

Sistema de Apoyo para Entrenamiento en Radiología Mamaria

Ricardo Antonio Salas Letelier

Universidad Técnica Federico Santa María

Av. Vicuña Mackenna 3939, 8940897 San Joaquín, Región Metropolitana

`ricardo.salas@usm.cl`

Resumen: Actualmente, la mamografía es el estándar más efectivo para la detección precoz del cáncer de mama, que exige a los radiólogos un alto grado de especialización, destreza y rigor analítico. La especialización en radiología mamaria tradicionalmente se desarrolla bajo una estructura de maestro-aprendiz. Este tipo de estructura, aunque efectiva en la transferencia de destrezas clínicas, presenta limitaciones sistémicas en el entorno sanitario. Para mitigar las limitaciones identificadas, se propone el diseño de un sistema de apoyo al entrenamiento en radiología mamaria que utilice técnicas de inteligencia artificial para guiar de forma autónoma el aprendizaje de los médicos residentes. La estrategia seguida en este trabajo corresponde al uso de redes neuronales convolucionales (CNN) con y sin entrenamiento. La metodología implementada consiste en: a) preprocesamiento de imágenes, b) evaluación del desempeño del modelo/dataset, c) capacidad de predicción y d) recomendación a partir de los modelos analizados. Dentro de los resultados obtenidos se destaca la relevancia del preprocesamiento de imágenes para detección de densidad/hallazgos. Se logró observar un rendimiento dispar entre los modelos, donde DenseNet, presentó el rendimiento más robusto y consistente, alcanzando un accuracy de 0.65. Le sigue modelo SwinV2, cuya capacidad para clasificar la densidad mamaria resultó insuficiente. Tanto ResNet50 como VGG16, modelos sin reentrenar, no lograron detectar eficazmente los hallazgos de Mass y Suspicious Calcification, respectivamente.

Palabras Clave: radiología, mamografía, búsqueda por similitud, sistema de recomendación.



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☐ Memoria o trabajo de título ☒ Tesis de Postgrado

Título del trabajo: SISTEMA DE APOYO PARA ENTRENAMIENTO EN RADIOLOGIA ULTRASONIA

Nombre del candidato(a): Ricardo Antonio Salas Letelier

Carrera / Grado: Magister en Tecnologías de la Información

Campus: San Joaquín Departamento: Informática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Julio Sotelo Parraguez, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

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

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

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

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un período de confidencialidad (**embargo**) por (marcar una opción):

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

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

4.- FIRMAS

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

Fecha: 07/08/2026

Firma: 

Estudiante o Candidato(a):

Fecha: 07-08-2026

Firma: 

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

1 Introducción

1.1 Contexto, motivación y problemática

En el ámbito de las ciencias de la salud, la formación de especialistas posee un componente pragmático fundamental basado en el modelo tradicional de maestro y aprendiz. En este entorno (típicamente hospitales y centros clínicos), la transferencia de conocimiento se produce mediante la resolución supervisada de casos prácticos. La especialidad de Radiología se rige bajo este paradigma de aprendizaje experiencial [1] [2] [3].

En el contexto epidemiológico de Chile, el cáncer de mama constituye la principal causa de mortalidad femenina. Dada la complejidad de reducir su incidencia mediante prevención primaria, los esfuerzos sanitarios se orientan hacia la disminución de la mortalidad a través del diagnóstico precoz, etapa en la que las intervenciones clínicas presentan una mayor eficacia [4]. Para ello, es imperativo realizar estudios imagenológicos periódicos en mujeres asintomáticas a partir de los 40 años [5] [6]. Actualmente, la mamografía es el estándar de detección más efectivo para identificar la patología en sus estadios iniciales [7]. No obstante, la interpretación visual de estos estudios es una labor crítica que exige un alto grado de especialización, destreza y rigor analítico.

En consecuencia, el análisis de mamografías requiere de radiólogos con un entrenamiento avanzado. Sin embargo, se observa que este modelo formativo genera una dependencia directa de la disponibilidad horaria del especialista o maestro, un recurso sumamente escaso en la práctica radiológica actual. En este escenario, el presente trabajo propone un sistema de recomendación de imágenes de mamografía digital para el entrenamiento radiológico a través de búsqueda por similitud de mamografías, teniendo como estrategia el uso de redes neuronales convolucionales (CNN) con y sin entrenamiento [8] [9] [10] [11].

1.2 Definición del problema

El proceso de especialización en radiología mamaria enfrenta actualmente un problema crítico debido a la naturaleza de su modelo de enseñanza. El aprendizaje basado en la relación maestro-aprendiz, aunque efectivo en la transferencia de destrezas clínicas, presenta limitaciones estructurales en el entorno sanitario [12] [7].

- (a) **Dependencia de la Supervisión.** La formación del residente depende estrictamente de la disponibilidad y el tiempo del radiólogo especialista. Dada la alta carga asistencial en los centros de salud, el tiempo dedicado a la tutoría académica se ve frecuentemente comprometido, lo que ralentiza el ciclo de aprendizaje del estudiante y limita la cantidad de casos que puede analizar bajo supervisión experta.
- (b) **Subjetividad y Curva de Aprendizaje.** La interpretación de mamografías es una tarea visual de alta complejidad que conlleva un grado inherente de subjetividad. El entrenamiento requiere la exposición a una vasta y diversa casuística para desarrollar la agudeza necesaria. Sin una herramienta que proporcione retroalimentación inmediata y objetiva, el residente depende exclusivamente de la validación diferida del maestro, lo que puede consolidar sesgos o errores interpretativos durante las fases iniciales del entrenamiento.
- (c) **Necesidad de Estandarización.** Se identifica la falta de sistemas tecnológicos de apoyo que actúen como un facilitador del entrenamiento autónomo. Es necesario un mecanismo que permita al aprendiz (residente) enfrentarse a casos complejos de manera sistemática, recibiendo una guía que emule el criterio del especialista sin requerir su presencia física constante.

1.3 Estrategia de solución adoptada

Para mitigar las limitaciones identificadas en el modelo de enseñanza tradicional, se propone el diseño del Sistema de Apoyo para Entrenamiento en Radiología de Mamas, cuya columna vertebral será un sistema de recomendación a través de búsqueda por similitud de mamografías.

El diseño de nuestro sistema se fundamenta en la integración de dos pilares tecnológicos:

- **Deep Learning.** Implementado para la recuperación de imágenes médicas basadas en contenido (CBIR – Content-Based Image Retrieval), permitiendo al usuario contrastar sus hallazgos con casos similares validados.

Se explora la aplicación de redes neuronales convolucionales (CNN) para la resolución de problemas de clasificación y recuperación en lo que respecta a las imágenes de mamas. La idea es usar las CNN para la extracción de vectores de características de las imágenes [10] [13]. Las CNN usadas se clasifican en: (i) Aquellas que se entrenaron en imágenes naturales (*ImageNet* [14]) y (ii) Aquellas que usan los pesos de otra CNN previamente entrenada [11]. De esta manera las CNN usadas no serán reentrenadas [15].

- **Sistemas de Recomendación.** Diseñados para personalizar la ruta de aprendizaje del residente, identificando y sugiriendo casos clínicos específicos que refuercen las áreas donde se detecten discrepancias diagnósticas.

A partir de los vectores de características obtenidos de las imágenes se pueden realizar búsquedas por similitud, de tal manera de proponer una o más imágenes que tengan un parecido más cercano [16].

En instancias donde la interpretación del residente no sea concordante con la "verdad fundamental" del sistema, la plataforma instará o recomendará al usuario a reevaluar nuevas imágenes que posean características de una imagen de referencia o query, promoviendo un proceso de reflexión crítica y aprendizaje correctivo inmediato.

1.4 Objetivos

Con el propósito de materializar la solución propuesta, se definen los siguientes objetivos.

Objetivo General

Diseñar un sistema de apoyo al entrenamiento en radiología mamaria que utilice técnicas de inteligencia artificial para guiar de forma autónoma el aprendizaje de los médicos residentes.

Objetivos Específicos

- Implementar un método de recuperación de imágenes mamográficas basado en la similitud de histogramas, que permita identificar y recomendar de forma automatizada casos con características de distribución de intensidad análogas a una imagen de referencia.
- Implementar arquitecturas de Deep Learning previamente entrenadas con imágenes naturales, para la extracción de vectores de características (embeddings) y la recuperación de imágenes de mamografías mediante métricas de similitud visual.
- Implementar arquitecturas de Deep Learning previamente entrenadas con imágenes mamográficas, para la extracción de vectores de características (embeddings) y la recuperación de imágenes de mamografías mediante métricas de similitud visual. Contrastar el desempeño diagnóstico de los modelos entrenados con

imágenes mamográficas vs los no entrenados, utilizando métricas de Accuracy, F1-Score y Recall, para determinar la viabilidad de su implementación en entornos clínicos.

- Diseñar un algoritmo de recomendación basado en el filtrado por contenido, orientado a personalizar la ruta de aprendizaje del residente, facilitando el reforzamiento sistemático de aquellas competencias diagnósticas que presenten un menor grado de consolidación.

1.5 Dataset

Para la validación de los modelos de Deep Learning, se utilizará el conjunto de datos público VinDr-Mammo [17]. Este es un dataset de referencia a gran escala que comprende mamografías digitales.

Características principales del conjunto de datos:

- **Volumen.** Contiene 5000 exámenes, cada uno con las cuatro vistas estándar de un examen de mamografía, lo que equivale a un total de 20000 imágenes en formato DICOM. Las vistas corresponden a la combinación de las siguientes características: lateralidad: Izquierda (L) y Derecha (R) y posición de la vista: Cráneo Caudal (CC) y Medio Lateral Oblicuo (MLO). Entonces las cuatro vistas son: CC - L, CC - R, MLO - L y MLO - R.
- **Validación.** Cada examen fue interpretado de forma independiente por dos radiólogos; en caso de discrepancias, los resultados fueron arbitrados por un tercer especialista para garantizar la fiabilidad de las anotaciones.
- **Estándar Clínico.** Los hallazgos y reportes siguen el léxico y esquema del estándar BI-RADS (Breast Imaging-Reporting and Data System) [18].
- **Anotaciones.** El dataset proporciona etiquetas detalladas sobre la densidad mamaria (categorías A, B, C y D) y hallazgos específicos como masas, calcificaciones, asimetrías y distorsiones arquitectónicas, entre otras.

Las siguientes tablas muestran la estadística de hallazgos (Tabla 1) y densidad (Tabla 2) en el dataset.

Tabla 1. Estadística de los hallazgos. Se proporciona la cantidad de hallazgos para los conjuntos de entrenamiento, de prueba y de datos completo. Basada en [17].

Hallazgo	Split		Total
	Entrenamiento	Prueba	
Masa	989 (80,67%)	237 (19,33%)	1.226
Calcificación sospechosa	428 (78,82%)	115 (21,18%)	543
Asimetría	77 (79,38%)	20 (20,62%)	97
Asimetría focal	216 (80,30%)	53 (19,79%)	269
Asimetría global	20 (76,92%)	6 (23,08%)	26
Distorsión arquitectónica	95 (79,83%)	24 (20,17%)	119
Engrosamiento cutáneo	45 (78,95%)	12 (21,05%)	57
Retracción cutánea	15 (93,33%)	3 (16,67%)	18
Retracción del pezón	30 (81,08%)	7 (18,92%)	37
Ganglio linfático sospechoso	46 (80,70%)	11 (19,30%)	57
Total	1.961 (80,70%)	488 (19,93%)	2.449

Tabla 2. Estadística de densidad mamaria. Basada en [17].

	Densidad Mamaria				Total
	A	B	C	D	
Entrenamiento	40 (0,50%)	764 (9,55%)	6.116 (76,45%)	1.080 (13,50%)	8.000
Prueba	10 (0,50%)	190 (9,50%)	1.530 (76,50%)	270 (13,50%)	2.000
Total	50 (0,50%)	954 (6,54%)	7.646 (76,46%)	1.350 (13,50%)	10.000

1.6 Definición de hipótesis

Para el diseño de una solución tecnológica que impacte en la formación radiológica, el sistema debe integrar funcionalidades que permitan: establecer una línea base de desempeño mediante evaluaciones iniciales, visualizar estudios en formato estándar DICOM, y proveer recomendaciones automatizadas de material de reforzamiento.

Sobre esta base funcional, se formula la siguiente hipótesis de trabajo:

"El sistema de apoyo para entrenamiento en radiología de mamas es capaz de recomendar recursos de aprendizaje pertinentes para un residente con un grado de precisión igual o superior a modelo preentrenado en mamografía a través de búsqueda por similitud de mamografías, teniendo como estrategia el uso de CNN que no serán reentrenadas. El cumplimiento de este umbral indicado permitirá proporcionar una retroalimentación relevante al radiólogo, fortaleciendo significativamente su capacidad diagnóstica."

Variables del Estudio

Para la validación experimental, se definen las siguientes variables:

- **Variable Independiente.** Los datos de entrada y la arquitectura de las Redes Neuronales Convolucionales (CNN) empleadas.
- **Variable Dependiente.** El grado de precisión alcanzado por el modelo en sus funciones de clasificación y recomendación.

1.7 Metodología de trabajo aplicada

La ejecución del trabajo se estructura en cuatro fases:

Fase 1. Preprocesamiento de Imágenes. Se realiza un análisis del conjunto de datos (dataset) para implementar técnicas de preprocesamiento específicas. En esta etapa se normalizan los datos y luego se aplican distintos preprocesamientos, generando distintas bases de datos, para posteriormente realizar búsquedas por histogramas.

Fase 2. Evaluación de Desempeño Modelo/Dataset. Se lleva a cabo un proceso de validación cruzada para determinar qué modelos presentan un mejor rendimiento sobre bases de datos específicas. El objetivo es establecer una correlación entre modelo y base de datos.

Fase 3. Capacidad de predicción. A partir de la definición de subconjuntos de datos de referencia y prueba (test), se calcula las inferencias de los modelos. Se utilizan embeddings para determinar la capacidad de predicción diagnóstica de los modelos.

Fase 4. Recomendación a partir de modelos. En la etapa final, se integran los modelos para ejecutar la clasificación de imágenes y a partir de esto buscar y determinar imágenes similares a recomendar.

1.8 Estructura del informe

En el segundo capítulo se revisa el marco teórico y estado del arte de los conceptos claves que se utilizarán en desarrollo del presente trabajo.

En el tercer capítulo, se aborda la ejecución de la metodología propuesta con los experimentos realizados y los resultados obtenidos. Se presenta una propuesta de visión arquitectónica del sistema, considerando la solución propuesta. Finalmente, se presenta una conclusión respecto de los resultados obtenidos, proponiendo consideraciones y próximos pasos que seguir.

2 Marco teórico y estado del arte

Para abordar el problema planteado se examinará el estado del arte a través de trabajos que traten un problema similar y comparándolos con la propuesta de solución. Los conceptos que se consideraran para resolver el problema planteado son: a) Modelo/Proceso de aprendizaje, b) Búsqueda de imágenes por similitud, c) Sistema de recomendación (SR).

