



**Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria  
en imágenes de mamografía.**

Milka Mora Morales

Informe de práctica presentado para optar al título de Bioingeniero

Asesor

Jazmín Ximena Suarez Revelo, Doctora (PhD) en Ingeniería Electrónica y de Computación  
Isabella Ariza Cuberos, Bioingeniera, Universidad de Antioquia

Universidad de Antioquia  
Facultad de ingeniería  
Bioingeniería  
Medellín, Antioquia, Colombia  
2024

Cita	Mora Morales [1]
<b>Referencia</b> Estilo IEEE (2020)	[1] M. Mora Morales, “Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.”, Trabajo de grado profesional, Bioingeniería, Universidad de Antioquia, Medellín, Antioquia, Colombia, 2023.



Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - [www.udea.edu.co](http://www.udea.edu.co)

**Rector:** John Jairo Arboleda Céspedes.

**Decano/Director:** Julio César Saldarriaga Molina.

**Jefe departamento:** John Fredy Ochoa Gómez.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

## **Dedicatoria**

A mi padre, por darme su apoyo incondicional en esta experiencia tan gratificante.

A mi madre, por siempre enseñarme a luchar para que se cumplan todos los anhelos de mi corazón.

A Michelle, por ser mi fuente de inspiración en este camino, mi apoyo y mi confidente día a día.

A Guelito y a Kale, quienes me motivan a ser el mejor ejemplo para ustedes y a inspirarlos a seguir sus propios sueños.

A Tata, porque siempre me diste un empujoncito para seguir estudiando cada día.

## **Agradecimientos**

A mi familia que ha estado presente en cada paso del camino.

Al profesor John Fredy Ochoa, por su guía y acompañamiento a lo largo de mi trayectoria académica.

A mis asesores Jazmín Ximena Suárez, Jon Edison Duque y Juan Gabriel Castrillón por brindarme su orientación invaluable.

A SURA, por darme la oportunidad de ser parte de este gran equipo de trabajo.

A mis amigos, que hicieron parte de este recorrido desde el primer día.

## TABLA DE CONTENIDO

RESUMEN.....	8
ABSTRACT .....	9
I. INTRODUCCIÓN .....	10
II. OBJETIVOS.....	13
A. Objetivo general .....	13
B. Objetivos específicos .....	13
III. MARCO TEÓRICO .....	14
IV. METODOLOGÍA .....	16
V. RESULTADOS .....	20
VI. ANÁLISIS .....	37
VII. CONCLUSIONES .....	42
REFERENCIAS .....	44

## LISTA DE TABLAS

Tabla 1. Distribución porcentual de la evaluación de la segmentación de los datos evaluados. ...	23
Tabla 2. Número de estudios con buena segmentación en cada categoría ACR. ....	26
Tabla 3. Estudios como casos y controles en cada categoría ACR.....	27
Tabla 4. Porcentaje de acierto del porcentaje de densidad en cada categoría. ....	29
Tabla 5. Porcentaje de acierto del porcentaje de densidad en cada categoría y los intervalos de confianza. ....	30
Tabla 6. Rango de porcentaje predicha por el modelo en cada clasificación ACR. ....	30
Tabla 7. Comparación por categoría del software de LIBRA y de MONAI. ....	35

## LISTA DE FIGURAS

Figura 1. Flujo de trabajo para lectura de MG .....	20
Figura 2. Lista lectura imágenes .....	21
Figura 3. Distribución porcentual de la evaluación de la segmentación de los datos evaluados. ...	22
Figura 4. Distribución porcentual de la evaluación de la segmentación de acuerdo con modelo de cada equipo.....	24
Figura 5. Distribución porcentual de la evaluación de la segmentación de acuerdo con el tipo de equipo.....	24
Figura 6. Distribución porcentual de la evaluación de la segmentación de acuerdo con la clasificación ACR. ....	25
Figura 7. Distribución porcentual de la evaluación de la segmentación de acuerdo la clasificación BIRADS. ....	26
Figura 8. Distribución porcentual de los casos y controles en cada categoría ACR de los estudios con buena segmentación.....	27
Figura 9. Distribución de la clasificación ACR en cada categoría.....	28
Figura 10. Matriz de confusión de la regresión logística con el porcentaje de densidad y la clasificación ACR. ....	28
Figura 11. Matriz de confusión eliminando el efecto del scanner. ....	29
Figura 12. A la izquierda el diagrama de cajas inicial y a la derecha el diagrama de cajas de las predicciones del modelo de regresión logística para cada clasificación ACR.....	31
Figura 13. Imagen del proceso de DeepLIBRA. A la derecha la imagen de la base de datos empleada y a la izquierda la imagen empleada por los autores de DeepLIBRA [26].....	33
Figura 14. Matriz de confusión para el modelo de MONAI con todos los 2098 datos.....	34
Figura 15. Comparación de matrices de confusión. A la izquierda la matriz obtenida del modelo LIBRA y a la derecha la matriz obtenida del modelo de MONAI.....	35
Figura 16. Porcentaje de las clasificaciones correctas e incorrectas de los scanners usados en MONAI. ....	36
Figura 17. Frecuencias de clasificaciones correctas e incorrectas de los scanners empleando MONAI. ....	36
Figura 18. Mapas de intensidad de las imágenes. ....	37

## SIGLAS, ACRÓNIMOS Y ABREVIATURAS

<b>ACR</b>	American college of radiology
<b>IA</b>	Inteligencia Artificial
<b>BI-RADS</b>	Breast imaging-reporting and data system
<b>CC</b>	Craneo-caudal
<b>MLO</b>	Medio-lateral oblicua
<b>MG</b>	Mamografía
<b>PD</b>	Porcentaje de densidad
<b>dCNN</b>	Red neuronal convolucional profunda

---

RESUMEN

El proyecto se enfoca en abordar el cáncer de mama como un problema de salud pública significativo a nivel mundial, con especial atención en Colombia. El cáncer de mama representa la principal causa de muerte por cáncer en mujeres, por lo que su detección temprana es fundamental. Mediante la evaluación de modelos de inteligencia artificial para calcular la densidad mamaria en mamografías, es posible centrarse en una prevención primaria que permita evaluar el riesgo individual de presentar cáncer de mama esporádico. Se propone evaluar el desempeño de la tecnología LIBRA y otro modelo de IA (MONAI) para estimar la densidad mamaria y calcular su relación con los valores de la clasificación ACR. Ambos modelos sirvieron como una propuesta de evaluación tanto cuantitativa como cualitativa del cálculo de la densidad mamaria, brindando mayor información al radiólogo y agilizando los tiempos del proceso de diagnóstico y la toma de decisiones. Los hallazgos y resultados de este proyecto ayudan a transformar el enfoque actual de atención y gestión del cáncer de mama, proporcionando estrategias más efectivas y proactivas para prevenir su aparición y mejorar los resultados en salud.

***Palabras clave — Cáncer de mama, mamografías, porcentaje de densidad mamaria, BI-RADS, clasificación ACR.***

---

ABSTRACT

The project focuses on addressing breast cancer as a significant public health problem worldwide, with special attention to Colombia. Breast cancer represents the leading cause of cancer death in women, so its early detection is critical. By evaluating artificial intelligence models to calculate breast density in mammograms, it is possible to focus on primary prevention to assess the individual risk of presenting sporadic breast cancer. We propose to evaluate the performance of LIBRA technology and another AI model (MONAI) to estimate breast density and calculate its relationship with ACR classification values. Both models served as a proposal for both quantitative and qualitative evaluation of breast density calculation, providing more information to the radiologist and speeding up the diagnostic process times and decision making. The findings and results of this project will help transform the current approach to breast cancer care and management, providing more effective and proactive strategies to prevent its occurrence and improve health outcomes.

***Keywords*** — *Breast cancer, mammograms, breast density percentage, BI-RADS, ACR classification.*

---

## I. INTRODUCCIÓN

El cáncer de mama representa la principal causa de muerte por cáncer en mujeres a nivel mundial, este tipo de cáncer es uno de los más frecuentes [1]. Según datos de la Organización Mundial de la Salud (OMS) en 2020, se diagnosticaron aproximadamente 2,3 millones de casos de cáncer de mama en todo el mundo, y 685,000 mujeres perdieron la vida debido a esta enfermedad [2]. En el mismo año, se reportaron más de 210.000 nuevos casos de cáncer de mama en América Latina y el Caribe, con casi 68.000 fallecimientos [1]. Estas cifras resaltan la importancia de abordar el cáncer de mama como un problema de salud pública significativo y la necesidad de tomar medidas preventivas y de detección temprana.

En Colombia, se presenta un riesgo intermedio de desarrollar cáncer de mama, con una tasa de incidencia estandarizada por edad de 44,1 por cada 100.000 mujeres [3]. A pesar de las recomendaciones nacionales de realizar tamizaje mamográfico en mujeres de 50 a 69 años, examen clínico de mama anual y detección temprana en mujeres sintomáticas sin importar la edad, lamentablemente no se ha logrado reducir la mortalidad por esta enfermedad en el país. En efecto, el cáncer de mama ocupa el primer lugar en incidencia entre las mujeres en Colombia. [4]

La detección temprana de signos de cáncer de mama, como microcalcificaciones y masas, se puede lograr mediante el uso de imágenes mamográficas de alta resolución [5]. Estas imágenes proporcionan una visualización detallada de la estructura interna de la mama, donde el tejido mamario aparece blanco debido a una mayor atenuación de los rayos X por parte del epitelio y el estroma mamario en comparación con la grasa circundante, que se presenta de manera más oscura [6]. La mamografía es un área potencial en la que el procesamiento avanzado de imágenes puede mejorar las probabilidades de detectar el cáncer de mama en forma temprana [7].

La proporción de tejido epitelial y conectivo presente en una mamografía determina el porcentaje de densidad mamaria (PD%), y se ha demostrado que un mayor porcentaje de dicha densidad está asociado con un mayor riesgo de desarrollar cáncer de mama. Esta relación, hace que el (PD%) se convierta en uno de los factores de riesgo más significativo para desarrollar este tipo de cáncer [8]. Existen múltiples enfoques para medir la densidad mamaria, siendo el sistema de

## **Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.**

---

clasificación BI-RADS el más comúnmente utilizado. Sin embargo, se ha observado que las mediciones asistidas por computadora proporcionan resultados más confiables y precisos en la evaluación de la densidad mamaria [9].

Según un estudio realizado, se encontró que la tasa de acuerdo en la categorización BI-RADS entre los médicos radiólogos del Instituto Nacional de Cancerología de Colombia y los médicos radiólogos de otras instituciones, fue del 54,9%. Además, el índice kappa ( $k$ ) obtenido fue de 0,21, lo cual indica una concordancia débil en la categorización BI-RADS [10]. Estos resultados sugieren una baja concordancia en los informes mamográficos lo cual resalta la variabilidad que puede existir en la valoración de cada observador al momento de interpretar las mamografías a pesar del sistema BI-RADS estandarizado.

