: <https://cuentadealtocosto.org/cancer/dia-mundial-del-cancer-2023/>. [Último acceso: 15 05 2023].
- [3] A. C.-R. M. H. M.E MEJIA-ROJAS, «BIOMEDICA,» INSTITUTO NACIONAL DE SALUD , 15 06 2020. [En línea]. Available: <https://revistabiomedica.org/index.php/biomedica/article/view/4971>.
- [4] NTELIGENCIA ARTIFICIAL EN SALUD,» UAM ESPAÑA, 20 06 2022. [En línea]. Available: <https://www.iic.uam.es/lasalud/realidad-inteligencia-artificial-salud/>
- [5] A. C. Society, «American Cancer Society,» 15 10 2019. [En línea]. Available: <https://www.cancer.org/es/cancer/tipos/cancer-de-seno/pruebas-de-deteccion-y-deteccion-temprana-del-cancer-de-seno/>
- [6] N. H. C. ORG, «NPH HEALTH CARE ORG Medical Services. (2015, Nov. 9). ¿Qué es el cáncer de mama? [Online]. Available: <https://medicalservices.nph.org/2015/11/09/que-es-el-cancer-de-mama/>
- [7] H. Demuth and B. De Jesús, Neural Network Design 2nd Edition. Accessed: Jun. 03, 2023. [Online]. Available: <https://hagan.okstate.edu/NNDesign.pdf>
- [8] J. L. Sarmiento-Ramos, “Aplicaciones de las redes neuronales y el deep learning a la ingeniería biomédica,” Revista UIS Ingenierías, vol. 19, no. 4, pp. 1–18, Jun. 2020, doi: 10.18273/revuin.v19n4- 2020001.
- [9] C. Alberto Ruiz Marta Susana Basualdo Autor and D. Jorge Matich, “Cátedra: Informática Aplicada a la Ingeniería de Procesos-Orientación I Redes Neuronales: Conceptos Básicos y Aplicaciones.”
- [10] J. Pérez Guerrero Sevilla, J. de, T. por, and R. Pino Mejías, “REDES RECURRENTE.”

- [11] Álvaro Artola Moreno, "Clasificación de imágenes usando redes neuronales convolucionales en Python".
- [12] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556v16-19.
- [13] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (pp. 770-778)R50.
- [14] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Hinton, G. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 vit.
- [15] A. Quiñones-Espín, M. Perez-Diaz, R. Espín-Coto, D. Rodriguez-Linares, and J. Lopez-Cabrera, "Automatic detection of breast masses using deep learning with YOLO approach," Health and Technology, vol. 13, no. 6, pp. 915-923, 2023.
- [16] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 807-814.
- [17] Cleverpy, "Como crear una red convolucional con PyTorch–Cleverpy," 2018 <https://cleverpy.com/2018/01/05/red-convolucional-pytorch/>.
- [18] National Institute of Biomedical Imaging and Bioengineering |, "Mamografía,2022" <https://www.nibib.nih.gov/sites/default/files/2022-05/Fact-Sheet-Mamograf%C3%ADa.pdf>.
- [19] Y. Wang et al., "Supervised breast cancer prediction using integrated dimensionality reduction CNN," PLOS ONE, vol. 16, no. 3, 2021. [Online]. Available: <https://doi.org/10.1371/journal.pone.0248572>. [Accessed: Jul. 15, 2024].
- [20] A. Author et al., "Breast cancer detection using artificial intelligence techniques," University of Sharjah, 2021. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/34008163/>. [Accessed: Jul. 15, 2024].

- [21] M. Rossi et al., "Convolutional Networks and Transformers for Mammography Classification," *Journal of Imaging*, vol. 9, no. 2, 2023. [Online]. Available: <https://doi.org/10.3390/jimaging9020080>. [Accessed: Jul. 15, 2024].
- [22] H. Tanaka et al., "Quality control system for mammographic breast positioning using deep learning," *Scientific Reports*, vol. 13, no. 1, 2023. [Online]. Available: <https://doi.org/10.1038/s41598-020-72663-w>. [Accessed: Jul. 15, 2024].
- [23] A. B. Nassif, M. A. Talib, Q. Nasir, Y. Afadar and O. Elgendy, "Breast cancer detection using artificial intelligence techniques: Universidad de Sharjah en los Emiratos Árabes Unidos 2021" <https://pubmed.ncbi.nlm.nih.gov/35430037/>
- [24] A. Kowalski et al., "Automatic detection of breast masses using deep learning with YOLO approach," Springer, 2023. [Online]. Available: <https://link.springer.com/article/10.1007/s12553-023-00783>. [Accessed: Jul. 15, 2024].
- [25] Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., Taylor, P., Betal, D., & Savage, J. (1994). The mammographic image analysis society digital mammogram database. En Suckling, J. & Parker, J. (Eds.), *MIAS Database*. 10_1. Base de datos MIAS de mamografías:
- [26] Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer Jr., P. (2000). The digital database for screening mammography. En Heath, M. & Bowyer, K. (Eds.), *DDSM Database*. 11_2. Base de datos digital para mamografías de detección (DDSM):
- [27] Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., & Cardoso, J. S. (2012). INbreast: Toward a full-field digital mammographic database. En Moreira, I. C. & Cardoso, J. S. (Eds.), *INbreast Database*. 3-3. Base de datos INbreast:
- [28] VinDr-Mammo -Nguyen, T. T., Nguyen, H. T., Nguyen, T. H., Nguyen, T. N., Nguyen, D. H., Tran, Q. H., ... & Pham, T. D. (2021). VinDr-Mammo: A large-scale benchmark for mammogram analysis and diagnosis. medRxiv 4. base CBIS-DDSM: .
- [29] VinDr, "VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography", 2022. [Online]. Available: 1. [Accessed: 03-Oct-2023]. 5. VinDr-Mammo
- [30] H. T. Nguyen, H. Q. Nguyen, H. H. Pham, K. Lam, L. T. Le, M. Dao, V. Vu, "VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography," medRxiv, 2022.
- [31] Zhai, Xiaohua and Unterthiner, Thomas and Dehghani, Mostafa and Minderer, Matthias and Heigold, Georg and Gelly, Sylvain and others