2.1 Deep learning (DL)

Deep Learning o aprendizaje profundo, es una técnica de aprendizaje automático perteneciente al campo de la inteligencia artificial (IA), en específico al subcampo del Machine Learning (ML) o aprendizaje automático. Los algoritmos de Deep Learning se caracterizan por el uso de arquitecturas jerárquicas capaces de aprender abstracciones de alto nivel. Estas arquitecturas están formadas por capas de unidades de procesamiento apiladas, y se denominan redes neuronales artificiales (RNA) dado que su funcionamiento está inspirado por las redes de neuronas biológicas en los seres vivos [19].

2.1.1 Redes neuronales convolucionales (CNN)

David H Hubel y Torsten Wiesel [20], a través de una serie de experimentos, hacia fines de la década de 1950, demostraron que muchas de las neuronas de la corteza visual, solo reaccionan a estímulos visuales localizados en una región limitada del campo visual. En consecuencia, los campos receptivos de diversas neuronas presentan solapamientos, que de manera conjunta, configuran una representación integral y continua del campo visual a modo de mosaico. Estos estudios de la corteza visual inspiraron lo que ahora se conoce como Redes Neuronales Convolucionales (CNN, Convolutional Neural Networks).

CNN, es un tipo especializado de algoritmo de aprendizaje profundo diseñado principalmente para tareas que requieren reconocimiento de objetos, como la clasificación, la detección y la segmentación de imágenes.

Las CNN son una de las redes más importantes y estudiadas en el campo de DL, logrando resultados notables en áreas como la visión por computador y procesamiento del lenguaje [21].

2.1.2 Métricas de Deep Learning

Matriz de Confusión. Para crear una matriz de confusión, se necesitan cuatro atributos: a) Verdaderos positivos (VP): el modelo predijo una etiqueta y coincide correctamente con la verdad fundamental. b) Verdaderos

negativos (VN): el modelo no predice la etiqueta y no forma parte de la verdad fundamental. c) Falsos positivos (FP): el modelo predijo una etiqueta, pero no forma parte de la verdad fundamental (error de tipo I). d) Falsos negativos (FN): el modelo no predice una etiqueta, pero forma parte de la verdad fundamental (error de tipo II) [22].

Real	Positivo	Verdadero Positivo (VP)	Falso Positivo (FP)
	Negativo	Falso Negativo (FN)	Verdadero Negativo (VN)
		Positivo	Negativo
		Predicción	

Figura 1. Matriz de confusión. Basada en [23].

Accuracy: mide el porcentaje de predicciones correctas realizadas por el modelo sobre el total de predicciones. Aunque es una métrica ampliamente utilizada, puede ser engañosa en casos donde las clases están desbalanceadas, como ocurre frecuentemente en los datos médicos.

$$Accuracy = \frac{(VP + VN)}{(VP + VN + FN + FP)} \quad (1)$$

Precisión: evalúa la proporción de predicciones correctas sobre el total de predicciones positivas realizadas por el modelo. En otras palabras, indica la capacidad del modelo para evitar falsos positivos. Es especialmente relevante en el diagnóstico médico, donde los falsos positivos pueden generar costos innecesarios y estrés para los pacientes.

$$Precisión = \frac{VP}{(VP + FP)} \quad (2)$$

Recall: también llamado sensibilidad mide la capacidad del modelo para identificar todas las instancias positivas, evaluando qué tan bien el modelo detecta los verdaderos positivos en el conjunto de datos. Es crucial en aplicaciones médicas donde es importante reducir los falsos negativos, ya que estos representan casos en los que una condición, por ejemplo, una alta densidad mamaria, podría pasar desapercibida.

$$Recall = \frac{VP}{(VP + FN)} \quad (3)$$

F1-Score: es la media armónica de la Precisión y el Recall, proporcionando un único valor que equilibra ambas métricas. Es útil cuando se busca un balance entre Precisión y Recall. Un valor alto de F1-Score indica que el modelo tiene un buen balance entre la Precisión y el Recall, lo que lo hace adecuado para situaciones con clases desbalanceadas. Esta métrica es crítica en la clasificación de densidad mamaria, donde tanto los falsos positivos como los falsos negativos pueden tener implicaciones significativas en el manejo clínico.

$$F1 - Score = \frac{(2 * Precisión * Recall)}{(Precisión + Recall)} \quad (4)$$

2.2 Búsqueda de imágenes por similitud

La recuperación de imágenes similares con base en su contenido (Content-Based Image Retrieval, CBIR, por sus siglas en inglés) surgió a partir de la necesidad de clasificar, indexar y recuperar imágenes de interés de una gran base de datos sin la necesidad de utilizar metadatos, los cuales han demostrado ser ambiguos y subjetivos [24]. Los sistemas CBIR dependen crucialmente de dos elementos: embedding y medidas de similitud.

Las redes neuronales convolucionales (CNN) junto con el aprendizaje profundo ofrecen varias soluciones en el campo de la visión por computadora, ya que las redes CNN modelan abstracciones de alto nivel. La manera de utilizar las redes CNN en los sistemas CBIR es utilizar la salida de una de las últimas capas completamente conectadas (FC6, FC7 o FC8 en el caso de Alexnet) de la red como un vector característico (embedding) para describir cada una de las imágenes en la base de datos [25]. Estos embeddings son almacenados y cuando se requiere hacer una consulta, la red CNN extrae el vector de características de la imagen de consulta y éste es comparado con cada uno de los embeddings en la base de datos; los vectores que tengan la distancia más corta corresponderán a las imágenes de la base de datos que serán similares a la imagen de consulta.

De lo anterior, se puede apreciar que los sistemas CBIR, pueden operar tanto con CNN entrenadas en imágenes naturales (*ImageNet*) o con aquellas que usan los pesos de otra CNN reentrenada.

Los sistemas CBIR trabajan en dos etapas, la primera es la generación de los vectores característicos llamados descriptores, donde cada una de las imágenes de la base de datos es representada por un vector numérico; dichos vectores son guardados en una base de datos. En la segunda etapa, una imagen de consulta es dada, posteriormente, se representa con un vector característico, el cual, es comparado utilizando una medida de distancia con cada uno de los vectores de la base de datos; los vectores que tengan una distancia menor corresponderán a imágenes similares a la imagen de consulta dada.

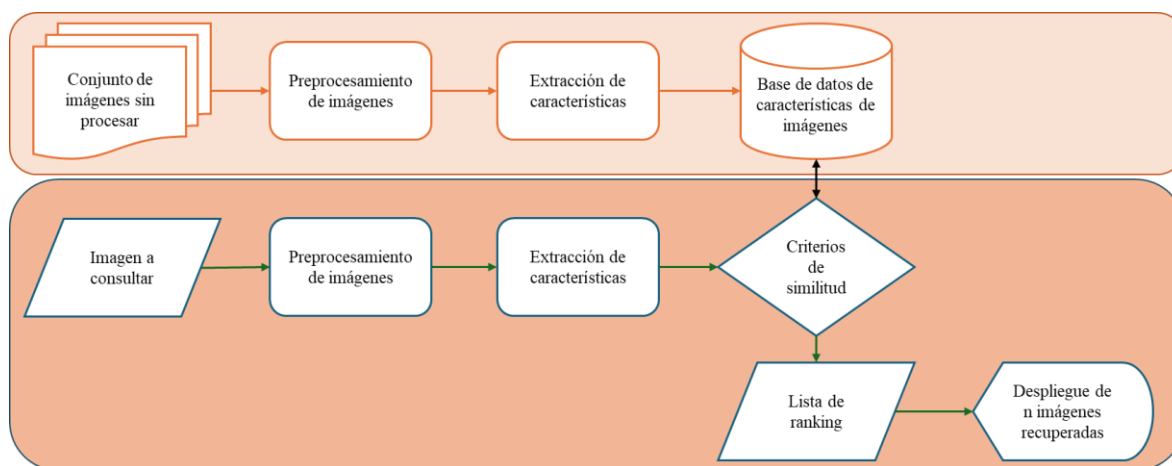


Figura 2. Arquitectura del sistema CBIR. Basado en [26].

2.3 Sistema de recomendación

Los sistemas de recomendación analizan los intereses de los usuarios y recomiendan algo en función de esos intereses. El objetivo de estos sistemas es reducir la sobrecarga de información recuperando los artículos similares en función del interés del cliente. El uso principal de estos sistemas es la toma de decisiones, la

maximización de beneficios y la reducción de riesgos. Esto reduce el esfuerzo y el tiempo del cliente en la búsqueda de información. Funciona como un filtro que sugiere alternativas en función de los datos [27].

Un sistema de recomendación aprende el comportamiento del usuario y predice sus preferencias actuales (recomienda) en función de sus preferencias anteriores. La implementación de la IA como sistema de recomendación mejora la precisión de esta predicción y resuelve los problemas de arranque en frío y escasez de datos [28].

Las aplicaciones de hoy ocupan diferentes algoritmos de Machine Learning (ML), como K-nearest neighbor (KNN), Naive Bayes, Random Forest, Adaboost, Singular value decomposition (SVD), entre otros. La evolución del esquema de recomendación ha llevado a la aplicación de algoritmos de ML e Inteligencia Artificial (IA) para una predicción y precisión efectivas [29].

La idea de los sistemas de recomendación es brindar recomendaciones al usuario de acuerdo con su comportamiento o perfil. Analizar el interés del usuario de forma dinámica para que cuando el usuario realice acciones, recomiende de acuerdo con sus preferencias.

Los sistemas de recomendación se pueden dividir principalmente en tres categorías [30]:

- **Filtrado colaborativo.** Compara usuarios con preferencias similares y recomienda probar elementos que otros usuarios han probado.
- **Filtrado basado en contenido.** Modelo analiza otros elementos con atributos similares a los de la búsqueda y los recomienda al usuario.
- **Filtrado híbrido.** Estos sistemas de recomendación se basan en los datos del elemento y en la información del usuario.

2.4 Estándar BI-RADS

El estándar BI-RADS, desarrollado por el Colegio Americano de Radiología (ACR), constituye el marco de referencia universal para la estandarización del reporte e interpretación de imágenes mamarias. Su implementación tiene como objetivo principal reducir la ambigüedad en los informes radiológicos y facilitar la comunicación interdisciplinaria mediante un léxico técnico unificado y una categorización del riesgo bien definida [18].

En el contexto de este trabajo, el estándar BI-RADS es relevante en dos dimensiones diagnósticas:

- **Evaluación de la Densidad Mamaria:** La densidad de la mama se clasifica en cuatro categorías (A, B, C y D), basándose en la proporción de tejido fibroglandular frente al tejido graso. Esta clasificación es crítica, ya que una mayor densidad (categorías C y D) puede enmascarar lesiones pequeñas, disminuyendo la sensibilidad de la mamografía y aumentando la complejidad del entrenamiento para el residente.
- **Categorización de Hallazgos y Conducta Clínica:** El BI-RADS asigna una escala numérica (del 0 al 6) que correlaciona los hallazgos visuales con una probabilidad estadística de malignidad. Esta escala dicta la conducta médica a seguir, desde el control anual de rutina hasta la necesidad de una biopsia inmediata.

La Figura 3, muestra ejemplos de imágenes con densidad categoría D (a), Suspicious Calcification (b) y Mass (c). Obsérvese que para las imágenes (b) y (c), el hallazgo es acotado por un rectángulo. Llama la atención lo pequeño del rectángulo para la imagen con Suspicious Calcification.

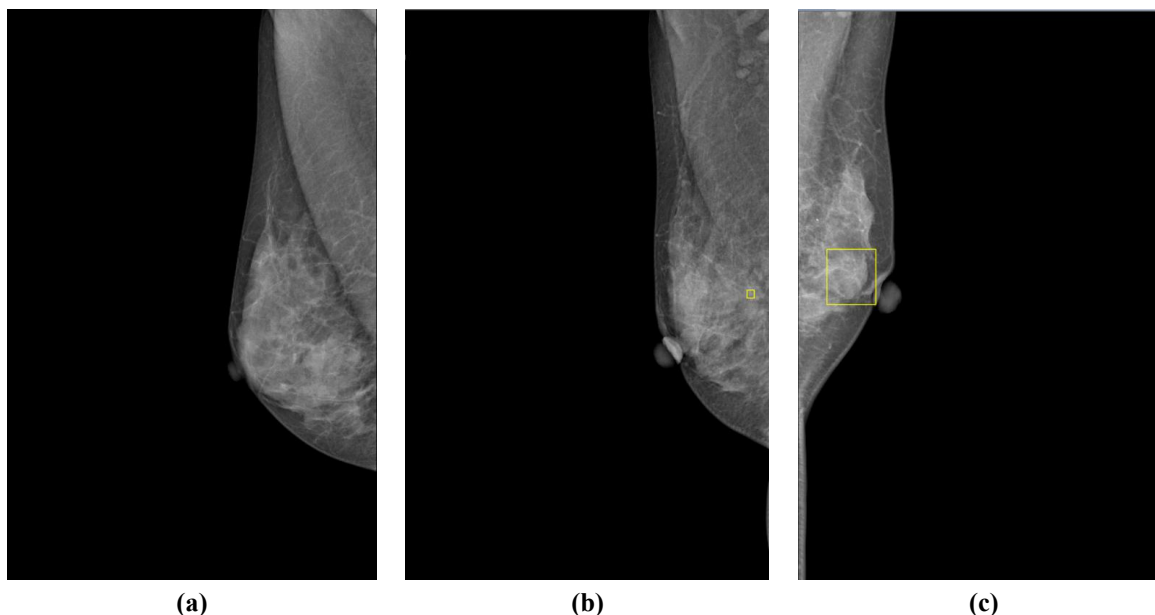


Figura 3. (a) Mama con densidad categoría D (lateralidad derecha, vista medio lateral oblicuo) (b) Mama con hallazgos de Suspicious Calcification (lateralidad derecha, vista medio lateral oblicuo). (c) Mama con hallazgos de Mass (lateralidad izquierda, vista medio lateral oblicuo). Fuente [17]

2.5 Modelo/Proceso de aprendizaje

De acuerdo con [31], existen dos enfoques en la educación en radiología: el enfoque de **abajo hacia arriba** ocurre cuando los estudiantes obtienen experiencia de primera mano con los casos al diagnosticarlos ellos mismos. Y el enfoque de **arriba hacia abajo** es cuando el médico tratante diagnostica el caso y luego enseña a los estudiantes los puntos de aprendizaje relevantes. Este último enfoque se ha visto ser una mejor herramienta de aprendizaje para los becarios de radiología.

En la Figura 4, describen el término llamado “educación médica de precisión” que es una posible utilidad de la inteligencia artificial (IA) para brindar conocimientos personalizados al estudiante de radiología en función de sus habilidades y conocimientos previos, que plantea un conjunto de pasos para la aplicación de la IA en el entrenamiento en radiología [1].