Recientemente se han propuesto otros métodos para intentar cuantificar la densidad mamaria. Estas herramientas emplean algoritmos basados en la segmentación y umbralización de forma semiautomática (interactiva) y completamente automática (LIBRA, DeepLibra, Cumulus, etc) [10], [11], [12]. LIBRA (Laboratory for Individualized Breast Radiodensity Assessment) es un paquete de software desarrollado por el Computational Breast Imaging Group (CBIG) de la Universidad de Pensilvania. Es una solución de software de estimación de densidad mamaria completamente automática basada en un algoritmo publicado que funciona con imágenes de mamografía digital [12].

Actualmente en SURA, se está realizando un proyecto que consiste en la validación clínica de una prueba de estratificación imagenómica de riesgo en cáncer de mama esporádico, orientada al diseño de un modelo de servicio de prevención primaria personalizada en mujeres colombianas entre los 40 y 69 años afiliadas a SURA. Dentro de los objetivos de este proyecto, se encuentra el establecer adecuados valores de referencia de la densidad mamaria mediante el sistema automático LIBRA.

Se desarrolló un estudio retrospectivo de casos y controles de pacientes con estudios de MG realizados en la compañía. Dos médicas radiólogas especialistas en imagen mamaria realizaron una

## **Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.**

12

---

segunda revisión de las mamografías para confirmar los hallazgos y diagnósticos de las pacientes. Utilizaron los criterios de la 5ta edición BI-RADS de 2013 para establecer la densidad mamaria de forma individual. También, se determinó el porcentaje de densidad mediante el sistema automático LIBRA (Laboratory for Individualized Breast Radiodensity Assessment) [10]. Ambas radiólogas revisaron y evaluaron la segmentación realizada por el software. A partir de estos datos es posible establecer rangos de referencia para la densidad mamaria expresada en porcentaje en el sistema automático LIBRA.

Con el objetivo de mejorar la precisión en la evaluación del porcentaje de la densidad mamaria y abordar la variabilidad en la interpretación de las imágenes mamográficas, se llevó a cabo un proyecto de prácticas en el cual se apoyó la iniciativa actual que se está llevando a cabo en SURA y se estimó el desempeño del modelo de inteligencia artificial LIBRA para el cálculo del porcentaje de densidad. A partir de lecturas y valoraciones realizadas por especialistas, se calculó la relación entre dicho porcentaje arrojado por LIBRA y el rango de clasificación ACR y se evaluó la necesidad de implementar nuevos modelos de IA como MONAI.

Para abordar el proyecto y lograr dichos objetivos, se siguió una metodología en la cual se realizó una revisión bibliográfica, un análisis exploratorio y descriptivo de los datos existentes y se establecieron valores adecuados de referencia del porcentaje de densidad mamaria con el fin de validar el desempeño y utilidad del software LIBRA, esto fue reportado en un informe para su posterior entrega al Ministerio de Ciencia, Tecnología e Innovación y de igual forma, se evaluó otro modelo para realizar la estimación de densidad (MONAI).

---

## II. OBJETIVOS

### *A. Objetivo general*

Implementar modelos de inteligencia artificial para el cálculo de densidad mamaria a partir de imágenes de mamografía de pacientes SURA.

### *B. Objetivos específicos*

- Estimar el desempeño del modelo LIBRA implementado en la compañía a partir de las lecturas y valoraciones realizadas por los médicos radiólogos.
- Calcular la relación entre la densidad mamaria calculada por el modelo y la clasificación ACR realizada por los médicos radiólogos.
- Evaluar un nuevo modelo para la estimación de densidad mamaria a partir de imágenes de mamografía de una base de datos abierta.

---

### III. MARCO TEÓRICO

La mamografía es una técnica de imagen que utiliza rayos X para obtener una imagen detallada del tejido mamario [13]. Esta exploración radiográfica proporciona una mejor visualización de la estructura interna del seno. La sensibilidad de la mamografía de detección se ve afectada por la calidad de la imagen y el nivel de experiencia del radiólogo [14]. Aunque la mamografía sigue siendo el método principal de detección y diagnóstico de cáncer de mama, presenta limitaciones en casos de mayor densidad mamaria, lo que puede disminuir su sensibilidad para detectar lesiones pequeñas y visualizar el cáncer subyacente [15]. Otras de sus limitantes están asociadas a la subjetividad, la variabilidad entre evaluaciones y la alta tasa de resultados falsos positivos y falsos negativos [16].

El cáncer de mama es una enfermedad caracterizada por el crecimiento anormal y desordenado de células malignas en el tejido mamario. Esta proliferación celular descontrolada puede dar lugar a la formación de tumores en la mama, los cuales tienen la capacidad de invadir otros órganos del cuerpo [2]. La detección de este tipo de cáncer se puede hacer mediante un tamizaje, el cual es una estrategia utilizada para identificar tempranamente la presencia o ausencia de enfermedades en una población específica que se encuentran en buen estado de salud. Esta técnica permite una pronta gestión e intervención para reducir los efectos secundarios asociados, como el dolor y la mortalidad. El Ministerio de Salud y Protección Social de Colombia, indica que las tecnologías para la detección temprana de este cáncer incluidas en el Plan Obligatorio de Salud (POS) son el examen clínico de mama recomendado para mujeres mayores a 20 años y una mamografía para mujeres entre 50 y 69 años [17]. Para el tratamiento y diagnóstico también se incluye la rehabilitación, el cuidado paliativo y los servicios de apoyo social del cáncer de mama.

En la actualidad, SURA ofrece un programa denominado 'Tiempo para ti', el cual tiene como objetivo realizar un análisis completo y oportuno a los pacientes, con el fin de aumentar las posibilidades de cura del cáncer de seno mediante un diagnóstico integral [18].

El porcentaje de densidad mamaria (PD%) es la cantidad de tejido fibroglandular presente en la mama, estimada mediante mamografía. Se considera uno de los factores de riesgo más significativos para desarrollar cáncer de mama [19]. Para cuantificar esta densidad existen métodos muy subjetivos como un análisis visual. Sin embargo, existen métodos semiautomatizados para analizar y realizar una evaluación de las imágenes mamográficas la cual se puede realizar mediante la segmentación del tejido fibroglandular del tejido graso, lo que permite calcular el porcentaje de densidad mamaria [20].

Existen sistemas estandarizados para describir los resultados de las mamografías, entre ellos se encuentra el BI-RADS el cual no es un diagnóstico, sino que representa un lenguaje estándar entre radiólogos que permite categorizar los hallazgos y ver cómo proceder en el tratamiento. Tiene como objetivo estandarizar la terminología y la sistemática del informe

## Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.

15

---

mamográfico, categorizar las lesiones estableciendo el grado de sospecha y asignar la actitud a tomar en cada caso [21]:

- BI-RADS 0: Resultado no concluyente debido a una evaluación incompleta o a una mamografía mal tomada o de mala calidad. Se recomienda realizar estudios adicionales de imágenes o comparar con mamografías anteriores para obtener un diagnóstico más preciso.
- BI-RADS 1: Resultado negativo o mama sin hallazgos. Esta categoría se asigna cuando no se detectan anomalías mamográficas que cumplan con los criterios de las categorías 2 a 5 en términos de sospecha. Se sugiere realizar una mamografía de seguimiento en un plazo de dos años.
- BI-RADS 2: Hallazgo benigno. Hay nódulos y calcificaciones en el grado de sospecha típicamente benignas. Se recomienda una mamografía en dos años.
- BI-RADS 3: Hallazgo posiblemente benigno. Se aplica a un conjunto de lesiones mamarias que presentan características que no se consideran definitivamente benignas según los estándares establecidos. Se sugiere realizar un seguimiento a los 6 y 12 meses al igual que a los 2 años para establecer la benignidad.
- BI-RADS 4: Anomalía sospechosa probablemente maligna. Esta categoría presenta 3 subcategorías:
  - Categoría 4-A: Descubrimiento mamográfico que amerita biopsia, pero con una sospecha mínima de malignidad.
  - Categoría 4-B: Sospecha intermedia de malignidad.
  - Categoría 4-C: Inquietud moderada, pero no típica de malignidad como ocurre en la categoría 5. Para esta categoría se sugiere una derivación hospitalaria.
- BI-RADS 5: Anomalía altamente sospechosa de ser un hallazgo maligno, es decir, un posible cáncer de mama. Se puede presentar una morfología irregular al igual que microcalcificaciones y nódulos. Se sugiere una derivación hospitalaria.
- BI-RADS 6: Confirmación histológica de malignidad en resultados de biopsia. Esta categoría se utiliza cuando se han obtenido resultados de biopsia que demuestran la presencia de malignidad. Es importante destacar que esta categoría se asigna antes de iniciar el tratamiento y se basa en evidencia histológica sólida que confirma la presencia de cáncer.

La densidad mamaria es una medida de la proporción de tejido mamario en comparación con la grasa en los senos de una mujer [22]. Cuanto mayor es la proporción de tejido mamario, mayor es la densidad mamaria. Se ha observado que las mujeres con mayor densidad mamaria

tienen un mayor riesgo de desarrollar cáncer de mama [23]. Sin embargo, la presencia de tejido mamario denso puede dificultar la detección del cáncer mediante mamografías. Esto se debe a que tanto el tejido mamario denso como los tumores aparecen como áreas blancas en una mamografía, lo que puede dificultar la identificación de anomalías y aumentar la posibilidad de falsos negativos [24].

Esta densidad mamaria está estrechamente relacionada con la clasificación ACR (Categorías para clasificar la densidad de la mama) ya que, en la detección mamográfica, los informes generalmente se formulan según el catálogo o sistema ACR BI-RADS (American College of Radiology Breast Imaging-Reporting and Data System), el cual se actualizó por última vez en noviembre de 2015. En el sistema ACR BI-RADS, la densidad mamaria se clasifica en cuatro subcategorías [25]:

- A: Predominio de tejido adiposo. Se refiere a mamas que tienen una cantidad insignificante de tejido fibroglandular denso.
- B: Tejido glandular disperso. Son mamas que contienen algunas áreas pequeñas de tejido fibroglandular denso, pero la mayor parte de la mama es grasa.
- C: Predominio de tejido fibroglandular heterogéneo (Tejido heterogéneamente denso que pueden ocultar masas pequeñas). Son mamas con una cantidad moderada de tejido fibroglandular denso, lo que puede dificultar la visualización de masas pequeñas o anomalías en la mamografía.
- D: Mamas totalmente fibroglandulares (Tejido extremadamente denso que disminuyen la sensibilidad de la mamografía). En estas mamas, la densidad puede ser tan alta que dificulte la detección de masas o anomalías en la mamografía.

#### IV. METODOLOGÍA

En SURA se llevó a cabo un proyecto para la validación clínica de una prueba de estratificación imagenómica de riesgo en cáncer de mama esporádico. Para el desarrollo del proyecto, SURA alistó un equipo conformado por dos radiólogas especialistas en imagen mamaria, quienes revisaron los estudios tanto de los casos como de los controles. Cada radióloga estableció de forma individual la densidad mamaria bajo los criterios de la 5ta edición BI-RADS de 2013, donde se establece una denominación de la composición del tejido comprendido entre las letras, A hasta D. Los estudios se anonimizaron y se ejecutaron en el software LIBRA. Cada una de las especialistas revisó y evaluó la segmentación realizada por el software LIBRA, clasificándola en tres categorías: buena, regular y mala.

**1. Estimación del desempeño del modelo actual implementado en la compañía a partir de las lecturas y valoraciones realizadas por los médicos radiólogos.**

Se realizó una revisión bibliográfica centrada en la implementación de modelos de inteligencia artificial para calcular el porcentaje de densidad mamográfica y su relación con la clasificación ACR BI-RADS. Se analizaron las técnicas y enfoques que existen en la actualidad para abordar la problemática.

Se llevó a cabo un análisis exploratorio de los datos disponibles en SURA sobre los cuales se puede validar el software LIBRA teniendo en cuenta la evaluación de las radiólogas involucradas en el proyecto de la calidad de la segmentación realizada por el software y el tipo de scanner empleado en la mamografía.

Se realizaron análisis estadísticos y validaciones que permitieron evaluar el rendimiento del modelo en los diferentes equipos de mamografía y se compararon los resultados de las clasificaciones de los expertos con la clasificación del modelo. Adicionalmente, se realizó un análisis univariado, bivariado y multivariado de los datos con el fin de analizar gráficamente la distribución de las variables y la concentración de las distintas categorías.

Se analizaron los resultados del desempeño del software LIBRA con el fin de redactar un informe para el Ministerio de Ciencias, Tecnología e Innovación. El objetivo del informe fue presentar una evaluación detallada de la efectividad y utilidad del software en el contexto de la segmentación de imágenes mamográficas de acuerdo a los diferentes proveedores y modelos de scanners.

**2. Cálculo de la relación entre la densidad mamaria calculada por el modelo y la clasificación ACR realizada por los médicos radiólogos.**

Cada radióloga estableció de forma individual la clasificación ACR y la valoración de la segmentación de densidad entregada por el software LIBRA, clasificándola en tres categorías: buena, regular y mala. A los estudios que contaban con confirmación de segunda lectura se les sometió a un análisis de concordancia en relación con la clasificación ACR y otro análisis para las

## Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.

evaluaciones realizadas por ambas radiólogas en cuanto a la segmentación del software. La validación del software se realizó sobre aquellas lecturas en las que la segmentación fue clasificada como ‘Buena’ y se empleó el criterio de la clasificación ACR de la radióloga con más experiencia.

Se implementaron análisis de clusters, y una regresión lineal y logística de los datos para estimar la relación entre el porcentaje de densidad mamario arrojado por el software LIBRA con respecto a la clasificación ACR.

En primer lugar, se realizó la regresión lineal de los datos usando como variable en el eje “y” el porcentaje de densidad y el modelo del scanner empleado y como variable objetivo la clasificación ACR. Posteriormente, se llevó a cabo una división de los datos en una proporción de 80-20 para crear conjuntos de entrenamiento y prueba respectivamente. Se empleó el modelo de regresión lineal Ordinary Least Squares (OLS) [26].

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_1 x_i - b_0)^2 = \sum_{i=1}^n (\hat{\epsilon}_i)^2 = \min \quad [26]$$

A partir del modelo se obtuvieron los residuales los cuales se emplearon como el nuevo porcentaje de densidad para el entrenamiento del modelo de regresión logística junto con la clasificación ACR. Se utilizaron los datos de prueba para realizar predicciones y se construyó una matriz de confusión para evaluar el desempeño del modelo. Esta matriz proporcionó información valiosa sobre cómo las predicciones del modelo se compararon con las etiquetas reales, permitiendo una evaluación detallada de su rendimiento. Utilizando un modelo de regresión logística, se determinaron los intervalos del porcentaje de densidad mamaria obtenidos mediante el software y su correlación con cada una de las clasificaciones ACR realizadas por los médicos radiólogos. Esta matriz permitió obtener una métrica de exactitud en las predicciones, la cual se esperaba que sea mayor a 0.70.

Utilizando los rangos de clasificación ACR, se usaron cajas de bigotes para asignar cada valor de porcentaje de densidad mamaria a la categoría correspondiente para estratificar los valores de densidad mamaria.

---

**3. Evaluación de un nuevo modelo para la estimación de densidad mamaria a partir de imágenes de mamografía de una base de datos abierta.**

Una vez evaluado el rendimiento y desempeño actual del modelo LIBRA, se evaluó un nuevo modelo para estimar la densidad mamaria utilizando una base de datos abierta.

Para el desarrollo de esta fase se realizó una recopilación de estudios con imágenes de mamografías. Se inició con la implementación de DeepLIBRA, el cual es un método de inteligencia artificial para una cuantificación de la densidad mamaria con validación independiente en la evaluación del riesgo de cáncer de mama [27]. La diferencia es que está implementado en Python.

Posteriormente se empleó el software de MONAI el cual es un modelo previamente entrenado para la clasificación de la densidad mamaria, que no da un valor cuantitativo sino un array de probabilidades de cada clasificación ACR [28, 29].

Se emplearon las mismas imágenes con segmentación ‘Buena’ de LIBRA y se les realizó un cambio de formato de DICOM a .png el cuál es el formato que acepta MONAI, de igual forma, se les realizó un redimensionamiento de la imagen a 299x299 y se eliminaron las imágenes que contaban con artefactos, todo esto con el fin de comparar las dos matrices de confusión de ambos modelos y la precisión por cada categoría.

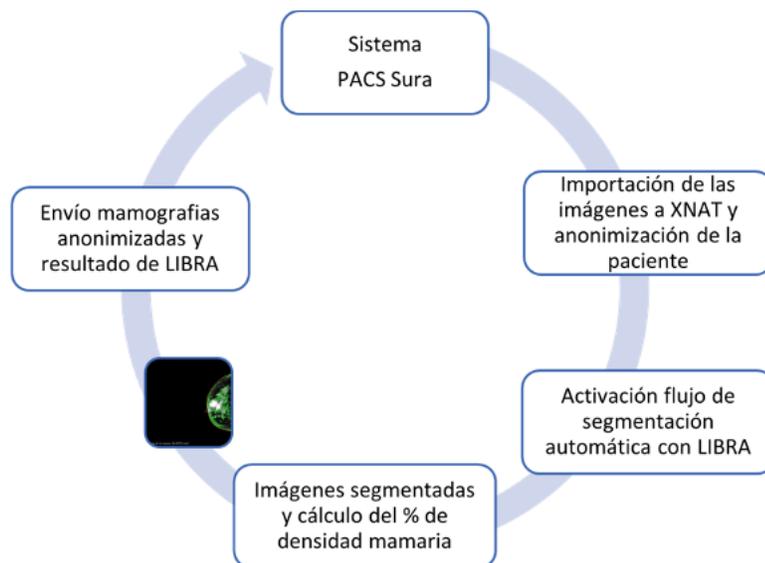
De igual forma, se usaron todas las imágenes para verificar si existía un efecto del scanner en la predicción del modelo.

---

## V. RESULTADOS

El flujo de trabajo implementado se describe en la figura 1. Este incluye la importación de las imágenes desde el PACS a la central de postprocesamiento XNAT en la que se realiza la anonimización de los archivos y se ejecuta el algoritmo de segmentación automática LIBRA para obtener el cálculo de la densidad mamaria. Finalmente, las imágenes anonimizadas y el resultado de la segmentación se envían nuevamente al PACS para la lectura que realizan las radiólogas. Esta lectura se consignó en una lista de sharepoint (Figura 2).

Cada radióloga estableció la densidad mamaria bajo los criterios de la 5ta edición BI-RADS de 2013 y la evaluación de la segmentación LIBRA. Toda la valoración se llevó a cabo en pantallas óptimas para la interpretación de los estudios. En total, se leyeron 2160 estudios de mamografía.



*Figura 1. Flujo de trabajo para lectura de MG*

Lectura de imágenes ☆

Title	Identificación	Sujeto:birad	Clasificación ...	Evaluación de ...	Porcentaje d
PRS010824	PRS010824	BIRADS 6	B	Regular	2.84%
PRS005447	PRS005447	BIRADS 6	B	Regular	2.58%
PRS011677	PRS011677	BIRADS 6	C	Regular	12.50%
PRS011502	PRS011502	BIRADS 5	B	Regular	2.75%
PRS011666	PRS011666	BIRADS 6	C	Buena	14.50%
PRS010818	PRS010818	BIRADS 5	C	Mala	0.00%
PRS005331	PRS005331	BIRADS 5	D	Buena	78.40%
PRS004855	PRS004855	BIRADS 5	C	Regular	6.76%
PRS001572	PRS001572	BIRADS 5	C	Buena	20.75%
○ PRS007070	PRS007070		B	Buena	4.30%