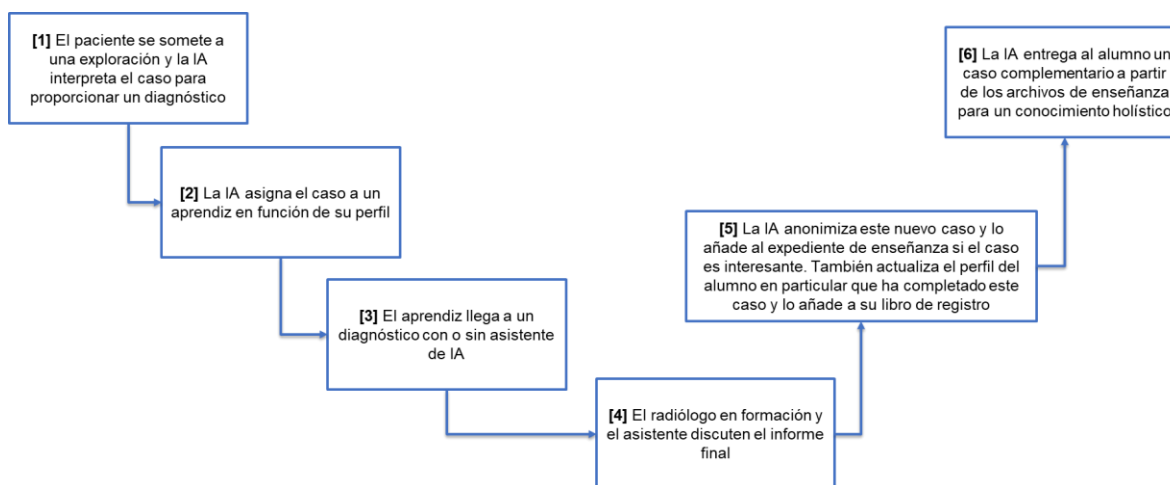


Figura 4. Pasos para la aplicación de la IA en el entrenamiento en radiología. Basado en [1].

2.6 Estado del arte

Para abordar el estado del arte se hizo foco en tres aspectos a revisar: el proceso de aprendizaje usado para la enseñanza; casos de uso de CNN sin reentrenar; y el uso de CNN para resolver problemas de clasificación y recuperación; en el contexto de imágenes de mamas.

Existen esfuerzos para mejorar el proceso de aprendizaje de los residentes, por ejemplo, a través de “aprender haciendo”. El trabajo [32], propone un nuevo paradigma basado en simulación, con el objetivo de mejorar la experiencia de los estudiantes de medicina y brindar una oportunidad de aprendizaje activo como parte de su rotación de radiología. A través de un software de procesamiento de texto común, un PACS de uso gratuito e instructores residentes, se dispone a los estudiantes en un entorno de sala de lectura simulada. Los estudiantes interpretan casos preseleccionados, dictan informes y discuten los hallazgos con la sesión informativa de los residentes. En la validación del paradigma, los estudiantes le otorgan una alta calificación, proporcionando una experiencia novedosa de una manera rentable. Dentro de sus limitaciones se tiene la necesidad de ampliar la biblioteca de casos para abarcar más patología, sutileza de los hallazgos o presentaciones complejas, potencialmente de múltiples modalidades, podría mejorar la experiencia de aprendizaje.

Desde una perspectiva tecnológica, ha habido un amplio desarrollo de entornos de formación radiológica computarizados que combinan oportunidades de aprendizaje tradicionales con plataformas avanzadas de aprendizaje electrónico, con el objetivo de apoyar la adquisición de experiencia radiológica. MamODB [33], a través del paradigma aprendizaje basado en problemas, logra la adquisición de conocimiento clínico mediante las anotaciones emitidas por un panel de radiólogos expertos (almacenados en una base de datos) y una estrategia de aprendizaje, para la formación de radiólogos (es decir, analizando los casos clínicos y comparando los hallazgos con las anotaciones expertas). Las consultas a MamODB se realizan a través de un formulario web y proporcionan al usuario una lista de casos clínicos de acuerdo con los parámetros de consulta, tanto incluyendo la imagen como los archivos de anotación.

Con el objetivo de facilitar la tarea de docentes en la creación de casos, donde las imágenes radiológicas son el componente principal el trabajo [34] crea un marco de trabajo RadEd (Radiology Education), para la enseñanza basado en la web para crear ejercicios basados en casos que permiten la interacción con imágenes. Requiere

que los docentes preparen material teniendo en cuenta que se accederá en línea, y esto requiere un esfuerzo extra para ser presentado de una manera más pedagógica.

En el trabajo de [35] se desarrolla el concepto de recuperación de casos basada en imágenes para la educación radiológica (IBCR-RE, de las siglas en ingles de Image-Based Case Retrieval for Radiological Education). La capacitación diagnóstica de la IBCR-RE está integrada en un marco didáctico basado en el enfoque de aprendizaje basado en problemas (ABP). El sistema de aprendizaje diagnóstico desarrollado ayuda a los residentes en la exploración casos de pacientes similares. El trabajo usa el marco de trabajo de recuperación de imágenes en aplicaciones médicas, IRMA (siglas de Image Retrieval in Medical Applications) [36].

La recuperación de imágenes con base en su contenido (Content-Based Image Retrieval, CBIR, por sus siglas en inglés), hoy juegan un rol importante al momento de recomendar una imágenes. Un elemento relevante en un sistema CBIR, es la CNN que se use para la extracción de vectores de características de las imágenes.

Desde la perspectiva de la estrategia de solución de uso de CNN sin reentrenar, el trabajo de [15], explora la capacidad de una CNN para identificar diferentes tipos de patologías en imágenes de rayos X de tórax, utilizando una CNN que fue entrenada con *ImageNet* [14]. El rendimiento obtenido, área bajo la curva (AUC) fue de 0.93 para la detección de derrame pleural derecho, 0.89 para la detección de agrandamiento cardíaco y 0.79 para la clasificación entre radiografía de tórax sana y anormal. Estos resultados, permiten indicar que el aprendizaje profundo con entrenamiento en *ImageNet* puede ser suficiente para tareas generales de reconocimiento de imágenes médicas.

De manera parecida en el trabajo de [11], se examinan los vectores de características extraídas por CNN y analiza su efectividad en la recuperación de imágenes médicas similares. Uno de sus resultados indica que para una misma arquitectura de CNN, no hay diferencias significativa en el rendimiento de recuperación, independientemente de si las características extraídas son con los pesos de *ImageNet* o sin.

En el contexto de proyecto [37] los trabajos de [38] y [39] los modelos DenseNet y SwinV2_tiny_window8_256, son reentrenados usando imágenes de mamografías, para obtener un mejor resultado en la búsqueda de densidad mamaria.

Entendiendo que las CNN juegan un rol relevante en un sistema CBIR, la pregunta que surge es que arquitecturas de CNN, son las más usadas en la recuperación de imágenes médicas. A pesar de que la literatura no es taxativa al respecto, a continuación se presentan algunos trabajos que usan CNN.

Dentro de las arquitecturas de CNN, para el análisis de imágenes médicas, el trabajo [40], señala que las CNN han logrado un gran progreso en las aplicaciones de clasificación, segmentación y detección en imágenes médicas. Gracias a arquitecturas como VGG, ResNet e Inception, las CNN han mejorado logrando ser capaces de detectar enfermedades, segmentar órganos y clasificar tumores con bastante precisión.

En el contexto de búsqueda de imágenes basada en contenido, usando tomografías computarizadas de pulmón el estudio [41] analiza modelos de aprendizaje profundo preentrenados conocidos, como la red neuronal profunda VGG16 y la red neuronal residual ResNet50.

Usando un enfoque dividido en 3 pasos, preprocesamiento de imágenes, extracción de características y clasificación, para clasificar tumores de cáncer de mamas, el trabajo [42], usa modelos de CNN preentrenados como extractores de características, como ResNet50, DenseNet, y VGG16, entre otros. Obtiene mejores rendimientos con los CNN ResNet50 y DenseNet.

Los trabajos mencionados anteriormente tratan principalmente el problema de apoyo al entrenamiento de los radiólogos. Las tecnologías usadas son variadas desde aplicaciones web hasta técnicas de Deep Learning. El presente trabajo se centra en generar apoyo a los radiólogos en base a la búsqueda de imágenes similares y recomendaciones de filtrado por contenido, teniendo como estrategia el uso de CNN sin reentrenar. La combinación de estas técnicas debe mejorar la capacidad del sistema de predecir con precisión los recursos de aprendizaje para el aprendiz.

3 Desarrollo de solución

Para la detección temprana de cáncer de mamas, es relevante considerar el grado de densidad mamaria (Density A, B, C, y D). La elección de otros hallazgos que ayuden a la detección temprana quedó determinada por la cantidad de datos (imágenes) que se tienen, por lo cual, se consideraron los siguientes hallazgos: Mass y Suspicious Calcification. Así, estos elementos son centrales en el presente trabajo, pues los modelos deben apoyar en la búsqueda de imágenes que contengan una densidad específica y/o Mass y/o Suspicious Calcification. En consecuencia con lo indicado en sección anterior, los modelos con los cuales se decidió trabajar, que se entrenaron en imágenes naturales (*ImageNet*) fueron: ResNet50 y VGG16. Y aquellos que usan los pesos de otra CNN reentrenada con imágenes mamográficas como lo son: DenseNet y SwinV2, estas últimas especialmente entrenadas para clasificar la densidad mamaria.

3.1 Preprocesamiento de imágenes

Para mejorar la detección de los elementos señalados, se utilizaron técnicas de procesamiento de imágenes y establecieron la mejora de la imagen de la mamografía [43]. A partir de esto se crearon 4 base de datos de imágenes, a saber, como muestra la Figura 5.

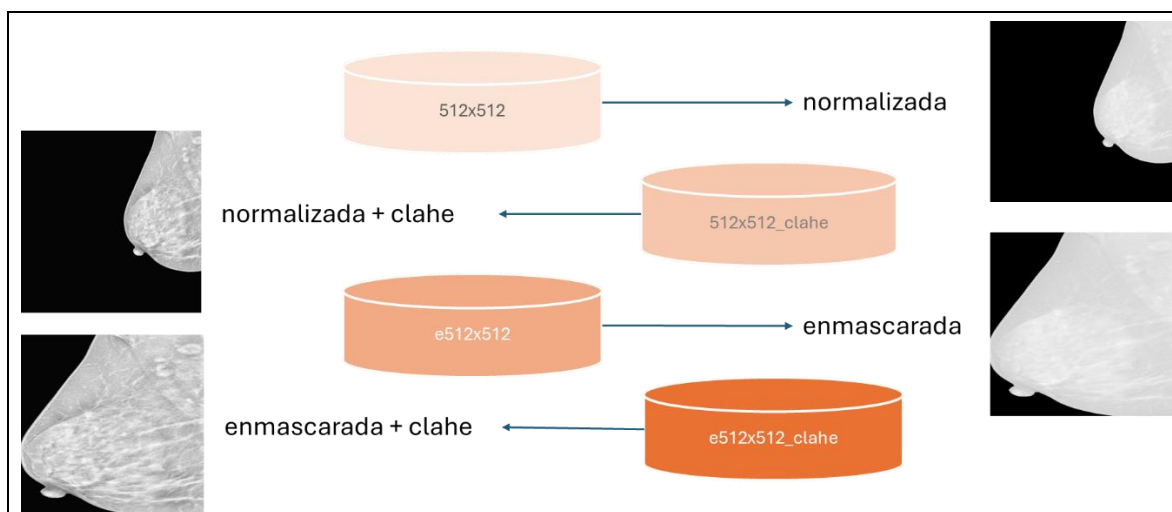


Figura 5. Base de datos generadas a partir del preprocesamiento de imágenes, con ejemplo de imagen de lateralidad derecha y proyección medio lateral oblicuo (MLO). Fuente propia.

Las bases de datos están compuestas por: a) Imágenes en el tamaño de 512 x 512 píxeles. b) Normalización a 8 bits, con valores de píxeles en el rango [0 255]. c) Uso de tres canales en escala de grises. La base de datos “512x512_clahe”, se generó a partir de la base de datos “512x512”, a la cual se le aplicó CLAHE (Contrast

Limited Adaptive Histogram Equalization), que permite mejorar el contraste local, especialmente en imágenes con regiones muy oscuras o brillantes. A diferencia de la ecualización de histograma global (HE), que aplica una única transformación a toda la imagen, CLAHE divide la imagen en pequeñas regiones y realiza una ecualización local. Esto ayuda a mejorar los detalles en áreas de bajo contraste sin amplificar excesivamente el ruido en las regiones uniformes [44]. La base de datos “e512x512”, se generó a partir de la base de datos “512x512”, a cuyas imágenes se les aplicó enmascaramiento, que de forma no destructiva permite controlar la visibilidad del contenido de la imagen. Específicamente se aplicó para eliminar áreas de la imagen sin información. La base de datos “e512x512_clahe”, se generó a partir de la base de datos “e512x512”, a cuyas imágenes se les aplicó CLAHE.

3.2 Evaluación de Desempeño Modelo/Dataset

Esta sección presenta la evaluación de desempeño de los modelos implementados, con el propósito de establecer una línea base (baseline) de referencia para los experimentos posteriores. Los modelos evaluados fueron:

DenseNet, entrenado con imágenes de mamografías digitales [38], se utilizó para predecir el grado de densidad mamaria con un F1-Score de 0.815. Para nuestro propósito se utilizaron los pesos de modelo. Esta red se utilizó para recomendar imágenes por su grado de densidad mamaria.

SwinV2_tiny_window8_256, entrenado con imágenes de mamografías digitales [37], al igual que el anterior se utilizó para predecir el grado de densidad mamaria con un F1-Score de 0.578. Para nuestro propósito se utilizaron los pesos de modelo. Esta red se utilizó para recomendar imágenes por su grado de densidad mamaria.

ResNet50, entrenamiento con imágenes de ImageNet [14]. Para nuestro propósito se utilizaron los pesos del modelo. Esta red se utilizó para recomendar imágenes por sus hallazgos Mass/Suspicious Calcification que son los más comunes dentro de la base de datos analizada.

VGG16, entrenamiento con imágenes de ImageNet [14]. Para nuestro propósito se utilizaron los pesos del modelo. Esta red se utilizó para recomendar imágenes por sus hallazgos Mass/Suspicious Calcification que son los más comunes dentro de la base de datos analizada.

Con el fin de evaluar el desempeño de cada modelo en función de las distintas bases de datos, se llevó a cabo el siguiente procedimiento:

- Para cada base de datos y modelo:
 - El embedding de cada una de las imágenes dentro de una base de datos (los cuales actuarán como el embedding a consultar), se comparó con los embeddings del resto de las imágenes dentro de la misma base de datos, a través de la distancia euclidiana.
 - Se seleccionaron las 100 menores distancias.
 - Se determinó la posición en que se encuentra la primera imagen que contiene la misma densidad/hallazgo de la imagen de entrada, dentro de las 100 menores distancias.
- Para comparar los modelos, se utilizó la métrica de **posición promedio**. Esta medida se interpreta de la siguiente manera:
 - A partir de la posición determinada en el punto anterior, esta fue almacenada en un arreglo, para cada una de las imágenes dentro de la base de datos. Posteriormente esta información se promedió respecto al tipo de densidad/hallazgo que poseía la imagen de entrada.
 - Un valor bajo indica que la imagen se encuentra en primeras posiciones y que, por lo tanto, el desempeño del modelo es mejor.
 - Un valor alto señala que es más costoso encontrar la densidad/hallazgo.