Recuento  
2160

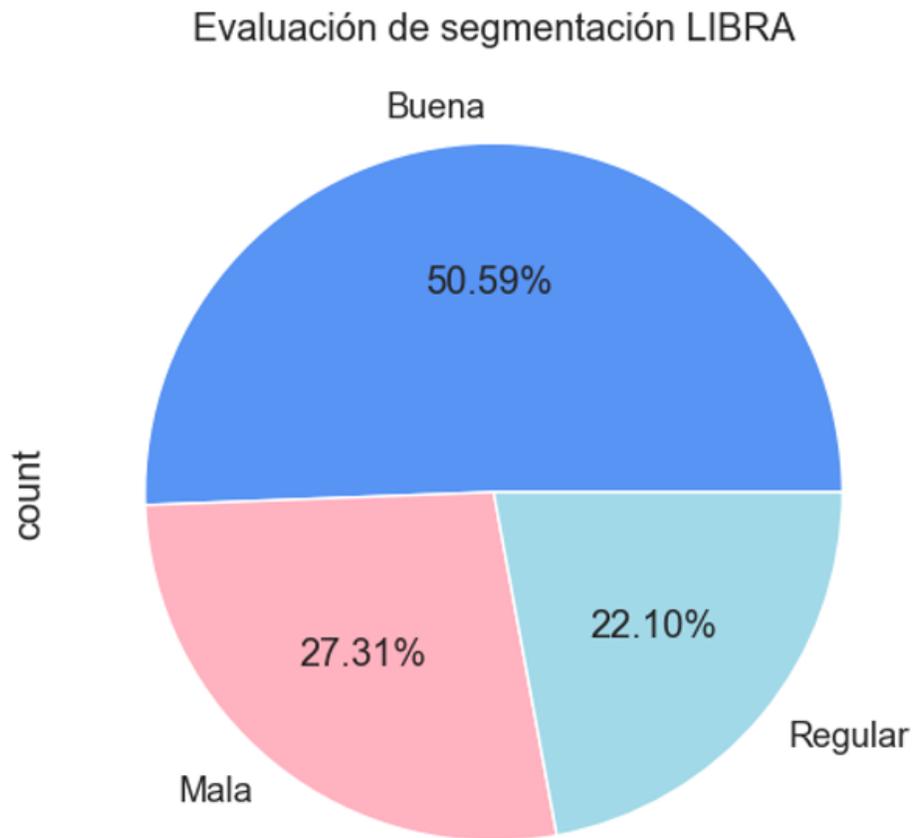
Figura 2. Lista lectura imágenes

## a. Análisis de concordancia clasificación ACR

Del total de pacientes leídas, 2013 estudios tuvieron confirmación de segunda lectura. Sobre esos estudios, se realizó un análisis de concordancia de la clasificación ACR utilizando la medida Kappa de Cohen [30] obteniendo un valor de concordancia de 0.398, que indica un acuerdo medio o aceptable. De igual forma, se realizó el análisis de concordancia para la evaluación de la segmentación de LIBRA dada por ambas radiólogas, obteniendo un índice Kappa con un valor de 0.493, el cual indica una concordancia moderada. De acuerdo con la literatura, los valores del coeficiente kappa de Cohen proporcionan las siguientes interpretaciones: malo ( $\kappa < 0$ ), leve ( $\kappa = 0.0-0.20$ ), aceptable ( $\kappa = 0.21-0.40$ ), moderado ( $\kappa = 0.41-0.60$ ), sustancial ( $\kappa = 0.61-0.80$ ) y casi perfecto ( $\kappa = 0.81-1.00$ ) [31].

## b. Estimación del desempeño del modelo actual implementado en la compañía a partir de las lecturas y valoraciones realizadas por los médicos radiólogos.

2123 estudios fueron procesados por el software LIBRA y contaban con confirmación de segunda lectura. Basándose en los resultados de concordancia previamente reportados, se tomó la clasificación realizada por la radióloga más experimentada para los siguientes análisis. Se seleccionaron los datos de las pacientes que presentaban una evaluación de la segmentación clasificada como “Buena”, la cual representa un 50.59% del total de los datos, es decir un total de 1069 estudios (Figura 3).



*Figura 3. Distribución porcentual de la evaluación de la segmentación de los datos evaluados.*

Del total de pacientes leídas, 1965 estudios tuvieron información completa referente a la evaluación de la segmentación e información del equipo. La tabla 1 muestra la distribución de la evaluación de la segmentación en los diferentes tipos de escáner, y sus respectivos modelos y en la figura 4 y 5 se muestra esa misma distribución en términos porcentuales.

**Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.**

**TABLA 1**

*Distribución de la evaluación de la segmentación según los equipos de mamografía.*

<b>Tipo de escáner</b>	<b>Modelo de escáner</b>	<b>Buena</b>	<b>Mala</b>	<b>Regular</b>	<b>Suma</b>
<b>Agfa</b>	CR30-Xm	0	13	5	18
	DR_18M	1	53	8	62
	DR_24M	10	70	27	107
<b>CARESTREAM</b>	CLASSIC CR	34	2	3	39
<b>GE HEALTHCARE</b>	Senographe Pristina	2	91	6	99
	Senographe_Crystal	59	145	28	232
<b>GE MEDICAL SYSTEMS</b>	Selenia Dimensions	133	47	102	282
	Senographe Essential ADS_56.22.2	1797	17	114	328
	Senographe Essential DBT Reconstruction Box_01.10.4	0	1	0	1
	Senographe Essential VERSION ADS_54.20	16	1	8	25
	Senographe Essential VERSION ADS_55.31.10	18	3	6	27
	Senographe Pristina	44	25	25	94
	Senographe_Crystal	129	10	247	186
<b>HOLOGIC, Inc.</b>	Lorad Selenia	29	4	22	55
	Selenia Dimensions	319	15	30	364
<b>KODAK</b>	CLASSIC CR	33	3	5	41
<b>None</b>	None	3	0	2	5
<b>Suma</b>	-	<b>1027</b>	<b>500</b>	<b>438</b>	<b>1965</b>

# Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.

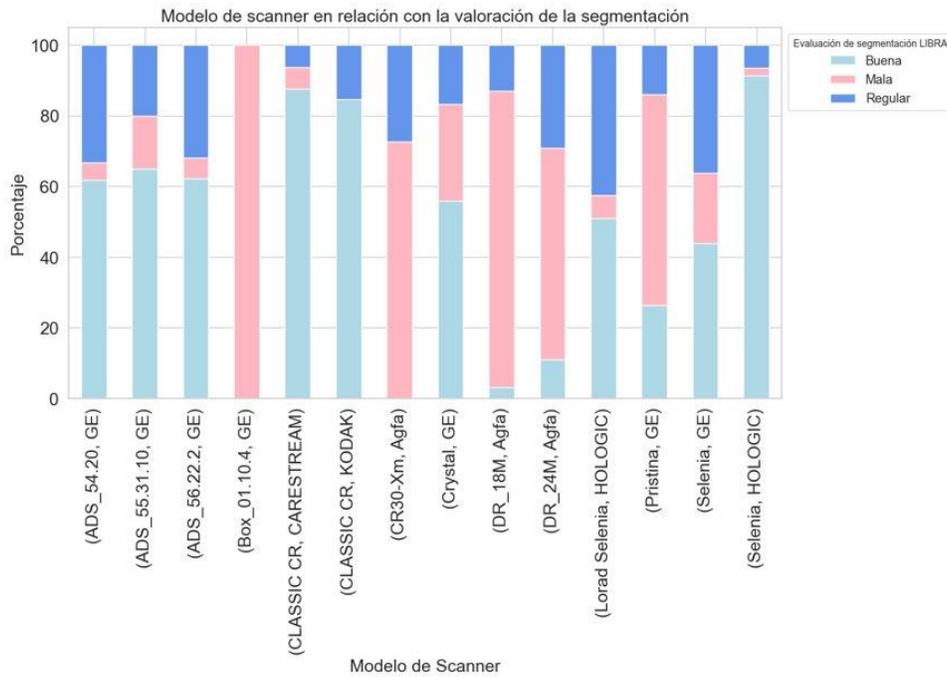
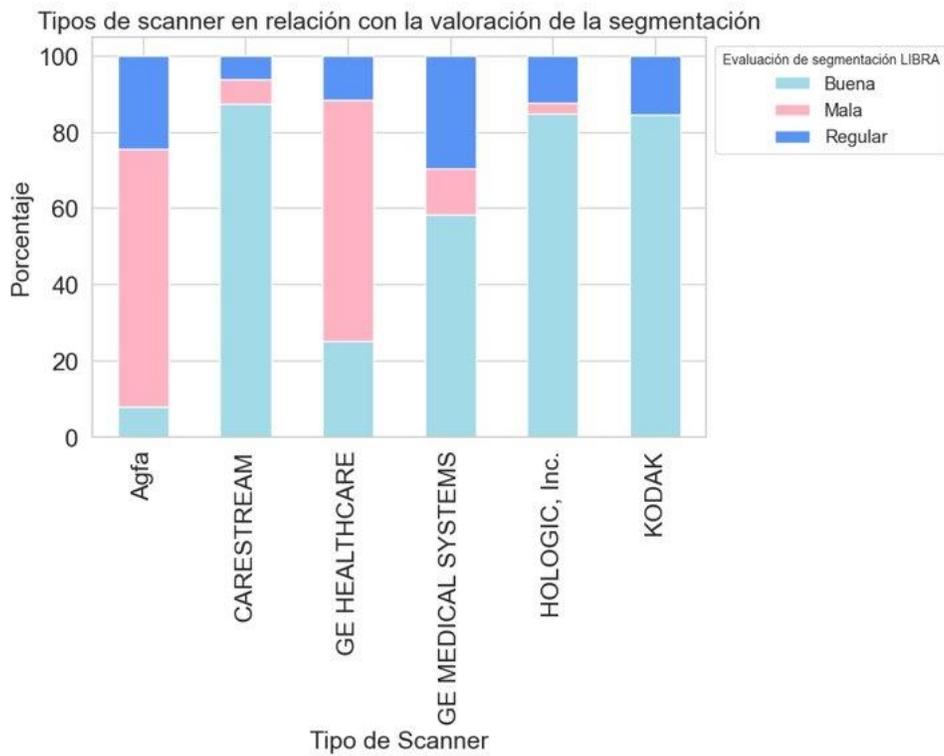


Figura 4. Distribución porcentual de la evaluación de la segmentación de acuerdo con modelo de cada equipo.



## Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.

Figura 5. Distribución porcentual de la evaluación de la segmentación de acuerdo con el tipo de equipo.

De lo anterior, se encuentra que el algoritmo LIBRA tuvo un mejor desempeño para equipos específicos, entre los que se encuentran HOLOGIC con el modelo Selenia Dimension con un 87.64% de las segmentaciones clasificadas como buenas, seguido del equipo CARESTREAM con el modelo CLASSIC CR con un 87.18% y finalmente KODAK con el mismo modelo CLASSIC CR con un 80.49 % de las segmentaciones buenas.

La figura 6 muestra la distribución porcentual de la evaluación de la segmentación en las diferentes categorías de densidad, y la figura 7 indica la relación entre la evaluación de la segmentación con la clasificación BIRADS. Se pueden observar distribuciones similares en los porcentajes de evaluación a lo largo de las categorías, lo que afirma la dependencia del desempeño del algoritmo con el equipo de adquisición.

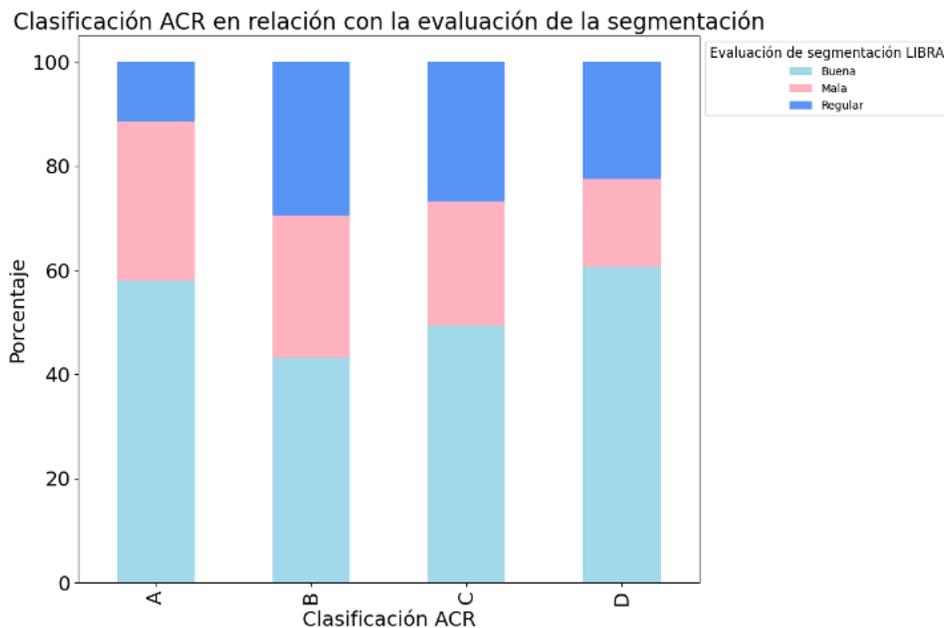
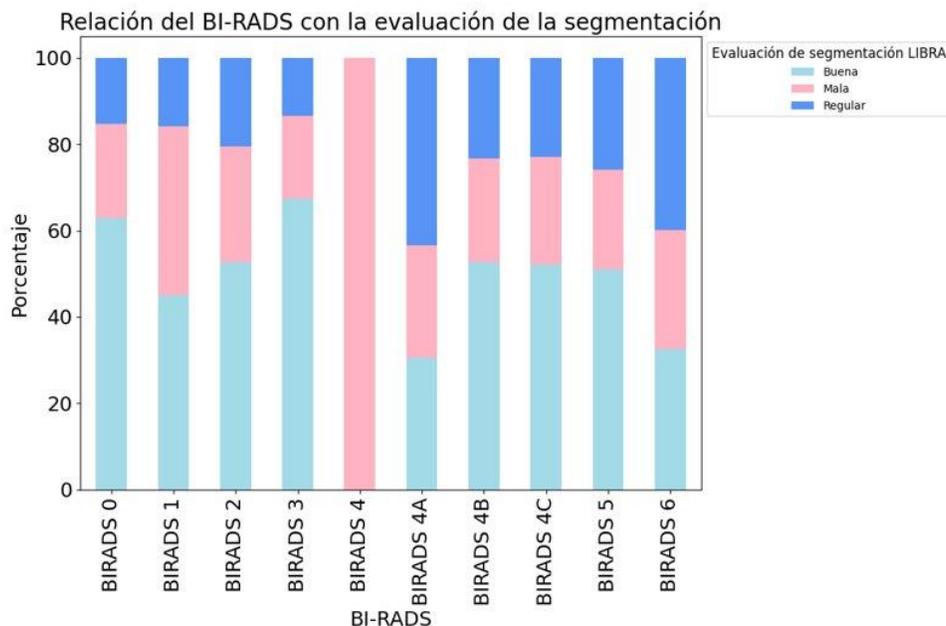


Figura 6. Distribución porcentual de la evaluación de la segmentación de acuerdo con la clasificación ACR.



*Figura 7. Distribución porcentual de la evaluación de la segmentación de acuerdo la clasificación BIRADS.*

**c. Cálculo de la relación entre la densidad mamaria calculada por el modelo y la clasificación ACR realizada por los médicos radiólogos.**

Para el análisis de relación entre las categorías del ACR y el porcentaje de densidad calculado por LIBRA, se tomaron en cuenta únicamente los estudios que tuvieron una evaluación buena de la segmentación (1065 estudios). La tabla 2 muestra el número de estudios y su distribución porcentual por cada categoría.

**TABLA 2**  
*Número de estudios con buena segmentación en cada categoría ACR.*

Clasificación	A	B	C	D	Total
ACR					
Total de estudios	440	355	216	54	1065
Porcentaje	41.44%	33.21%	20.30%	5.05%	100%

Con estos estudios, también fue posible evidenciar el comportamiento de los casos y controles en cada categoría de la clasificación ACR (Figura 8 y tabla 3).

## Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.

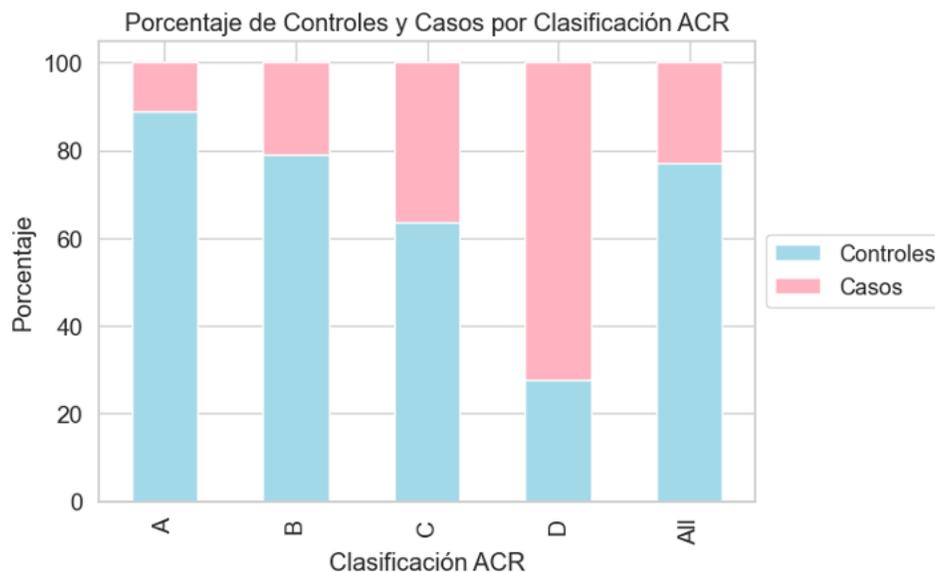


Figura 8. Distribución porcentual de los casos y controles en cada categoría ACR de los estudios con buena segmentación.

**TABLA 3**  
Estudios como casos y controles en cada categoría ACR.

Clasificación ACR	Caso	Contol	Suma
A	39	312	351
B	64	242	306
C	64	112	176
D	34	13	47
<b>Suma</b>	<b>201</b>	<b>679</b>	<b>880</b>

Para estimar la relación entre la clasificación ACR y el porcentaje de densidad obtenido con LIBRA se llevó a cabo un análisis de regresión logística. Este tipo de regresión permite calcular la predicción de una variable categórica (ACR) en función de una variable cuantitativa (% densidad).

Antes de aplicar el modelo, se realizó un sobremuestreo de los datos para abordar la problemática del desbalance las clasificaciones ACR descrita en la figura 9.

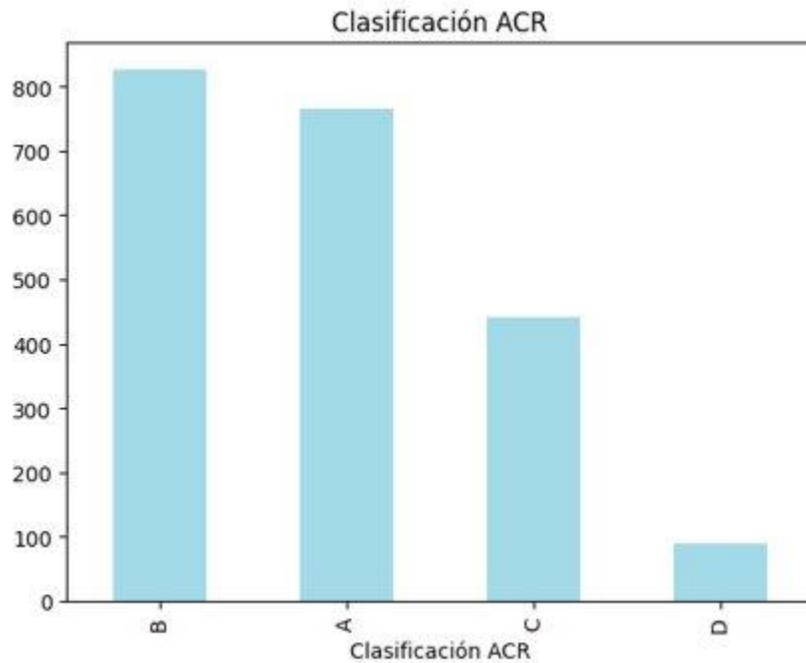
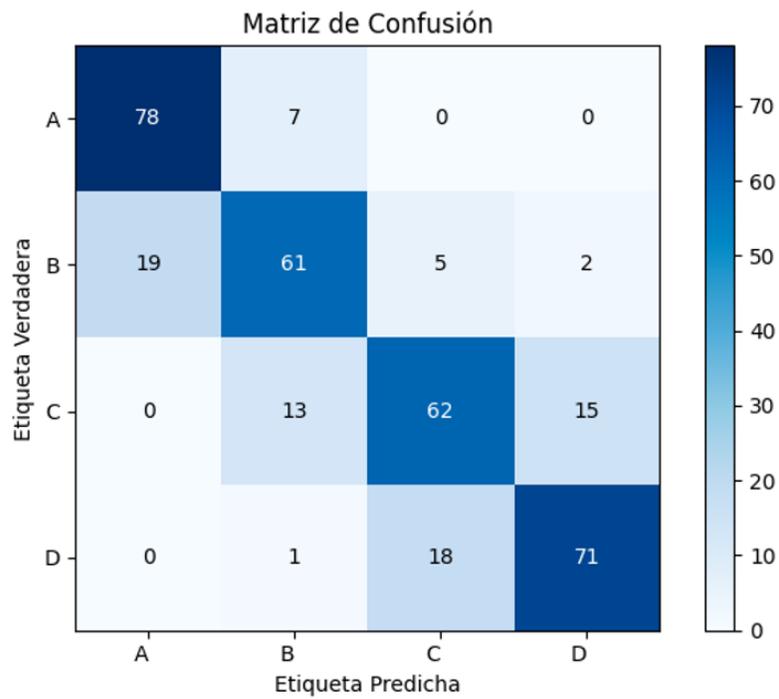


Figura 9. Distribución de la clasificación ACR en cada categoría.

En la figura 10 se muestra la matriz de confusión obtenida sin tener en cuenta el efecto evidente del scanner.



## Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.

Figura 10. Matriz de confusión de la regresión logística con el porcentaje de densidad y la clasificación ACR.