En la Figura 6, se muestra cómo se comporta la métrica de posición promedio, en la categoría densidad. En las bases de datos, donde se aplicó CLAHE (512x512_clahe y e512x512_clahe), se aprecia que la métrica de posición promedio tiende a ser más baja con respecto a las otras bases de datos (512x512 y e512x512). La excepción se aprecia en la categoría DENSITY C. El aplicar enmascaramiento a las imágenes, por sí solo no es suficiente para que los modelos tengan un mejor rendimiento, la métrica de posición promedio tiende a aumentar. Nuevamente la excepción se da en la categoría DENSITY C. El uso combinado de CLAHE y enmascaramiento, ayuda a los modelos a tener un mejor rendimiento, es decir, una métrica de posición promedio menor. Con los modelos DenseNet y SwinV2, se obtiene una métrica de posición promedio menor respecto de los otros modelos. La excepción se aprecia en la categoría DENSITY C, para modelo DenseNet. A partir de lo anterior se determina que los modelos DenseNet y SwinV2¹ tienen un mejor rendimiento en la base de datos enmascarada y con CLAHE, e512x512_clahe.

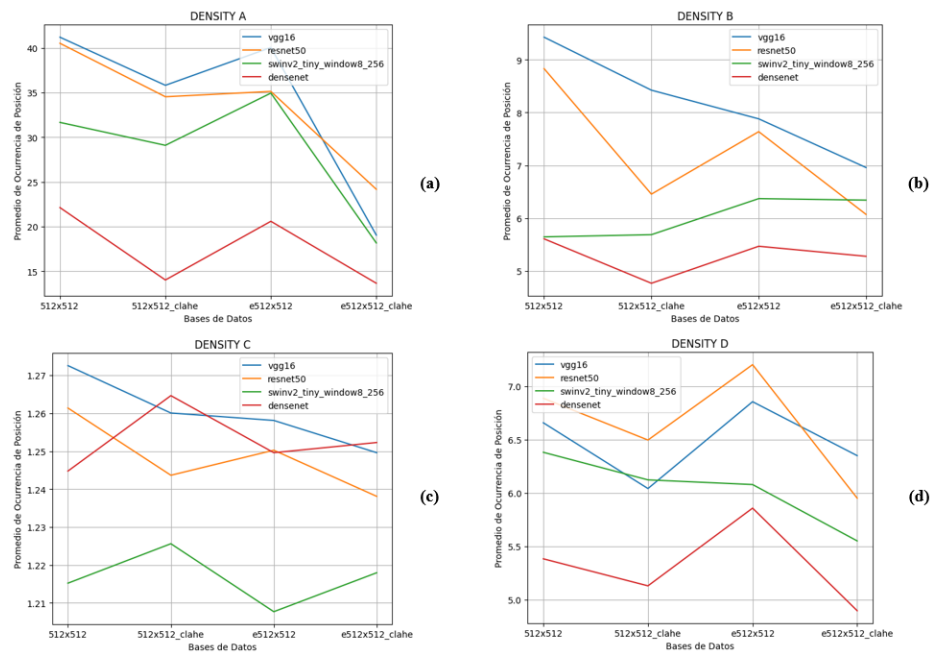
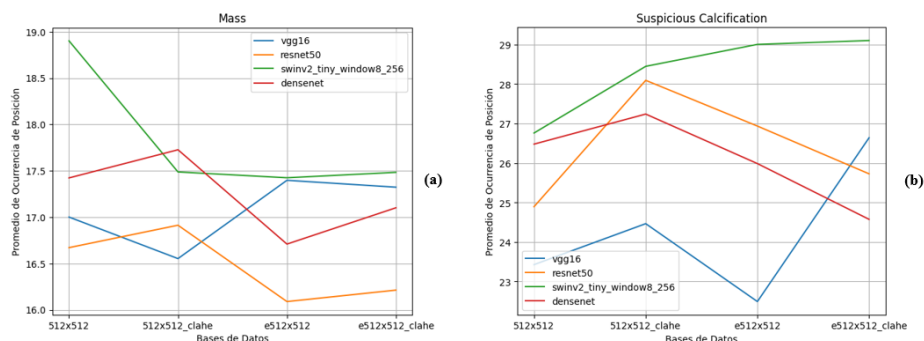


Figura 6. Promedio de ocurrencia de posición por densidad, modelo y base de datos. Fuente propia.

En la siguiente figura (Figura 7), se muestra cómo se comporta la métrica de posición promedio para los hallazgos Mass y Suspicious Calcification. Para el hallazgo Mass (a) se puede determinar que el modelo ResNet50 tiene un mejor rendimiento en la base de datos enmascarada y sin CLAHE (e512x512), para el reconocimiento del hallazgo Mass; esto dado que obtiene una métrica de posición promedio menor respecto de los otros modelos. Para el hallazgo Suspicious Calcification, se puede determinar que el modelo VGG16 tiene un mejor rendimiento en la base de datos enmascarada y sin CLAHE (e512x512), para el reconocimiento de hallazgo Suspicious Calcification; esto dado que obtiene una métrica de posición promedio menor respecto de los otros modelos.

¹ Nombre resumido, correspondiente a modelo: swinv2_tiny_window8_256

Figura 7. Métrica de posición promedio para los hallazgos Mass (a) y Suspicious Calcification (b) por modelo y base de



datos. Fuente propia.

3.3 Capacidad de predicción

3.3.1 Definir datos de referencia y test.

Para determinar la capacidad de predicción de los modelos utilizados se definieron dos subconjuntos de datos (Figura 8):

- **De referencia.** Datos etiquetados que servirán de referencia para compararse con los datos de Test. Estos datos fueron seleccionados desde la base de datos asociada al modelo (definido en sección anterior), es decir:
 - Para la clasificación de densidad mamaria, e512x512_clahe tuvo un mejor desempeño en los modelos DenseNet y SwinV2.
 - Para la clasificación de Mass, e512x512 tuvo un mejor desempeño en el modelo ResNet50.
 - Para la clasificación de Suspicious Calcification, e512x512 tuvo un mejor desempeño en el modelo VGG16
- **De Test.** Datos que servirán para probar la capacidad de predicción de los modelos. Las fuentes elegidas corresponden a imágenes que nunca han visto los modelos. Las fuentes para estos datos son: RSNA [45], para imágenes con categoría de densidad y CBIS-DDSM [46], para imágenes con hallazgo Mass y/o Suspicious Calcification.

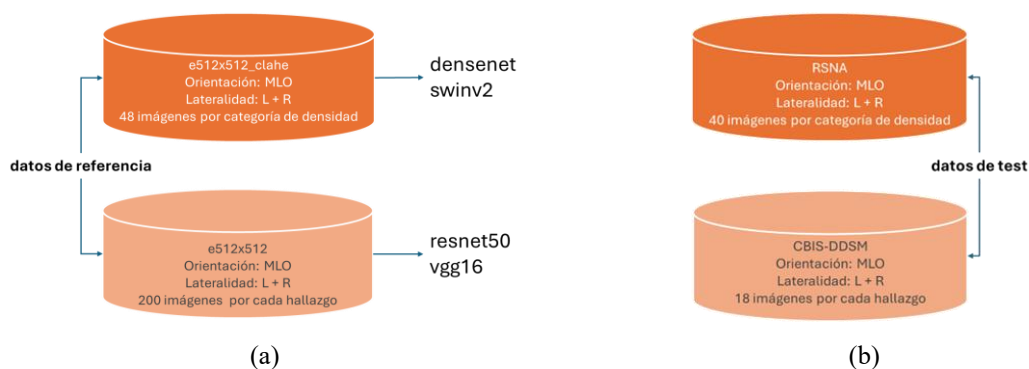


Figura 8. Detalles de los datos de referencia (a) y de test (b). Fuente propia.

La arquitectura de los conjuntos de datos de referencia y test se diseñó bajo un criterio de balanceo de clases (véase Tabla 3). Este enfoque garantiza la equidad estadística, impidiendo que una característica particular predomine sobre las demás y sesgue las métricas de rendimiento.

Tabla 3. Cantidad de datos usando como referencia y test. Fuente propia.

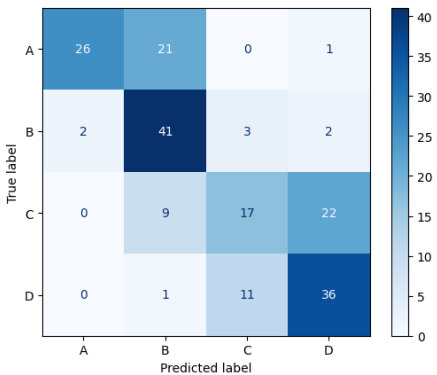
Cantidad de Datos			
	Densidad	Referencia	Test
	A	48	40
	B	48	40
	C	48	40
	D	48	40
	Hallazgo	Referencia	Test
	Mass	200	18
	Suspicious Calcification	200	18

3.3.2 Calcular la inferencia de los modelos con datos de referencia

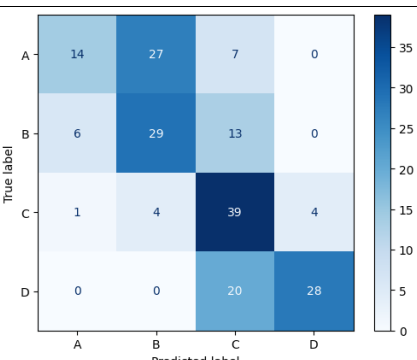
Con el objeto de tener un punto de comparación, al momento de establecer cuál es la capacidad de predicción diagnóstica de los modelos con los datos de test, es que a se procedió clasificar las imágenes de referencia a partir de los modelos descritos anteriormente.

En la Tabla 4, se muestra los resultados obtenidos para los modelos DenseNet y SwinV2, para clasificar imágenes en categoría de densidad mamaria. El modelo DenseNet presenta un mejor accuracy, respecto del modelo SwinV2. El modelo DenseNet, tiene la mayor Precisión, en la clase A (0.929), el Recall mayor se da en clase B (0.854). En las clases A, B y D, F1 Score es comparable, con valores más cercanos. El peor desempeño se obtuvo en la clase C. Por otro lado, en el modelo SwinV2, la mayor Precisión se da en la clase D (0.875). El Recall mayor se da en clase C (0.812). Con respecto a la métrica F1 Score, el mayor valor (0.700) se da en la clase D. Para la clase A, se distingue el menor valor para F1 Score. Para el resto de las clases (B, C y D) se aprecia que los valores están más distantes que los que muestra el modelo DenseNet.

Tabla 4. Resultados obtenidos de aplicar inferencia de densidad mamaria para modelos DenseNet y SwinV2. Fuente propia.

Densidad mamaria	DenseNet			
	Precisión	Recall	F1-Score	
A	0,929	0,542	0,684	
B	0,569	0,854	0,683	
C	0,548	0,354	0,430	
D	0,590	0,750	0,661	
Accuracy	0,625			

SwinV2			
Densidad mamaria	Precisión	Recall	F1-Score
A	0,667	0,292	0,406
B	0,483	0,604	0,537
C	0,494	0,812	0,614
D	0,875	0,583	0,700
Accuracy			
0,573			



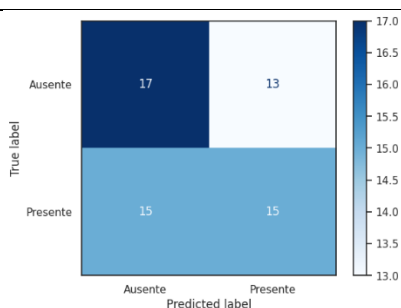
True \ Predicted	A	B	C	D
A	14	27	7	0
B	6	29	13	0
C	1	4	39	4
D	0	0	20	28

A diferencia de las inferencias hechas con los modelos DenseNet y SwinV2, en los modelos ResNet50 y VGG16, se utilizaron los vectores de clasificación de salida (de dimensión 1000) como descriptores de características. Para evaluar la proximidad en ese espacio, se usó el clasificador k-NN, parametrizado mediante la métrica Distancia Euclidiana, para garantizar una comparación coherente con la naturaleza normalizada de los datos.

En la Tabla 5, se muestra los resultados obtenidos para los modelos ResNet50 y VGG16, para clasificar imágenes en Mass y Suspicious Calcification, respectivamente. En modelo ResNet50, para la clase Presente se obtiene una Precisión de 0.536 y un Recall (0.500) que indica que 50% de los casos presentes no son detectados. En ambas clases, F1 Score se observa equilibrado respecto de los indicadores Precisión y Recall. Los valores de Precision y F1-score cercanos a 0.533 sugieren una capacidad discriminativa reducida. En modelo VGG16 se tiene Precisión de 0.576 y un Recall de 0.633, en la clase Presente. Al igual que modelo anterior, en ambas clases, F1 Score se observa equilibrado respecto de los indicadores Precisión y Recall. Los valores de accuracy de 58.3% y valores de F1-score entre 0.56 y 0.60, sugiere una capacidad de discriminación moderada.

Tabla 5. Resultados obtenidos de aplicar inferencia de Mass y Suspicious Calcification para modelos ResNet50 y VGG16, respectivamente. Fuente propia.

ResNet50			
Hallazgo Mass	Precisión	Recall	F1-Score
Ausente	0,531	0,567	0,548
Presente	0,536	0,500	0,517
accuracy			
0,533			

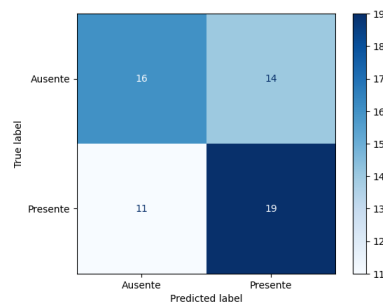


True \ Predicted	Ausente	Presente
Ausente	17	13
Presente	15	15

VGG16			
Hallazgo Suspicious Calcification	Precisión	Recall	F1-Score
Ausente	0,593	0,533	0,561
Presente	0,576	0,633	0,603

accuracy

0,583



3.3.3 A partir de embeddings establecer rendimiento de modelos con datos de test

Para establecer cuál es el rendimiento de los modelos, para cada set de datos de test, se calculó la matriz de confusión y sus métricas asociadas, la cual nos permitió tomar decisiones pertinentes.

En la siguiente figura (Figura 9) se muestra el proceso para calcular la matriz de confusión, este proceso consiste en lo siguiente: Se procede obtener los embeddings de cada dato de referencia y de test. Para cada embedding de las imágenes de test, se calcula la similitud usando como métrica la distancia euclidiana con respecto a los datos de referencia y se obtienen las **n** mejores distancias euclidianas (se usaron 10, 20, 30, 40, 50 y 100 mejores distancias). Posteriormente, se calcula la probabilidad de ocurrencia del hallazgo/densidad de la imagen de test en las imágenes de referencia. Esto se calcula contando la cantidad de veces que se repite el hallazgo/densidad en las **n** imágenes de referencia y se divide por **n**. Finalmente, con los datos anteriores se calcula matriz de confusión. De esto se deduce que por cada modelo se calcularon 6 matrices de confusión por cada modelo (ver Anexo A, Métricas de matrices de confusión calculadas).