El modelo de regresión logística tuvo una precisión de 0.77. El hecho de que la precisión del modelo sea alta se le puede atribuir a que únicamente se seleccionaron los datos de segmentación considerados como ‘Buenos’, lo cual elimina algunos modelos de scanner en los que todos los estudios tuvieron mala segmentación. En la tabla 4, se observa que la precisión por categoría es buena para A y D.

**TABLA 4**  
Porcentaje de acierto del porcentaje de densidad en cada categoría.

Clasificación ACR	Precisión
A	91.76%
B	70.11%
C	68.89%
D	78.89%

Ahora bien, para eliminar el efecto del scanner en el modelo, se empleó inicialmente una regresión lineal y se usaron los residuales como el nuevo porcentaje de densidad en el modelo de regresión logística obteniendo la matriz de la figura 11. Para este caso, se encontró la matriz de confusión para los datos de prueba sin realizar previamente un sobremuestreo de los datos, se evidencia que existe un desbalance de las clases.

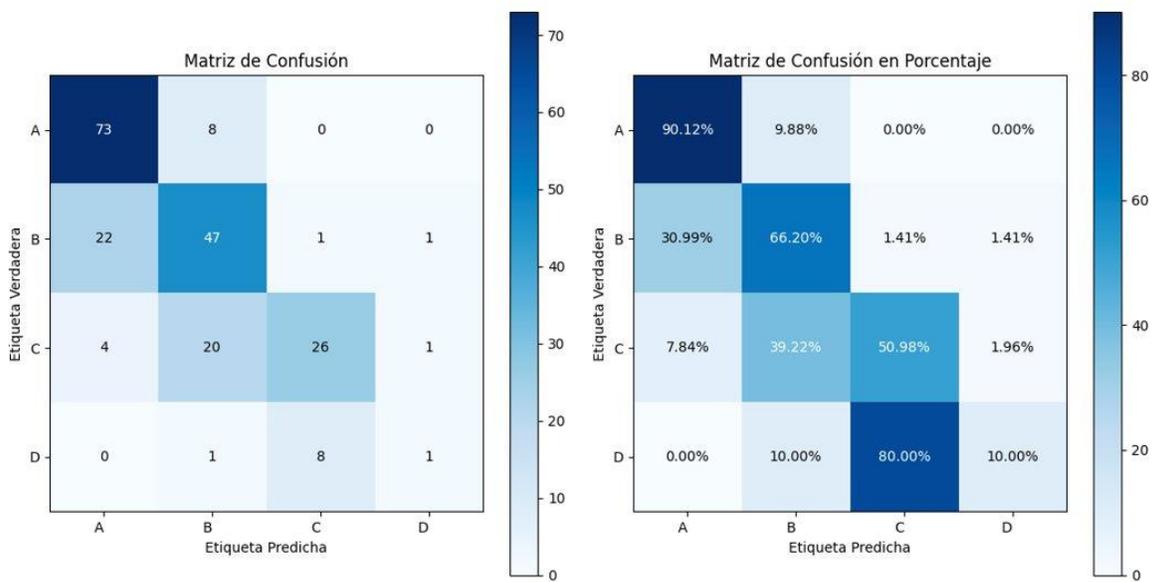


Figura 11. Matriz de confusión eliminando el efecto del scanner.

En la Tabla 5 se consolidan los porcentajes de acierto por cada categoría, obteniendo el mayor valor para la categoría A y siendo muy deficiente en la predicción de la categoría D. De igual forma, en la tabla 4 se observan los intervalos de confianza obtenidos para cada porcentaje de acierto.

**TABLA 5**

*Porcentaje de acierto del porcentaje de densidad en cada categoría y los intervalos de confianza.*

Clasificación ACR	Porcentaje de acierto	Intervalos de confianza
A	90.12%	[0.8435, 0.9539]
B	66.20%	[0.5736, 0.7440]
C	50.98%	[0.3900, 0.6106]
D	10.00%	[0.0000, 0.2500]

Se observa que el modelo de regresión eliminando el efecto del scanner tiene un buen rendimiento en A y en B, caso contrario para la clasificación C y D donde se obtiene un pobre rendimiento para esta categorías.

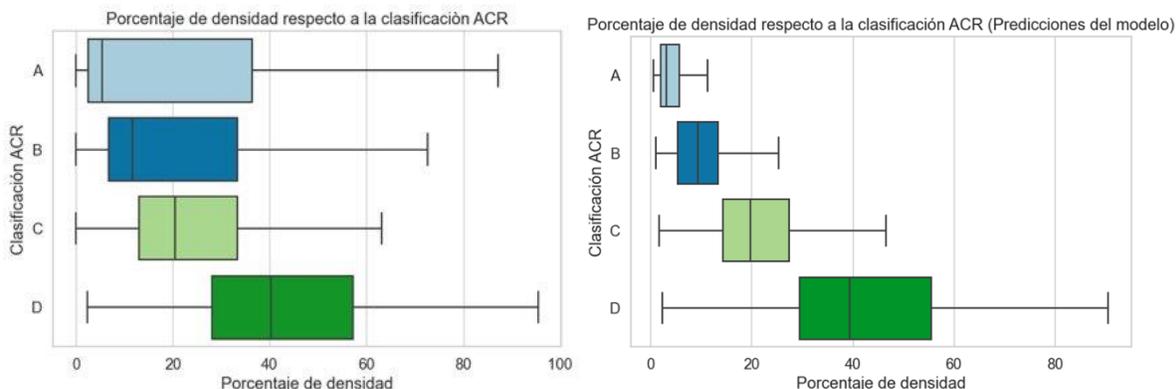
Finalmente, a partir del modelo de regresión logística eliminando el efecto del scanner fue posible establecer rangos del porcentaje de densidad arrojado por el software LIBRA para cada clasificación ACR (Tabla 6 y Figura 12).

**TABLA 6**

*Rango de porcentaje predicha por el modelo en cada clasificación ACR.*

Clasificación ACR	Rango de porcentaje
A	2.08% - 5.30%
B	7.41% - 12.80%
C	17.12% - 26.30%
D	28.16% - 47.86%

En la Figura 12 Se observa la comparación del diagrama de cajas previo la implementación del modelo de regresión lineal



*Figura 12. A la izquierda el diagrama de cajas inicial y a la derecha el diagrama de cajas de las predicciones del modelo de regresión logística para cada clasificación ACR.*

**d. Evaluar un nuevo modelo para la estimación de densidad mamaria a partir de imágenes de mamografía de una base de datos abierta.**

Se llevó a cabo la evaluación de un modelo para la estimación de densidad mamaria a partir de imágenes de mamografía, utilizando una base de datos abierta disponible en GitHub [27, 29]. Inicialmente, se implementó el software DeepLibra [27], alojado en dicha plataforma, el cual se distingue por su funcionalidad en Python en contraposición a LIBRA, que está desarrollado en MATLAB.

El funcionamiento de DeepLibra se destaca por su capacidad para llevar a cabo la eliminación del fondo y del músculo pectoral mediante el empleo de una red U-Net. Posteriormente, se generan superpíxeles, los cuales son clasificados por una Máquina de Vectores de Soporte (SVM) con base en las características extraídas, culminando en una segmentación del tejido denso.

Un inconveniente identificado en este modelo radicaba en la divergencia de etiquetas de los archivos DICOM utilizados. Mientras que las imágenes de prueba de DeepLibra contenían la etiqueta "FOR PROCESSING", las imágenes empleadas tenían la etiqueta "FOR

## **Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.**

32

---

PRESENTATION". Esta disparidad limitaba la efectividad del modelo exclusivamente a las imágenes de prueba, excluyendo las imágenes utilizadas en el proceso.

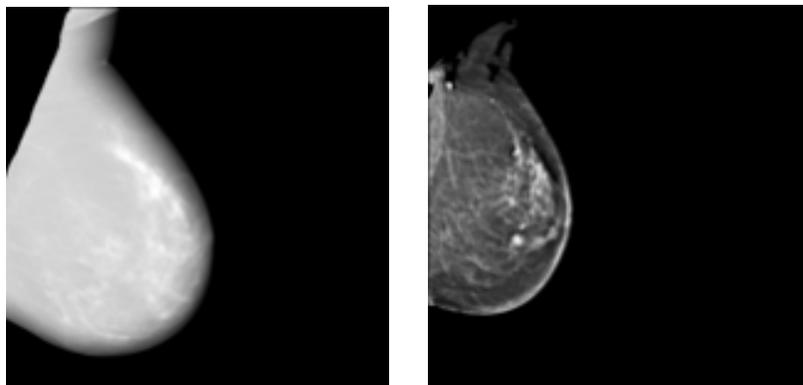
No obstante, a pesar de esta limitación, el modelo generaba carpetas intermedias con un flujo de trabajo específico:

- Carpeta "air\_net\_data": Contiene imágenes originales de mamografía en formato Full-Field Digital Mammography (FFDM) de 16 bits. Posteriormente, se genera una versión de la imagen con relleno de ceros a una resolución de intensidad de 8 bits, utilizada en la segmentación de fondo mediante U-Net.
- Carpeta "pec\_net\_data": Se emplean imágenes de vista mediolateral oblicua (MLO) para la segmentación del músculo pectoral mediante una Red Neuronal Convolutiva (CNN). Esta carpeta se basa en imágenes del conjunto de datos ds1, donde el músculo pectoral es más observable en vistas MLO.
- Carpeta "final\_images": Esta imagen resultante se utiliza para la generación de superpíxeles y el análisis de características radiómicas. Posteriormente, la SVM clasifica los superpíxeles según las características extraídas, logrando así una segmentación del tejido denso.
- Carpeta "superpixels": El módulo de cálculo de densidad mamaria divide la mama en superpíxeles, calculando características radiómicas globales y de superpíxeles. Estas características se emplean como entradas para modelos de aprendizaje automático, clasificando los superpíxeles como densos o no densos y calculando la densidad mamaria.
- Carpeta "breast\_density": Este módulo calcula un total de 101 características radiómicas para toda la imagen y 50 características adicionales en cada superpíxel. Se utiliza una SVM entrenada en ds3-a para clasificar los superpíxeles como densos o no densos, empleando características de textura como predictores.

En la última carpeta, "breast\_density", no se realiza la segmentación de la imagen, y por ende, no se obtiene un resultado en términos de un porcentaje de densidad. Las imágenes resultantes de la

## Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.

carpeta "final\_images" se conservan, ya que han sido procesadas para eliminar el fondo y el músculo pectoral, y se destinan a un uso potencial en modelos subsiguientes (Figura 13).



*Figura 13. Imagen del proceso de DeepLIBRA. A la derecha la imagen de la base de datos empleada y a la izquierda la imagen empleada por los autores de DeepLIBRA [27].*

Se procedió a emplear el modelo MONAI, en el cual los autores llevaron a cabo una modificación en las últimas capas del modelo Inception-V3 original. Estas fueron reemplazadas por una capa completamente conectada que albergaba 1024 nodos, seguida de cuatro nodos de salida con activación softmax. La implementación la realizaron con Keras y TensorFlow-1.10. La tasa de aprendizaje inicial fue establecida en 0,001 utilizando un optimizador de descenso de gradiente estocástico con un tamaño de lote de 8. El proceso de reentrenamiento se concluyó después de 50 épocas. Durante el entrenamiento/validación, los autores supervisaron el rendimiento del algoritmo mediante el monitoreo de la entropía cruzada binaria en el conjunto de validación por época. El modelo se guardaba conservando la mayor precisión hasta ese punto, y si la precisión de la validación aumentaba en épocas posteriores, se actualizaba el modelo. Cuatro modelos se entrenaron en base a diferentes protocolos de adquisición. [28, 29].

La salida del modelo consiste en la clasificación ACR de la imagen, proporcionando una evaluación cualitativa (A, B, C o D) en lugar de un valor cuantitativo. Para la preparación de los datos, se llevó a cabo una limpieza de imágenes con artefactos mediante un control de calidad. Este control generaba un resumen de todas las imágenes de cada paciente, que debían ser cuatro: una vista CC y una MLO para cada lateralidad. El estado del control de calidad era 200 si la paciente

## Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.

34

contaba con las cuatro imágenes y 249 si tenía más o menos. Esto facilitaba la identificación de imágenes con artefactos o daños para su eliminación.

Esta medida era esencial, ya que MONAI proporcionaba una clasificación por cada vista. Se promediaron todas las predicciones por paciente para obtener una predicción promedio general, considerando que el resultado era un array de probabilidades.

Se llevaron a cabo matrices de confusión al comparar con la clasificación ACR proporcionada por la doctora con mayor experiencia, utilizando todos los datos del proyecto que contaban con el dato del ACR, los cuales fueron 2098 (Figura 14).

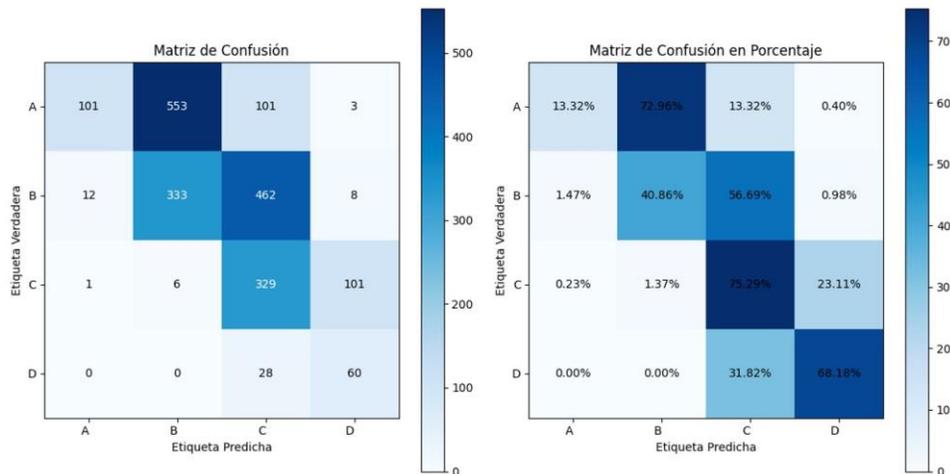


Figura 14. Matriz de confusión para el modelo de MONAI con todos los 2098 datos.

Para la comparación con LIBRA, se utilizaron exclusivamente los datos con segmentación 'Buena', los mismos empleados en LIBRA (1095), generando nuevamente la matriz. La precisión de esta fue de 0.392, considerablemente menor que en LIBRA (Figura 15).

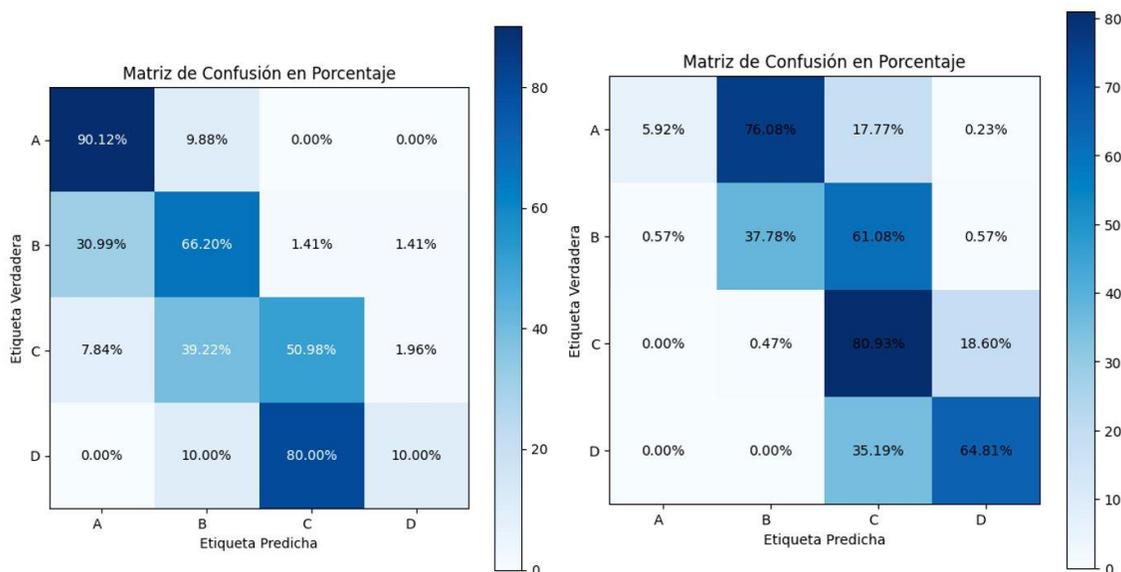


Figura 15. Comparación de matrices de confusión. A la izquierda la matriz obtenida del modelo LIBRA y a la derecha la matriz obtenida del modelo de MONAI.

La comparación por categoría también se realizó, destacando la mejor precisión en la categoría C con un 80.93% de exactitud para el caso de MONAI. Sin embargo, se obtuvo una precisión deficiente en la categoría A, con un 5.92% de exactitud, en contraste con lo observado en LIBRA en donde fue la máxima precisión con un 90.12% (Tabla 7)

**TABLA 7**  
*Comparación por categoría del software de LIBRA y de MONAI.*

Categoría	LIBRA	MONAI
A - 440	90.12%	5.92%
B - 355	66.20%	37.78%
C - 216	50.89%	80.93%
D - 54	10.00%	64.81%

**e. Análisis del efecto del scanner en MONAI**

Finalmente, se buscó analizar si existía un efecto del escáner en la predicción de MONAI mediante un análisis estadístico. A partir de las imágenes, se concluyó que no hubo un escáner donde

# Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.

MONAI funcionara mejor, a diferencia de lo observado en LIBRA donde se encontraron 3 en los cuales el rendimiento fue bueno. En todos los casos, el 50% o más de las predicciones fueron incorrectas en todos los scanner (Figura 16 y 17).

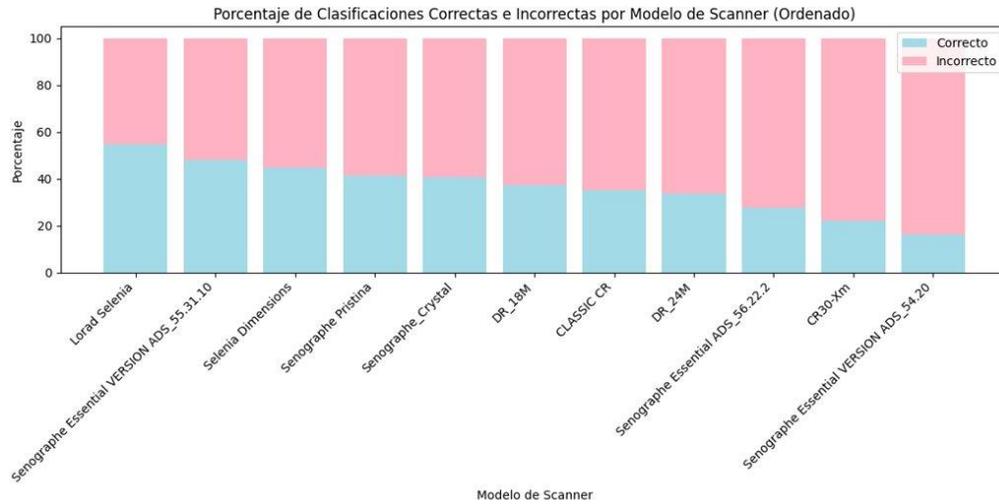


Figura 16. Porcentaje de las clasificaciones correctas e incorrectas de los scanners usados en MONAI.

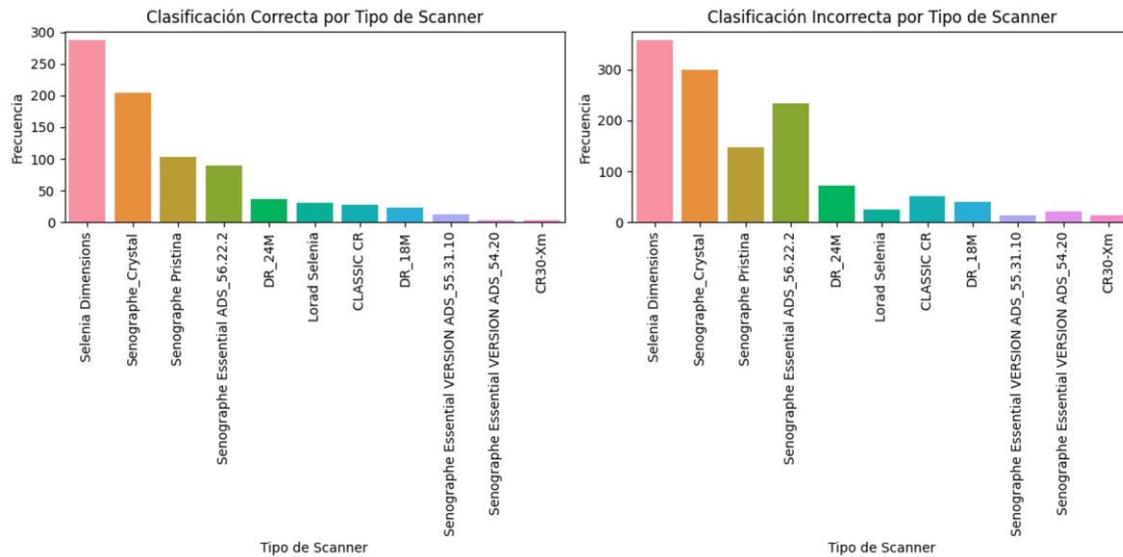
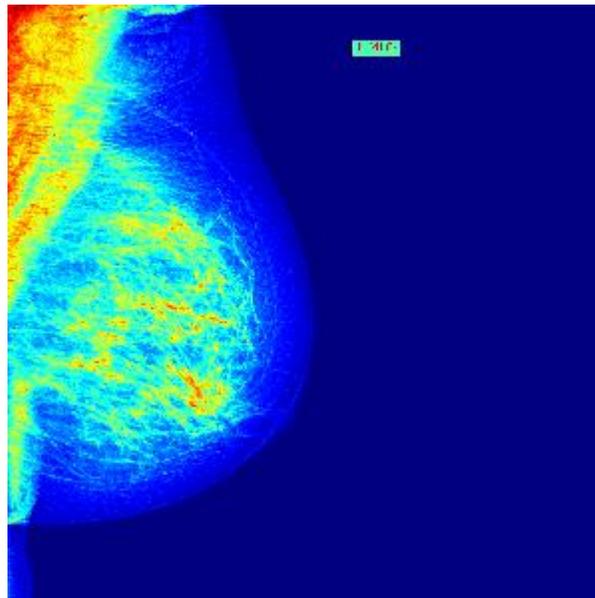


Figura 17. Frecuencias de clasificaciones correctas e incorrectas de los scanners empleando MONAI.