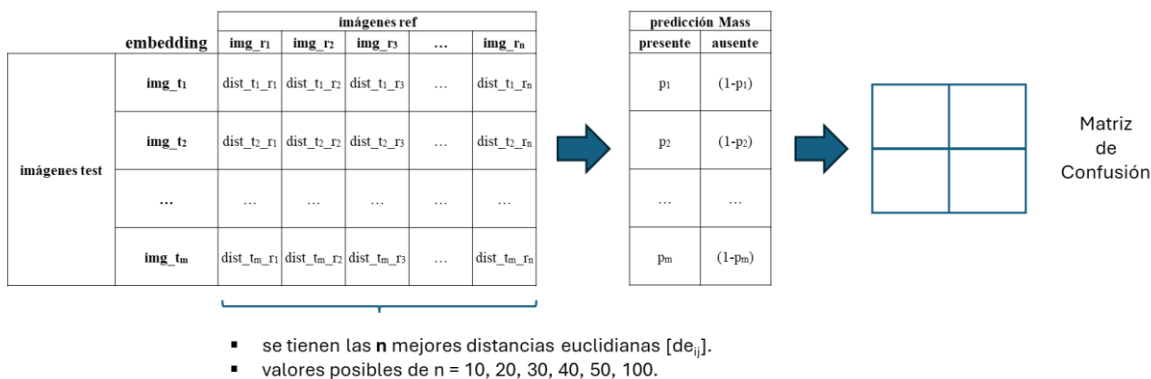


Figura 9. Proceso para el cálculo de matriz de confusión, tomando como ejemplo el hallazgo Mass. Fuente propia.

En la Figura 10, se muestra cómo se comporta la métrica accuracy, en función de los modelos y las **n** mejores distancias. El modelo DenseNet presenta un accuracy más alto, mientras que SwinV2, está en la vereda opuesta, un accuracy más bajo. Por otro lado, es el modelo DenseNet el que presenta valores más estables respecto de los otros modelos.

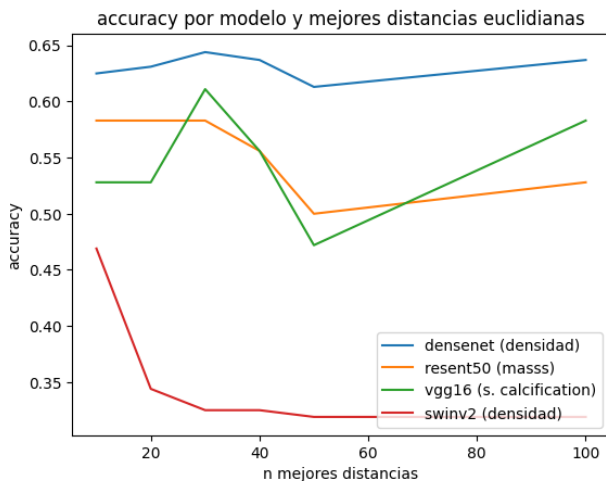


Figura 10. Accuracy en función de modelo y n mejores distancias. Fuente propia.

Teniendo en cuenta que la métrica accuracy, puede esconder algunas fallas en los modelos, se hizo una revisión de la métrica F1 Score. En la Figura 11, se muestra la métrica F1 Score en función de los modelos aplicados para clasificar densidad mamaria y las n mejores distancias. En el modelo DenseNet, el comportamiento de los datos de test es similar a la inferencia calculada con los datos de referencia (ver sección 3.3.2 Calcular la inferencia de los modelos). La clase C en el modelo DenseNet no tiene un buen rendimiento. En el modelo SwinV2, en general no tiene un buen comportamiento F1 Score, tiende a ser menor a 0.5 en todas las clases.

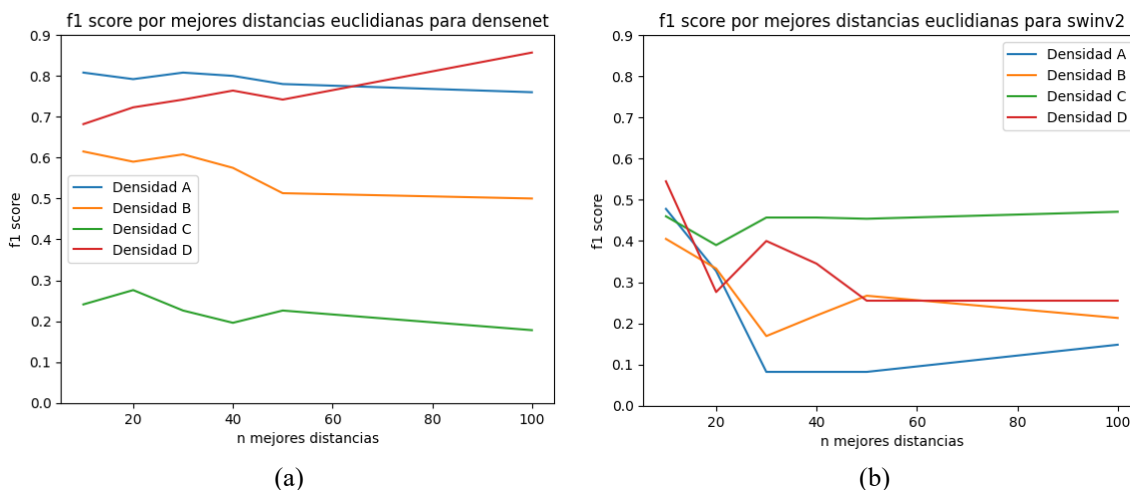


Figura 11. (a) F1 Score en función de las n mejores distancias para modelo DenseNet (b) F1 Score en función de las n mejores distancias para modelo SwinV2. Fuente propia.

En la Figura 12, se muestra cómo se comporta la métrica F1 Score, en función de las n mejores distancias, para los modelos ResNet50 y VGG16 para los hallazgos Mass y Suspicious Calcification, respectivamente. El modelo ResNet50, ratifica lo expresado en el cálculo de inferencia, el modelo no tiene buen rendimiento. En el caso del modelo VGG16, a pesar de que no tiene un buen rendimiento, llama la atención, la comparación en las 30 mejores distancias, donde F1 Score es 0.632 y 0.588, para las clases Suspicious Calcification presente y Suspicious Calcification ausente, respectivamente.

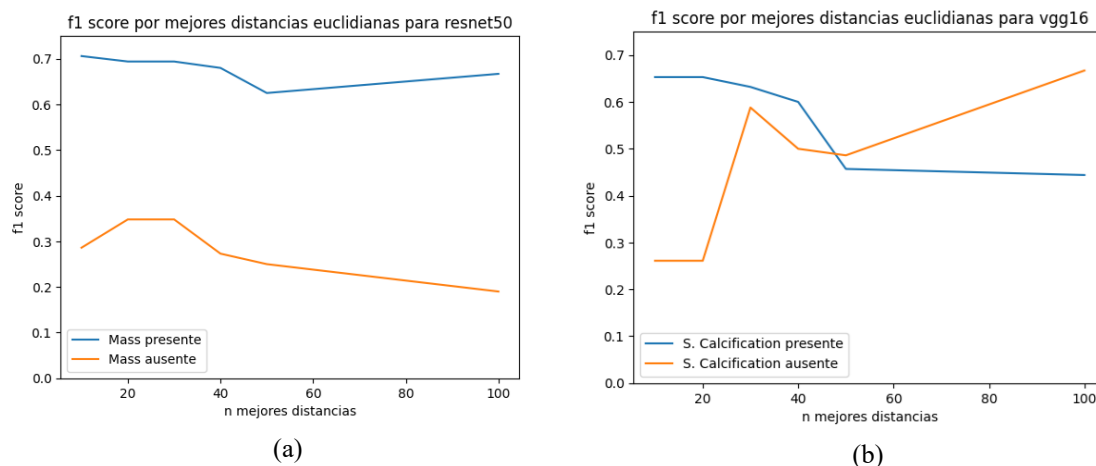


Figura 12. F1 Score en función de las n mejores distancias para modelos ResNet50 (a) VGG16 (b). Fuente propia.

3.4 Recomendar a partir de histogramas

Con las base de datos definidas, se hicieron búsquedas de imágenes por similitud por histogramas. Para esto se usaron los siguientes métodos de comparación: Correlación, Chi-Cuadrado, Intersección, Bhattacharya [47]. La Tabla 6, muestra los métodos y la transformación aplicada a cada una para poder usar de manera conjunta, obteniendo un nuevo rango de valores. Adicionalmente se observa el peso o ponderación que tendrá cada método para el cálculo de similitud, 25% cada una.

Tabla 6. Métodos usados para determinar similitud por histogramas y transformación aplicada para su uso conjunto. Fuente propia.

Método	Peso	Coincidencia exacta	Desajuste	Rango de valores	Transformación	Nuevo rango de valores
Correlación	25%	-1	1	$[-1, 1]$	$[-1, 1] + 1$	$[0, 2]$ 2, coincidencia exacta
Intersección	25%	0	1	$[0, 1]$	$[0, 1] * 2$	
Bhattacharya	25%	1	0	$[1, 0]$	$[1, 0] - 1 = [0, -1] * (-2)$	
Chi-Cuadrado	25%	0	2	$[0, 2]$	$[0, 2] - 2 = [-2, 0] * (-1)$	$[2, 0]$ 2, coincidencia exacta

En la Figura 13, se puede apreciar el proceso de cálculo de similitud de imágenes por histograma. A cada comparación de un histograma con otro se le calculo un vector, compuesto de 4 elementos, a saber el cálculo hecho con cada uno de los métodos mencionados. El vector calculado, se utilizó para calcular similitud contra el valor ideal, $[2, 2, 2, 2]$ (histogramas son iguales) a partir de la métrica “Distancia Euclidiana”. De esta manera se obtuvo un valor decimal, que representa la similitud entre histogramas. Cada imagen (query) de cada base de datos, se le calculo la similitud por histograma contra el resto de las imágenes de la base de datos, obteniendo las 500 distancias más cortas.

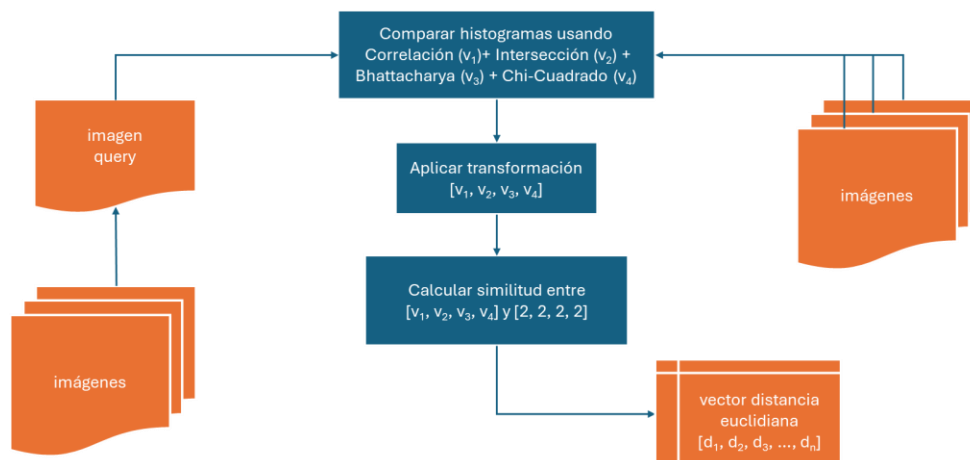


Figura 13. Proceso de cálculo de similitud de imágenes por histograma. Fuente propia.

En la siguiente figura (Figura 14), se muestra para una imagen query que contiene hallazgo Mass, las 5 imágenes más parecidas (menor distancia), junto con los histogramas asociados. Se puede evidenciar que a pesar de que la imagen query tiene lateralidad ‘R’ (derecha), dentro de las 5 mejores aparecen imágenes con lateralidad ‘L’ (izquierda).

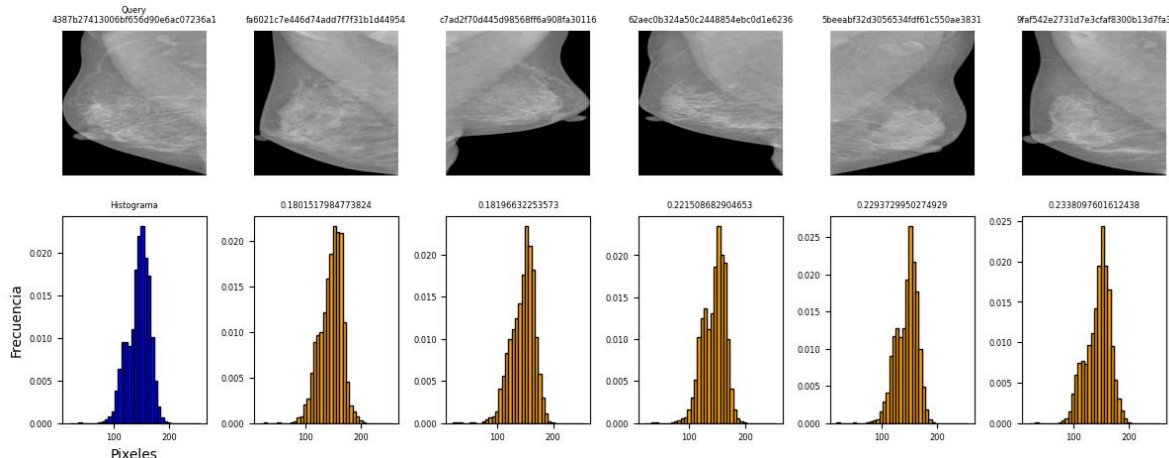


Figura 14. Búsqueda de imagen por histograma para una imagen query que contiene hallazgo Mass. Fuente propia.

3.5 Recomendar a partir de modelos

La recomendación a partir de modelos y teniendo como base una imagen de mama se lleva a cabo en dos pasos:
a) Clasificar la imagen y b) Buscar por similitud por histograma.

3.5.1 Determinar clasificación de imagen

Para evaluar el comportamiento del sistema en un entorno, se diseñó un flujo de clasificación (véase Figura 15).

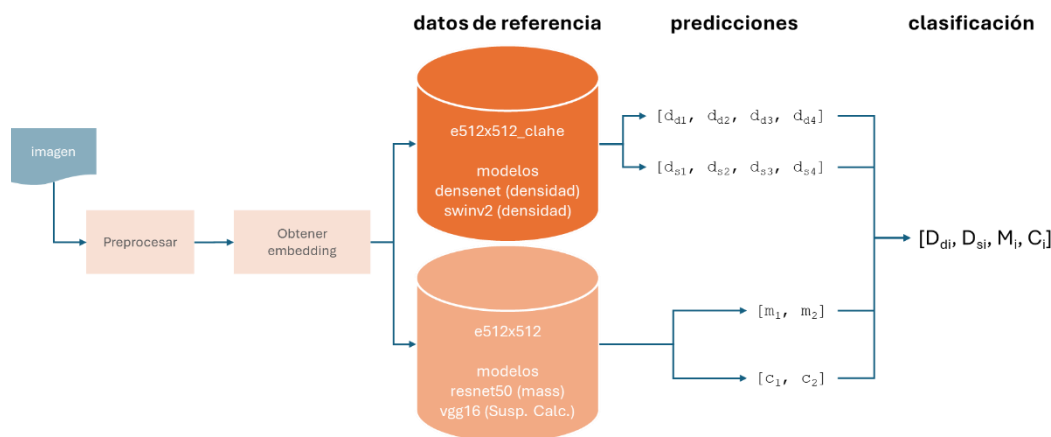


Figura 15. Flujo de clasificación de una imagen de mama. Fuente propia.

El flujo inicia con el preprocesamiento de la imagen, seguido de la extracción del *embedding* para generar las predicciones de densidad y hallazgos mediante los modelos seleccionados: **DenseNet**, **SwinV2**, **ResNet50** y **VGG16**.

La clasificación final se determina mediante el valor de máxima probabilidad en cada vector de predicción, obteniendo $[D_{di}, D_{si}, M_i, C_i]$, donde:

- D_{di} , corresponde a la densidad con modelo DenseNet
- D_{si} , corresponde a la densidad con modelo SwinV2
- M_i , corresponde a Mass con modelo ResNet50
- C_i , corresponde a Suspicious Calcification con modelo VGG16

A partir de un set de datos de validación, correspondiente a 12 mamografías, distintas a las mencionadas anteriormente, se procedió a corroborar el flujo de clasificación y sus resultados (Tabla 7 y 8). En la Tabla 7 se observa que las primeras 4 imágenes no se sabe cuál es la clasificación de densidad (columna “densidad real”). La columna “calce con valores reales”, muestra para cada modelo, si clasificación hecha es igual al valor real, 1, o no, 0. Así, los modelos en un 50% de los casos concuerdan en la clasificación de densidad C. Los calces entre densidad real y la clasificación que hacen los modelos es un 50% en modelo DenseNet y un 25% en modelo SwinV2. Adicionalmente se observa que los no calce se producen en una categoría de densidad vecina.

Tabla 7. Resultado de clasificación de imágenes con modelos DenseNet y SwinV2. Fuente propia.

densidad real	imagen	clasificación		calce con valores reales	
		DenseNet	SwinV2	DenseNet	SwinV2
-	1-026.png	C	C	-	-
-	1-171.png	C	D	-	-
-	1-204.png	C	C	-	-
-	1-296.png	C	D	-	-
C	282402171.png	B	B	0	0
B	440655856.png	A	A	0	0
D	441095399.png	D	C	1	0
A	560056097.png	A	B	1	0
B	1251063854.png	A	A	0	0

C	1547668985.png	C	C	1	1
A	1618424977.png	A	A	1	1
D	1976472767.png	C	C	0	0

Con respecto a la Tabla 8, la mayoría de las imágenes no se sabe si tienen hallazgo de Mass y/o Suspicious Calcification. La clasificación con el modelo VGG16 calza en un 100% con los valores reales. No así el modelo ResNet50 que calza 0%.

Tabla 8. Resultado de clasificación de imágenes con modelos ResNet50 y VGG16, para los hallazgos mass y suspicious calcification, respectivamente. Fuente propia.

valor real		imagen	clasificación		calce con valores reales	
mass	calcification		ResNet50 mass	VGG16 susp. calc.	ResNet50 mass	VGG16 susp. calc.
-	presente	1-026.png	ausente	presente	-	1
presente	-	1-171.png	ausente	presente	0	-
-	presente	1-204.png	presente	presente	-	1
-	presente	1-296.png	presente	presente	-	1
-	-	282402171.png	presente	ausente	-	-
-	-	440655856.png	ausente	ausente	-	-
-	-	441095399.png	presente	ausente	-	-
-	-	560056097.png	ausente	ausente	-	-
-	-	1251063854.png	ausente	ausente	-	-
-	-	1547668985.png	presente	presente	-	-
-	-	1618424977.png	presente	presente	-	-
-	-	1976472767.png	ausente	presente	-	-

3.5.2 Determinar imágenes a recomendar

Una vez obtenida la clasificación, se procedió a recomendar un caso (corresponde a un examen de un paciente que está constituido por cuatro imágenes [17]). La recomendación se hizo en base al contenido, a través del histograma de la imagen. Específicamente se buscó similitud por histograma, usando el mismo método indicado en sección “3.4 Recomendar a partir de histogramas”. A partir de las imágenes más parecidas en termino de histograma, se seleccionaron las 500 imágenes con menor distancia euclidiana y en este conjunto se buscó la primera imagen que contuviera la característica densidad o hallazgo entregada por cada uno de los modelos analizados. La Tabla 9 muestra el resultado de proceso de recomendación, para el conjunto de imágenes correspondiente a columna “imagen”, donde RC indica que se realiza la recomendación de un caso, es decir, se encuentra una imagen que contiene la característica de densidad/hallazgo, dentro de las primeras 500 imágenes con menor distancia euclidiana, en la búsqueda basada en histograma. NEI, no puede recomendar un caso, ya que no se encuentra una imagen que contenga la característica de densidad/hallazgo, dentro de las primeras 500 imágenes con menor distancia euclidiana, en la búsqueda basada en histograma. ISC, no se puede recomendar un caso, ya que la clasificación obtenida indica que característica está ausente.

Para abordar ciertos casos surgidos de la determinación de imágenes a recomendar, se definieron los siguientes criterios: para la densidad, en caso de NEI, se recomendará a partir de la siguiente categoría de densidad mamaria más alta. Cuando la recomendación de los modelos DenseNet y SwinV2, sean distintas prevalecerá la recomendación que haga DenseNet, es decir, se anula la recomendación de modelo SwinV2. Esta decisión se justifica, porque modelo DenseNet, tiene mejor rendimiento que modelo SwinV2.

Tabla 9. Resultado de recomendación de imágenes. RC: recomienda caso; NEI: no encontró imagen similar; ISC: imagen con clasificación 'ausente'. Fuente: propia.

imagen	clasificación				caso(imágenes) recomendado			
	DenseNet densidad	SwinV2 densidad	ResNet50 mass	VGG16 susp. calc.	DenseNet densidad	SwinV2 densidad	ResNet50 mass	VGG16 susp. calc.
1-026.png	C	C	ausente	presente	RC	RC	ISC	RC
1-171.png	C	D	ausente	presente	RC	RC	ISC	RC
1-204.png	C	C	presente	presente	RC	RC	RC	RC
1-296.png	C	D	presente	presente	RC	RC	RC	RC
282402171.png	B	B	presente	ausente	RC	RC	RC	ISC
440655856.png	A	A	ausente	ausente	RC	RC	ISC	ISC
441095399.png	D	C	presente	ausente	RC	RC	RC	ISC
560056097.png	A	B	ausente	ausente	RC	RC	ISC	ISC
1251063854.png	A	A	ausente	ausente	RC	RC	ISC	ISC
1547668985.png	C	C	presente	presente	RC	RC	RC	RC
1618424977.png	A	A	presente	presente	RC	RC	RC	RC
1976472767.png	C	C	ausente	presente	RC	RC	ISC	RC

4 Discusión

La discusión de los resultados obtenidos en este trabajo se centra en tres aspectos fundamentales: el impacto del preprocesamiento de imágenes en la calidad de los datos utilizados, el desempeño de los modelos de Deep Learning trabajados y el análisis de los resultados finales en términos de clasificación.

4.1 Preprocesamiento

Si bien la búsqueda de similitud basada en histogramas constituye una técnica establecida, esta puede resultar insuficiente para la identificación exhaustiva de imágenes análogas en el dominio radiológico. Tal como sostienen diversos autores, este enfoque carece de la información espacial necesaria para localizar con precisión las lesiones [48] [49]. Un ejemplo crítico de esta limitación se observa en la detección de asimetrías en mamografías bilaterales, un indicador clínico esencial cuya omisión podría derivar en la pérdida de hallazgos patológicos significativos si se depende exclusivamente del análisis de histogramas.

Por otro lado, el preprocesamiento de imágenes mamográficas es fundamental para optimizar la detección de anomalías, dado que mejora la calidad visual, reduce el ruido espurio y facilita la extracción de características mediante la eliminación de artefactos [50]. Estas técnicas son particularmente eficaces en la identificación de microcalcificaciones, masas y distorsiones arquitectónicas [51]. En el presente estudio, se implementó el algoritmo de ecualización adaptativa de histogramas con limitación de contraste (CLAHE) [44], alineándose con la metodología empleada en investigaciones previas orientadas a la mejora del diagnóstico mamográfico [52] [53] [54].

No obstante, al analizar la correlación óptima entre arquitectura y conjunto de datos (véase sección 3.2 Evaluación de Desempeño Modelo/Dataset), se observó una divergencia relevante en los resultados: mientras que para la clasificación de la densidad mamaria el dataset procesado con CLAHE exhibió el mejor desempeño, para la identificación de Mass y Suspicious Calcification el rendimiento superior se obtuvo mediante el uso de enmascaramiento sin aplicación de CLAHE. Este último hallazgo presenta una contradicción con la literatura convencional, la cual sugiere de manera generalizada que el uso de CLAHE incrementa la precisión en la clasificación de hallazgos patológicos [50] [51].

4.2 Desempeño

Al calcular la inferencia de los modelos utilizando el repositorio de referencia (sección 3.3.2 Calcular la inferencia de los modelos con datos de referencia), se identificaron los siguientes comportamientos: en la clasificación de densidad, el modelo DenseNet exhibió un desempeño superior en comparación con SwinV2. No obstante, el nivel de precisión (accuracy) de DenseNet no superó los estándares reportados en investigaciones previas, a pesar de utilizar la misma fuente de datos [38] [17]. Esta divergencia podría atribuirse a variaciones específicas en las técnicas de preprocesamiento de imágenes aplicadas en este estudio. En la clasificación de Mass y Suspicious Calcification, las arquitecturas ResNet50 y VGG16 mostraron un rendimiento limitado.

Tras evaluar el rendimiento en el conjunto de datos de test (sección 3.3.3 A partir de embeddings establecer rendimiento de modelos con datos de test), se consolidaron las siguientes observaciones: la consistencia en densidad, ratificando la superioridad de DenseNet sobre SwinV2 para la categorización de densidad mamaria. Limitaciones de SwinV2, el modelo presentó una tendencia de F1-Score inferior a 0.5 en la mayoría de las clases, lo que indica un desempeño insuficiente para esta tarea específica.

En el contexto de este trabajo, las arquitecturas ResNet50 y VGG16, trabajaron sobre hallazgos distintos, lo que invalida una comparación directa entre ambos. Sin embargo, su rendimiento resultó consistente con los resultados de inferencia obtenidos previamente. Es relevante notar que los pesos de las redes ResNet50 y VGG16 fueron preentrenados originalmente con la base de datos *ImageNet* [14]. Existe una brecha semántica significativa entre las fotografías cotidianas de ImageNet (objetos con texturas gruesas) y las mamografías, las cuales exigen la detección de estructuras más sutiles, como microcalcificaciones o distorsiones arquitectónicas finas.

Adicionalmente, los datos de test para los hallazgos Mass y Suspicious Calcification de los modelos ResNet50 y VGG16 respectivamente, son de tamaño reducido y no fueron fragmentados mediante una estrategia de partición estratificada. Esto aumenta la probabilidad de que la medición tenga sesgos.

En cuanto al bajo desempeño del modelo SwinV2, a pesar de haber sido entrenado con imágenes médicas, se postulan diversos factores determinantes reportados en la literatura [55] [56]: Sesgo en la muestra, el entrenamiento pudo verse afectado por una selección aleatoria de datos en lugar de una distribución equilibrada de las categorías de densidad. Heterogeneidad de formatos, la degradación del rendimiento suele ocurrir cuando los modelos se enfrentan a formatos digitales de mamografía distintos a los presentes en el set de entrenamiento. Integridad de las etiquetas, la calidad de la clasificación está intrínsecamente ligada a la precisión del etiquetado inicial; la presencia de imágenes con anotaciones imprecisas puede comprometer la capacidad de generalización del modelo.

El desempeño de SwinV2, contrasta con algunos trabajos de referencia, como el [57], que utilizó un modelo base de Swin Transformer [58] para la clasificación de la densidad mamaria en las cuatro categorías, logrando una precisión (accuracy) general del 74,96 %.

4.3 Resultados finales

Al contrastar las métricas consolidadas, se identifica que el modelo DenseNet presenta el rendimiento más robusto y consistente, alcanzando un accuracy de 0.65. No obstante, se observa una excepción crítica en la clase C (mamas heterogéneamente densas), donde el F1-Score desciende a 0.430. Este fenómeno revela una dicotomía en el comportamiento del modelo: aunque mantiene una Precisión aceptable (0.548), presenta un Recall deficiente (0.354). En términos clínicos, esto implica que el sistema omite una proporción significativa de casos reales, un resultado que se sitúa por debajo de los estándares establecidos en el estado del arte [38].

Por otro lado, los modelos ResNet50 y VGG16 (modelos sin reentrenar) no lograron detectar eficazmente los hallazgos de Mass y Suspicious Calcification, respectivamente. El resultado obtenido con estos modelos contrasta con algunas experiencias mencionadas en el estado del arte [11] [15]. Una limitación similar se presentó en el modelo SwinV2, cuya capacidad para clasificar la densidad mamaria resultó insuficiente. Para optimizar estas arquitecturas en futuras iteraciones, se deben considerar las siguientes directrices diagnósticas basadas en las métricas obtenidas: Sensibilidad baja (Recall bajo), indica la necesidad de incrementar el volumen de muestras en las clases subrepresentadas para que el modelo aprenda a generalizar mejor los patrones visuales. Precisión baja, sugiere una confusión interclase; por lo tanto, es imperativo refinar las características extraídas (features) que permiten diferenciar morfológicamente una categoría de otra.

La lógica del proceso de recomendación es estructuralmente buena y no se observa que requiera modificaciones funcionales.

5 Conclusiones

Con el objeto de apoyar el proceso de aprendizaje de radiólogos, el presente trabajo se centró en desarrollar un enfoque para clasificar y recomendar mamografías de acuerdo con su densidad, Mass y Suspicious Calcification, de manera automática mediante modelos de Deep learning. Por medio de técnica de preprocesamiento de imágenes, como enmascaramiento y CLAHE, se buscó resaltar detalles y mejorar su calidad para que los modelos tuvieran un buen rendimiento. En este contexto, se trabajó con distintas CNN: ResNet50, VGG16, DenseNet y SwinV2.

Los resultados, medidos a través de métricas de rendimiento como la accuracy, Precision, Recall y F1-Score, evidencian que este enfoque: el uso de modelos sin reentrenar, ResNet50 y VGG16, no es una estrategia adecuada, pues los hallazgos en mamografías no son detectados. Esto se evidencia al usar modelos reentrenados, DenseNet y SwinV2; que a pesar de tener resultados dispares, sus rendimientos son mejores que los alcanzados por ResNet50 y VGG16.

Como se indica en los resultados, el modelo preentrenado en mamografías DenseNet, a pesar de que logra el mejor rendimiento, este no supera los estándares reportados en investigaciones previas. En consecuencia la hipótesis planteada para este trabajo debe ser desestimada, dado que no se logró la precisión esperada.

Desde la perspectiva de diseño funcional del sistema SAERM, se observa que no influye directamente en los resultados de las métricas. Por lo tanto, las mejoras del diseño de SAERM, se deben orientar a: contar con un repositorio de referencia más extenso y diverso para garantizar la cobertura más amplia de variantes de mamografías y la optimización y reentrenamiento de los modelos de Deep Learning subyacentes.

5.1 Trabajo futuro

Este trabajo abre diversas oportunidades para trabajos futuros:

Ampliación del conjunto de datos. Incorporar nuevas bases de datos provenientes de diferentes regiones o poblaciones podría mejorar la robustez y generalización del modelo, con el objeto de que sea aplicable en un contexto clínico.

Optimización y reentrenamiento de los modelos. Junto con lo anterior, es relevante en el enfoque lograr mejorar las métricas, con el objeto de que pueda tener una validación clínica. Para esto, los modelos pueden mejorar ya sea haciendo modificaciones a su arquitectura o aplicando optimización mediante aprendizaje por transferencia y usando estrategias de ajuste fino.



Complementar enfoque propuesto haciendo Ensemble Learning, donde los resultados de varios modelos de CNN se pueden combinar para obtener el resultado.

Estos pasos fortalecerán el enfoque desarrollado, y serán un aporte al desarrollo de la inteligencia artificial en la medicina, apoyando el diagnóstico y tratamiento del cáncer de mama.

Agradecimientos.

Quiero agradecer en primer término a mi familia, María Eugenia mi esposa y mi hijo Vicente, por su paciencia en este camino recorrido. A Julio Sotelo, por su generosidad y sus constantes y valiosas observaciones, con las que ahora puedo ser mejor. A Angel (sin acento) e Iván, por haber formado un equipo insuperable y su amistad. Y a mis compañeros de trabajo por el aporte y colaboración en el desarrollo de este trabajo.

6 Referencias

- [1] M. T. Duong , A. M. Rauschecker , J. D. Rudie, P.-H. Chen, T. S. Cook, R. N. Bryan y S. Mohan, «Artificial intelligence for precision education in radiology,» *The British journal of radiology*, vol. 92, nº 1103, 2019.
- [2] L. Kathleen L., «Pedagogical Approaches to Diagnostic Imaging Education: A Narrative Review of the Literature,» *Journal of Chiropractic Humanities*, vol. 22, nº 1, pp. 9-16, 2015.
- [3] H. Muying (Lucy), S. Ethan, C. Andrew y K. Benjamin YM, «Exploring the integration of artificial intelligence in radiology education: A scoping review,» *Current Problems in Diagnostic Radiology*, vol. 54, nº 3, pp. 332-338, 2025.
- [4] A. S. L. Aibar, I. G. M.S. López- Criado, J. G. M.A. Calderón y J. Fernández -Parra, «Clasificación radiológica y manejo de las lesiones mamarias,» *Clínica e Investigación en Ginecología y Obstetricia*, vol. 38, nº 4, pp. 19-162, 2011.
- [5] C. Yinlu, B. Guy y W. Dongfeng, «Estimating key parameters in periodic breast cancer screening—Application to the Canadian National Breast Screening Study data,» *Cancer Epidemiology*, vol. 34, nº 4, pp. 429-433, 2010.
- [6] D. L. Monticciolo, M. Sharp F., F. Sarah M., E. Peter R., N. Mary S., M. Linda, D. Stamatia, L. Jessica W.T., H. R. Edward y S. Dana, «Breast Cancer Screening Recommendations Inclusive of All Women at Average Risk: Update from the ACR and Society of Breast Imaging,» *American College of Radiology*, vol. 18, nº 9, pp. 1280-1288, 2021.
- [7] Q. Huang, X. Huang, L. Liu, Y. Lin, X. Long y X. Li, «A case-oriented web-based training system for breast cancer diagnosis,» *Computer Methods and Programs in Biomedicine*, vol. 156, nº 0169-2607, pp. 73-83, 2018.
- [8] S. Jain y J. Dhar, «Image based search engine using deep learning,» de *Tenth International Conference on Contemporary Computing (IC3)*, 2017.
- [9] P. Lakshmi, S. K.Shiva, S. V.Sai, P.Shivani, K. Medikonda Asha, M. Pandita y T. Manyam, «Efficient Image Based Similarity Search Using CNN-VGG19,» de *3rd World Conference on Communication & Computing (WCONF)*, 2025.
- [10] A. Wong, «Medium,» 16 diciembre 2017. [En línea]. Available: <https://medium.com/data-science/building-a-similar-images-finder-without-any-training-f69c0db900b5>. [Último acceso: mayo 2026].
- [11] A. Mbilinyi y H. Schuldt, «Cross-Modality Medical Image Retrieval with Deep Features,» de *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2020.
- [12] S. C. d. Radiología, «SOCHRADI,» 2024. [En línea]. Available: <https://sochradi.cl/>. [Último acceso: 05 agosto 2024].

- [13] M. Yusuke, M. Ryosuke, R. Kenji y K. Kentaro, «A search system of retrieving images of the similar products without the requirement of the training of feature extractor and the purchase history,» *Computer Science*, 2017.
- [14] S. U. P. U. Stanford Vision Lab, «ImageNet,» 2020. [En línea]. Available: <https://www.image-net.org/about.php>. [Último acceso: 07 02 2026].
- [15] B. Yaniv, D. Idit, W. Lior y G. Hayit, «Deep learning with non-medical training used for chest pathology identification,» de *SPIE Medical Imaging*, Orlando, Florida, United States, 2015.
- [16] N. Khosla y V. Venkataraman, «Building image-based shoe search using convolutional neural networks,» 2015. [En línea]. Available: https://cs231n.stanford.edu/reports/2015/pdfs/nealk_final_report.pdf. [Último acceso: Marzo 2026].
- [17] Hieu T. Nguyen, H. Q. Nguyen, H. H. Pham, K. Lam, L. T. Le, M. Dao y V. Vu, «VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography,» *medRxiv*, 2022.
- [18] A. C. o. Radiology, «ACR BI-RADS® Atlas Breast Imaging Reporting and Data System,» Reston, USA, 2013.
- [19] E. Ordonez Ramos, «Deep learning para la visión artificial e identificación del personal administrativo y docente de la Universidad Nacional Micaela Bastidas de Apurímac 2018,» Febrero 2020. [En línea]. Available: <http://repositorio.unap.edu.pe/handle/20.500.14082/13523>. [Último acceso: octubre 2024].
- [20] D. H. Hubel y T. N. Wiesel, «Receptive Fields of Single Neurons in the Cat's Striate Cortex,» *The Journal of Physiology*, n° 148, pp. 574-591, 1959.
- [21] A. Geron, *Aprende Machine Learning con Scikit-learn, Keras y TensorFlow*, Madrid: Ediciones Anaya Multimedia, 2019.
- [22] D. M. W. Powers, «Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation,» *International Journal of Machine Learning Technology*, pp. 37-63, 2011.
- [23] D. Shah, «Mean Average Precision (mAP) Explained: Everything You Need to Know,» mayo 2022. [En línea]. Available: <https://www.v7labs.com/blog/mean-average-precision>. [Último acceso: octubre 2024].
- [24] X.-Y. Wang, Y.-J. Yu y H.-Y. Yang, «An effective image retrieval scheme using color, texture and shape features,» *Computer Standards & Interfaces*, vol. 33, n° 1, pp. 59-68, 2011.
- [25] JI WAN, W. DAYONG, . C. H. H. STEVEN, W. PENGCHENG, Z. JIANKE y Z. YONGDONG, «Deep Learning for Content-Based Image Retrieval: A Comprehensive Study,» de *MM '14: Proceedings of the 22nd ACM international conference on Multimedia*, Florida, Orlando, USA., 2014.

- [26] R. Vishraj, S. Gupta y Sukhwinder Singh, «A comprehensive review of content-based image retrieval systems using deep learning and hand-crafted features in medical imaging: Research challenges and future directions,» *Computers and Electrical Engineering*, Vols. %1 de %2104, Part A, n° 0045-7906, 2022.
- [27] A. Majumd, J. Lal Sarkar y A. Majumder, *Artificial Intelligence and Data Science in Recommendation System: Current Trends, Technologies and Applications*, Singapore: Bentham Science Publishers Pte. Ltd., 2023.
- [28] T. Iqbal, M. Masud, B. Amin, C. Feely, M. Faherty, T. Jones, M. Tierney, A. Shahzad y P. Vazquez, «Towards integration of artificial intelligence into medical devices as a real-time recommender system for personalised healthcare: State-of-the-art and future prospects,» *Health Sciences Review*, vol. 10, n° 100150, 2024.
- [29] P. M. Chawan y S. D. Kalkar, «Recommendation System using Machine Learning Techniques,» *International Research Journal of Engineering and Technology*, vol. 09, n° 09, 2022.
- [30] B. Ojokoh, Y. Folajimi y F. Isinkaye, «Recommendation systems: Principles, methods and evaluation,» *Egyptian Informatics Journal*, vol. 16, n° 3, pp. 261-273, 2015.
- [31] S. Biswas, «Artificial intelligence in radiology education,» *International Journal of Radiology & Radiation Therapy*, vol. 10, n° I, p. 3, 2023.
- [32] M. Vickery, E. Lanser, K. M. Koch, D. Pierce y J. Budovec, «SCRAPS: Introducing a Student-Centered Resident-Administered PACS Simulator for Medical Student Radiology Education,» *Academic Radiology*, 2024.
- [33] J. Martí, C. Mata, X. Lladó y A. Oliver, «MAMODB: A WEB-BASED TOOL FOR TRAINING RADIOLOGISTS IN THE DIAGNOSIS OF DIGITAL MAMMOGRAPHY,» de *3rd International Conference on Education and New Learning Technologies*, Barcelona, 2011.
- [34] X. Pau y B. Imma, «A new e-learning platform for radiology education (RadEd),» *Computer Methods and Programs in Biomedicine*, vol. 126, n° 0169-2607, pp. 63-75, 2016.
- [35] P. Welter, T. Deserno y B. Fischer, «Towards case-based medical learning in radiological decision making using content-based image retrieval,» *BMC Med Inform Decis Mak*, 2011.
- [36] T. Lehmann, M. Güld, C. Thies, B. Plodowski, D. Keysers, B. Ott y H. Schubert, «IRMA--content-based image retrieval in medical applications,» *Stud Health Technol Inform.*, vol. 107, n° 15360931, p. 842-846, 2004.
- [37] MAMAI-DENSE, *Cuantificación automática de densidad mamaria usando Inteligencia Artificial, como herramienta de apoyo en pesquisa de cáncer mamario: mamAI-dens.*, U. d. Valparaíso, Ed., Modelo entrenado para el proyecto ANID - FONDEF IT23I0040 de la Universidad de Valparaíso., 2025.

- [38] N. A. DÍAZ LATORRE, «CUANTIFICACIÓN DE DENSIDAD MAMARIA MEDIANTE TÉCNICAS DE PROCESAMIENTO DE IMÁGENES Y APRENDIZAJE PROFUNDO,» Universidad Técnica Federico Santa María, Santiago, Chile, 2024.
- [39] D. Mellado, «MamAI-Dens. Cuantificación de Densidad Mamaria utilizando modelos de Aprendizaje Profundo,» Universidad de Valparaíso, Valparaíso, 2025.
- [40] S. Mathi, D. Deb, A. Chaudhuri, L. Joseph y S. Biswas, «A Review Of Convolutional Neural Networks For Medical Image Analysis: Trends And Future Directions,» *Journal of Neonatal Surgery*, vol. 14, p. 90–97, 2025.
- [41] T. Rajasenbagam y S. Jeyanthi, «Semantic Content-Based Image Retrieval System Using Deep Learning Model for Lung Cancer CT Images,» *Journal of Medical Imaging and Health Informatics*, vol. 11, n° 10, pp. 2675-2682, 2021.
- [42] M. Benaouali, M. Bentoumi, M. Abed, M. Mimi y A. Taleb-Ahmed, «A Study on CNN-Based and Handcrafted Extraction Methods With Machine Learning for Automated Classification of Breast Tumors from Ultrasound Images,» *ELCVIA*, vol. 23, n° 2, p. 85–104, 2024.
- [43] V. H. C. de Albuquerque, S. K. Jagatheesaperumal, C. S. Nascimento y D. S. da Silva, «Mammogram Image Enhancement Techniques for Online Breast Cancer Detection and Diagnosis,» *MDPI*, n° 22: 8818, 2022.
- [44] F. Zavala, «Medium,» 24 06 2024. [En línea]. Available: <https://medium.com/imagecraft/histogram-equalization-clahe-algorithm-8841d402fc76>. [Último acceso: 15 07 2025].
- [45] Radiological Society of North America (RSNA), «RSNA Screening Mammography Breast Cancer Detection AI Challenge,» 2023.
- [46] «CBIS-DDSM Curated Breast Imaging Subset of Digital Database for Screening Mammography,» 14 Septiembre 2017. [En línea]. Available: <https://www.cancerimagingarchive.net/collection/cbis-ddsm/>. [Último acceso: 20 Noviembre 2025].
- [47] R. Szeliski, *Computer Vision Algorithms and Applications*, Cham, Switzerland: Springer, 2022.
- [48] Z. Wang, J. Xin y Y. Huang, «A similarity measure method combining location feature for mammogram retrieval,» *Journal of X-Ray Science and Technology*, n° 26, pp. 553-571, 2018.
- [49] S. Kenshi, M. Yusuke, Y. Hidetake y M. Junji, «Usefulness of histogram analysis of spatial frequency components for exploring the similarity and bilateral asymmetry in mammograms,» *Medical Imaging*, 2015.
- [50] J. P. T., S. G. S. y P. S. D., «Comparative Analysis of Different Preprocessing Techniques for Digital Mammograms Towards Breast Cancer Detection,» de *6th International Conference on Advances in Science and Technology (ICAST)*, Mumbai, India, 2023.

- [51] T. Sushreeta y S. Tripti, «Unified Preprocessing and Enhancement Technique for Mammogram Images,» *Procedia Computer Science*, vol. 167, n° 1877-0509, pp. 285-292, 2020.
- [52] L.-S. M. A., H.-R. Y. M., C.-M. O. E. y B.-M. R., «Comparative Analysis of CLAHE and VOI LUT for Enhanced Mammogram Contrast in Tumor Detection,» de *21st International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, Mexico City, Mexico, 2024.
- [53] S. S H y A. K A, «Using FC-CLAHE-ADF to enhance digital mammograms for detecting breast lesions,» de *The 6th International Conference of Science, Technology, and Interdisciplinary Research 2020 (IC-STAR 2020)*, Bandar Lampung, Indonesia, 2020.
- [54] C. P. Cunha, D. C. Lemos, A. A. Oliveira y P. A. C., «CLAHE Parameters Effects on the Quantitative and Visual Assessment of Dense Breast Mammograms,» *IEEE Latin America Transactions*, vol. 17, n° 5, pp. 851-857, 2019.
- [55] C. Ken, B. Andrew L, B. Laura, P. Jay B., S. Praveer, A. Nishanth T., H. Katharina V., G. Nathan, S. Meesam, P. Etta D., T. Mike, C. Laura P., D. Keith J., A. Bibb, A. Sheela y K. Jayashree, «Multi-Institutional Assessment and Crowdsourcing Evaluation of Deep Learning for Automated Classification of Breast Density,» *Journal of the American College of Radiology*, vol. 17, pp. 1653-1662, 2020.
- [56] A. A. Mohamed, W. A. Berg, H. Peng, Y. Luo, R. C. Jankowitz y S. Wu, «A deep learning method for classifying mammographic breast density categories,» *Medical Physics*, vol. 45, n° 1, pp. 314-321, 2018.
- [57] T. Kuen-Jang, Y. Wei-Cheng, K. Cheng-Yi, L. Ming-Wei, H. Chao-Ming, C. Hung-Ying, Y. Cheng-Yu y H. Shaw-Hwa, «Mammographic Breast Composition Classification Using Swin Transformer Network,» *Sensors and Materials*, vol. 36, n° 5, pp. 1951-1957, 2024.
- [58] L. Ze, L. Yutong, C. Yue, H. Hu, W. Yixuan y Z. Zheng, «Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,» de *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 2021.

Anexo A. Tablas de datos

1. Promedio de ocurrencia de posición por densidad/hallazgo

DENSITY A			DENSITY B		
Modelo	Base de datos	Promedio	Modelo	Base de datos	Promedio
densenet	images_512x512	22,1304	densenet	images_512x512	5,6097
densenet	images_512x512_clahe	14,0253	densenet	images_512x512_clahe	4,7694
densenet	images_e512x512	20,5942	densenet	images_e512x512	5,4695
densenet	images_e512x512_clahe	13,6571	densenet	images_e512x512_clahe	5,2795
swinv2	images_512x512	31,6667	swinv2	images_512x512	5,6471
swinv2	images_512x512_clahe	29,1014	swinv2	images_512x512_clahe	5,6886
swinv2	images_e512x512	34,9464	swinv2	images_e512x512	6,3706
swinv2	images_e512x512_clahe	18,1684	swinv2	images_e512x512_clahe	6,3421
resnet50	images_512x512	40,5227	resnet50	images_512x512	8,8291
resnet50	images_512x512_clahe	34,5385	resnet50	images_512x512_clahe	6,4553
resnet50	images_e512x512	35,1346	resnet50	images_e512x512	7,6373
resnet50	images_e512x512_clahe	24,1750	resnet50	images_e512x512_clahe	6,0723
vgg16	images_512x512	41,2051	vgg16	images_512x512	9,4242
vgg16	images_512x512_clahe	35,8200	vgg16	images_512x512_clahe	8,4220
vgg16	images_e512x512	40,0556	vgg16	images_e512x512	7,8798
vgg16	images_e512x512_clahe	19,0448	vgg16	images_e512x512_clahe	6,9585

DENSITY C			DENSITY D		
Modelo	Base de datos	Promedio	Modelo	Base de datos	Promedio
densenet	images_512x512	1,2448	densenet	images_512x512	5,3835
densenet	images_512x512_clahe	1,2646	densenet	images_512x512_clahe	5,1306
densenet	images_e512x512	1,2496	densenet	images_e512x512	5,8581
densenet	images_e512x512_clahe	1,2523	densenet	images_e512x512_clahe	4,8976
swinv2	images_512x512	1,2152	swinv2	images_512x512	6,3807
swinv2	images_512x512_clahe	1,2256	swinv2	images_512x512_clahe	6,1226
swinv2	images_e512x512	1,2077	swinv2	images_e512x512	6,0789
swinv2	images_e512x512_clahe	1,2180	swinv2	images_e512x512_clahe	5,5511
resnet50	images_512x512	1,2614	resnet50	images_512x512	6,8863
resnet50	images_512x512_clahe	1,2437	resnet50	images_512x512_clahe	6,4956
resnet50	images_e512x512	1,2503	resnet50	images_e512x512	7,2012
resnet50	images_e512x512_clahe	1,2381	resnet50	images_e512x512_clahe	5,9511
vgg16	images_512x512	1,2726	vgg16	images_512x512	6,6570
vgg16	images_512x512_clahe	1,2601	vgg16	images_512x512_clahe	6,0415
vgg16	images_e512x512	1,2581	vgg16	images_e512x512	6,8544
vgg16	images_e512x512_clahe	1,2496	vgg16	images_e512x512_clahe	6,3493

Mass			Suspicious Calcification		
Modelo	Base de datos	Promedio	Modelo	Base de datos	Promedio
densenet	images_512x512	17,4230	densenet	images_512x512	26,4780
densenet	images_512x512_clahe	17,7266	densenet	images_512x512_clahe	27,2383
densenet	images_e512x512	16,7093	densenet	images_e512x512	25,9852
densenet	images_e512x512_clahe	17,1010	densenet	images_e512x512_clahe	24,5739
swinv2	images_512x512	18,9012	swinv2	images_512x512	26,7592
swinv2	images_512x512_clahe	17,4872	swinv2	images_512x512_clahe	28,4467
swinv2	images_e512x512	17,4247	swinv2	images_e512x512	29,0021
swinv2	images_e512x512_clahe	17,4830	swinv2	images_e512x512_clahe	29,1000
resnet50	images_512x512	16,6711	resnet50	images_512x512	24,8946
resnet50	images_512x512_clahe	16,9128	resnet50	images_512x512_clahe	28,0923
resnet50	images_e512x512	16,0894	resnet50	images_e512x512	26,9349
resnet50	images_e512x512_clahe	16,2130	resnet50	images_e512x512_clahe	25,7259
vgg16	images_512x512	17,0008	vgg16	images_512x512	23,4276
vgg16	images_512x512_clahe	16,5528	vgg16	images_512x512_clahe	24,4644
vgg16	images_e512x512	17,3972	vgg16	images_e512x512	22,4927
vgg16	images_e512x512_clahe	17,3230	vgg16	images_e512x512_clahe	26,6387

2. métricas de matrices de confusión calculadas

Modelo DENSENET					
n mejores distancias	densidad	precision	recall	f1-score	accuracy
10	A	0,678	1,000	0,808	0,625
	B	0,632	0,600	0,615	
	C	0,389	0,175	0,241	
	D	0,644	0,725	0,682	
20	A	0,656	1,000	0,792	0,631
	B	0,605	0,575	0,590	
	C	0,444	0,200	0,276	
	D	0,698	0,750	0,723	
30	A	0,678	1,000	0,808	0,644
	B	0,615	0,600	0,608	
	C	0,462	0,150	0,226	
	D	0,673	0,825	0,742	
40	A	0,667	1,000	0,800	0,637
	B	0,575	0,575	0,575	
	C	0,455	0,125	0,196	
	D	0,694	0,850	0,764	
50	A	0,650	0,975	0,780	0,613
	B	0,526	0,500	0,513	
	C	0,462	0,150	0,226	
	D	0,673	0,825	0,742	
100	A	0,633	0,950	0,760	0,637
	B	0,477	0,525	0,500	
	C	0,800	0,100	0,178	
	D	0,765	0,975	0,857	

Modelo SWINV2

n mejores distancias	densidad	precision	recall	f1-score	accuracy
10	A	0,593	0,400	0,478	0,469
	B	0,441	0,375	0,405	
	C	0,356	0,650	0,460	
	D	0,692	0,450	0,545	
20	A	0,600	0,225	0,327	0,344
	B	0,318	0,350	0,333	
	C	0,289	0,600	0,390	
	D	0,444	0,200	0,276	
30	A	0,222	0,050	0,082	0,325
	B	0,194	0,150	0,169	
	C	0,320	0,800	0,457	
	D	0,600	0,300	0,400	
40	A	0,222	0,050	0,082	0,325
	B	0,242	0,200	0,219	
	C	0,320	0,800	0,457	
	D	0,556	0,250	0,345	
50	A	0,222	0,050	0,082	0,319
	B	0,286	0,250	0,267	
	C	0,317	0,800	0,454	
	D	0,467	0,175	0,255	
100	A	0,286	0,100	0,148	0,319
	B	0,229	0,200	0,213	
	C	0,333	0,800	0,471	
	D	0,467	0,175	0,255	

Modelo RESNET50

n mejores distancias	hallazgo presente?	precision	recall	f1-score	accuracy
10	Si	0,545	1,000	0,706	0,583
	No	1,000	0,167	0,286	
20	Si	0,548	0,944	0,694	0,583
	No	0,800	0,222	0,348	
30	Si	0,548	0,944	0,694	0,583
	No	0,800	0,222	0,348	
40	Si	0,531	0,944	0,680	0,556
	No	0,750	0,167	0,273	
50	Si	0,500	0,833	0,625	0,500
	No	0,500	0,167	0,250	
100	Si	0,515	0,944	0,667	0,528
	No	0,667	0,111	0,190	

Modelo VGG16

n mejores distancias	hallazgo presente?	precision	recall	f1-score	accuracy
10	Si	0,516	0,889	0,653	0,528
	No	0,600	0,167	0,261	
20	Si	0,516	0,889	0,653	0,528
	No	0,600	0,167	0,261	
30	Si	0,600	0,667	0,632	0,611
	No	0,625	0,556	0,588	
40	Si	0,545	0,667	0,600	0,556
	No	0,571	0,444	0,500	
50	Si	0,471	0,444	0,457	0,472
	No	0,474	0,500	0,486	
100	Si	0,667	0,333	0,444	0,583
	No	0,556	0,833	0,667	

3. Predicciones y clasificación de densidad mamaria

densidad real	imagen	predicción de densidad								clasificación	
		usando densenet				usando swinv2				densenet	swinv2
		A	B	C	D	A	B	C	D		
-	1-026.png	0,033	0,267	0,400	0,300	0,000	0,100	0,600	0,300	C	C
-	1-171.png	0,000	0,000	0,533	0,467	0,000	0,000	0,100	0,900	C	D
-	1-204.png	0,000	0,100	0,467	0,433	0,000	0,000	0,600	0,400	C	C
-	1-296.png	0,000	0,033	0,600	0,367	0,000	0,000	0,100	0,900	C	D
C	282402171.png	0,167	0,600	0,233	0,000	0,000	0,500	0,400	0,100	B	B
B	440655856.png	0,567	0,433	0,000	0,000	0,400	0,400	0,200	0,000	A	A
D	441095399.png	0,267	0,200	0,233	0,300	0,000	0,200	0,700	0,100	D	C
A	560056097.png	0,533	0,467	0,000	0,000	0,300	0,700	0,000	0,000	A	B
B	1251063854.png	0,700	0,300	0,000	0,000	0,500	0,300	0,200	0,000	A	A
C	1547668985.png	0,033	0,000	0,667	0,300	0,000	0,200	0,700	0,100	C	C
A	1618424977.png	0,600	0,367	0,033	0,000	0,500	0,400	0,100	0,000	A	A
D	1976472767.png	0,000	0,000	0,533	0,467	0,000	0,000	0,500	0,500	C	C

4. Predicciones y clasificación de hallazgos de mass y suspicious calcification

valor real			predicciones				clasificación	
			mass usando resnet50		susp. calc. usando vgg16		resnet50	vgg16
mass	calcification	imagen	presente	ausente	presente	ausente	mass	susp. calc.
-	presente	1-026.png	0,4600	0,5400	0,5400	0,4600	ausente	presente
presente	-	1-171.png	0,4800	0,5200	0,5400	0,4600	ausente	presente
-	presente	1-204.png	0,5800	0,4200	0,5200	0,4800	presente	presente
-	presente	1-296.png	0,5200	0,4800	0,5500	0,4500	presente	presente.
-	-	282402171.png	0,6200	0,3800	0,4100	0,5900	presente	ausente
-	-	440655856.png	0,4200	0,5800	0,4800	0,5200	ausente	ausente
-	-	441095399.png	0,5000	0,5000	0,4800	0,5200	presente	ausente
-	-	560056097.png	0,4000	0,6000	0,4800	0,5200	ausente	ausente
-	-	1251063854.png	0,4000	0,6000	0,4600	0,5400	ausente	ausente
-	-	1547668985.png	0,5400	0,4600	0,5200	0,4800	presente	presente
-	-	1618424977.png	0,5800	0,4200	0,6100	0,3900	presente	presente
-	-	1976472767.png	0,4600	0,5400	0,5200	0,4800	ausente	presente

5. Recomendaciones para densidad

imagen	clasificación		caso(imágenes) recomendado	
	densenet densidad	swinv2 densidad	densenet densidad	swinv2 densidad
1-026.png	C	C	67060319d052feb6d9c1d1784c174762	67060319d052feb6d9c1d1784c174762
1-171.png	C	D	5214f4386efcea56014510f1068da2a0	ee3551c7153c8549f6b423de9a7ae3c0
1-204.png	C	C	615db0ed3cdb232b21f04aa1f5df423a	615db0ed3cdb232b21f04aa1f5df423a
1-296.png	C	D	615db0ed3cdb232b21f04aa1f5df423a	85770f0bf3769aa63435c48de36443db
282402171.png	B	B	dc8eca8e41cb72804147da33e91917bb	dc8eca8e41cb72804147da33e91917bb
440655856.png	A	A	no encontró imagen similar	no encontró imagen similar
441095399.png	D	C	5fd8794c237e6dc35e1c1222c5edde59	615db0ed3cdb232b21f04aa1f5df423a
560056097.png	A	B	no encontró imagen similar	dc8eca8e41cb72804147da33e91917bb
1251063854.png	A	A	no encontró imagen similar	no encontró imagen similar
1547668985.png	C	C	6a7796ead3ac8071d7d1b04f1a09076	6a7796ead3ac8071d7d1b04f1a09076
1618424977.png	A	A	no encontró imagen similar	no encontró imagen similar
1976472767.png	C	C	628a6d10970a6b5c7032da06ba371b6c	628a6d10970a6b5c7032da06ba371b6c

6. Recomendaciones para hallazgos de mass y suspicious calcification

imagen	clasificación		caso(imágenes) recomendado	
	resnet50 mass	vgg16 susp. calc.	resnet50 mass	vgg16 susp. calc.
1-026.png	ausente	presente	imagen con clasificación 'ausente'	02bd0bd83c6d9fedc49b0df6ecd952c6
1-171.png	ausente	presente	imagen con clasificación 'ausente'	8dd986e8174fc84c2a984e66a416ff9f
1-204.png	presente	presente	da6ce640b22100d327694fac20224427	02bd0bd83c6d9fedc49b0df6ecd952c6
1-296.png	presente	presente	540ab71b28752c3e279be9bad638df79	02bd0bd83c6d9fedc49b0df6ecd952c6
282402171.png	presente	ausente	7eab495d8ae42610e89e53788db3e019	imagen con clasificación 'ausente'
440655856.png	ausente	ausente	imagen con clasificación 'ausente'	imagen con clasificación 'ausente'
441095399.png	presente	ausente	7eab495d8ae42610e89e53788db3e019	imagen con clasificación 'ausente'
560056097.png	ausente	ausente	imagen con clasificación 'ausente'	imagen con clasificación 'ausente'
1251063854.png	ausente	ausente	imagen con clasificación 'ausente'	imagen con clasificación 'ausente'
1547668985.png	presente	presente	7364e99bbbf25fca6f398ca2110b55b	93514476602890f55d1758d1f3e1a9b5
1618424977.png	presente	presente	60279474726275dcc024f3c67b4d51a3	60279474726275dcc024f3c67b4d51a3
1976472767.png	ausente	presente	imagen con clasificación 'ausente'	bba470c7ac302b5e4bfd2ed1a4515661