Finalmente, se realizaron mapas de intensidad de las imágenes que ingresaron a MONAI y se reportaron el total de píxeles (el cual era el mismo para todos por el resize que se realizó) y se halló una relación de los píxeles intensos respecto al total, hallando así un porcentaje de zonas densas o píxeles intensos (Figura 18). A pesar de esto, se observa que toma el pectoral como píxeles densos, lo que se traduce en una mala elección del umbral.



*Figura 18. Mapas de intensidad de las imágenes.*

## VI. ANÁLISIS

En el presente informe se abordó la evaluación del desempeño del software de segmentación automática LIBRA para el cálculo de la densidad mamaria en imágenes de mamografía. Se estableció un flujo de trabajo que permitía cuantificar la concordancia en la lectura de las imágenes, y la calificación de la segmentación obtenida con el software.

Los resultados de concordancia realizados mostraron un valor de 0.398 para la categorización ACR, lo cual indica un acuerdo aceptable. En la literatura se ha reportado que la tasa de acuerdo en la categorización BI-RADS entre los médicos radiólogos del Instituto Nacional de Cancerología de Colombia y los médicos radiólogos de otras instituciones, es de 54.9%, con

## **Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.**

38

una índice kappa ( $k$ ) de 0,21, lo cual indica una concordancia débil en la categorización BI-RADS [31, 32]. Otro estudio encontró que el valor del índice kappa para la clasificación BI-RADS fue 0.389, lo cual se considera aceptable de acuerdo a la literatura [31, 32], resaltando así la necesidad de llevar a cabo procesos de estandarización. Estos resultados sugieren una baja concordancia en los informes mamográficos lo cual resalta la variabilidad que puede existir en la valoración de cada observador al momento de interpretar las mamografías a pesar del sistema BI-RADS estandarizado [18]. Así mismo, se enfatiza la importancia de llevar a cabo análisis cuantitativos en tiempo real de las lesiones.

La subjetividad de la prueba de mamografía y la categorización de la densidad mamaria bajo un único observador (médico radiólogo) puede generar un sesgo. LIBRA es un sistema automático que según la literatura puede ayudar a reducir los sesgos de subjetividad en imágenes diagnósticas. Sin embargo, al incorporarlo al flujo de trabajo, se encontró que su desempeño depende la tecnología de adquisición empleada. De los equipos de mamografía evaluados en el proyecto, se encontró un buen desempeño del software para las marcas HOLOGIC y CARESTREAM, lo cual respalda lo reportado en la literatura sobre la validación del modelo la cual fue realizada en HOLOGIC, Inc y en GE [12]. Dicho esto, se esperaba que en General Electric el desempeño del modelo fuera similar a como lo fue en HOLOGIC, Inc, sin embargo, se encontró que el rendimiento fue bueno, únicamente en algunos modelos específicos. Por ejemplo, para GE Healthcare, el desempeño fue principalmente malo, caso contrario a GE Medical Systems.

Ahora bien, para el caso de la regresión lineal sin tener en cuenta el efecto del scanner, se observa que la estimación de la relación entre la clasificación ACR y el porcentaje de densidad obtenido con LIBRA muestra que se puede discernir más fácilmente la categoría A (predominio de tejido adiposo) asociada a un rango de densidad de 2.09% a 5.16%. Así mismo, la categoría D (mama totalmente fibroglandular) asociada a un rango de densidad de 33.23% a 54.27%, también puede predecirse con buen acierto. Sin embargo, la identificación y diferenciación entre las categorías B (tejido glandular disperso) y C (tejido fibroglandular heterogéneo) tuvo la menor precisión. Lo anterior plantea un desafío para la segmentación cuando la mama presenta un tejido heterogéneo entre graso y denso. Lo que lleva a considerar modelos que involucren el entrenamiento con este tipo de imágenes. Además, futuros análisis deberán tener en cuenta el efecto que puede tener el equipo de adquisición en la estimación.

## **Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.**

---

La mamografía ofrece ventajas, pero también presenta limitaciones, especialmente en mamas densas. La aplicación de la inteligencia artificial constituye una de las innovaciones más recientes diseñadas para superar las restricciones de la mamografía. En comparación con patrones menos densos, las mamas densas en una mamografía generan más resultados falsos negativos. Este fenómeno se atribuye al efecto de enmascaramiento, que reduce la sensibilidad mamográfica y triplica la tasa de recordatorio, la cual se refiere a la proporción de casos en los que se requiere una revisión adicional después de una mamografía inicial. Los resultados falsos negativos, en última instancia, provocan demoras en el diagnóstico; exposición a la radiación y sobre diagnóstico de tumores que podrían no representar una amenaza para la vida [33].

En la literatura se encuentran respaldos de LIBRA y del uso de esta herramienta de medición de densidad completamente automatizada y de acceso público. Estudios confirman que las medidas de densidad de LIBRA y otros modelos como Cumulus se asociaron significativamente tanto con el riesgo de cáncer de mama a corto como a largo plazo, y la magnitud de las asociaciones no se atenuó durante el período de seguimiento de 10 años. Estos autores exponen que, al proporcionar medidas de densidad confiables y robustas, LIBRA facilitará futuros estudios longitudinales a gran escala para abordar múltiples preguntas relacionadas con los determinantes de la densidad mamográfica y su relación con el riesgo de cáncer de mama, y cómo estas relaciones pueden variar con el tiempo [34].

Como se mencionó anteriormente, la densidad mamográfica es un factor de riesgo conocido para el cáncer de mama [8], donde las mujeres mayores presentan tasas más altas de incidencia de ocurrencia de este tipo de cáncer. En la literatura, se han usado modelos predictivos de clasificación de esta densidad mamaria mediante algoritmos de machine learning, entre ellos se encuentran modelos como Decision Tree (Tree), Support Vector Model (SVM) and k-Nearest Neighbour (kNN) [35]. El enfoque de estos autores demostró que podría ser potencialmente utilizado en entornos clínicos para ayudar a categorizar mamografías desde mamas extremadamente densas (Grupo D de BI-RADS) hasta mamas completamente adiposas (Grupo A de BI-RADS) lo que es comparable con los resultados obtenidos de la matriz de confusión sin quitar el efecto del scanner ya que igualmente es posible discernir entre ambos grupos, pero las predicciones fallan en las categorías B y C.

## **Implementación de modelos de inteligencia artificial para el cálculo de densidad mamaria en imágenes de mamografía.**

---

También existen modelos con un enfoque binario de detección de cáncer donde se aplica un etiquetado suave de positivos: en lugar de tener etiquetas binarias (positivas o negativas) para las imágenes de mamografía, el pipeline utiliza un enfoque de "etiquetado suave" para las muestras positivas. Esto significa que, en lugar de una etiqueta binaria, se asigna una probabilidad o puntaje continuo que indica la "positividad" de la muestra para finalmente obtener un resultado binario de 1 si la predicción es de cáncer y de 0 en el caso contrario [36]. Este tipo de enfoque se planteó, sin embargo, no se utilizó debido a que es más beneficioso obtener una medida cuantitativa de la densidad mamaria o la clasificación ACR directamente, en lugar de depender de un resultado binario.

En cuanto a la comparación de los modelos de LIBRA y MONAI se realizó una prueba estadística chi cuadrado, obteniendo que hay evidencia para rechazar la hipótesis nula, por lo que hay diferencias significativas entre los modelos.

Dada la posibilidad de que los tumores se vean enmascarados debido a la alta densidad del tejido mamario, resulta crucial realizar un seguimiento cercano a los pacientes con una alta densidad. Esto se debe a que la alta densidad mamaria puede dar lugar a falsos positivos, donde las áreas densas pueden ser malinterpretadas como anomalías, así como a falsos negativos, donde los tumores reales pueden pasar desapercibidos. Por lo tanto, un monitoreo más intensivo en este grupo de pacientes es esencial para mitigar la posibilidad de diagnósticos incorrectos y garantizar una detección temprana y precisa de posibles malignidades. En este caso, el modelo de MONAI presentó una mejor predicción para estas últimas categorías, pero un desempeño muy pobre para las categorías A y B.

En este escenario, se destaca la mayor relevancia clínica de las categorías más densas, como se mencionó anteriormente. Cabe destacar que LIBRA presenta ventajas significativas, entre las cuales se incluye su capacidad para llevar a cabo la segmentación del tejido mamario en la imagen. Además, proporciona un valor porcentual de dicha segmentación, ofreciendo así información valiosa para el radiólogo. Esta característica adicional puede ser particularmente útil en la interpretación clínica, permitiendo una evaluación más detallada y precisa del tejido mamario, y potencialmente mejorando la toma de decisiones en el diagnóstico.

Estos clasificadores de aprendizaje profundo han demostrado resultados prometedores en el diagnóstico médico basado en imágenes, sin embargo, al mismo tiempo sigue siendo un desafío

---

la interpretación de las salidas de estos, es por eso que se realizaron los mapas de intensidad de las imágenes. Estos modelos de segmentación o detección con anotaciones píxel a píxel que indican las ubicaciones de lesiones malignas ayudan a mejorar la toma de decisiones, sin embargo, el etiquetado de las imágenes toma tiempo y recursos. Los mapas de saliencia resaltan las regiones de la entrada más relevantes para la tarea de clasificación (por ejemplo, lesiones malignas en mamografías) utilizando solo etiquetas de nivel de imagen (por ejemplo, si el paciente tiene cáncer o no) durante el entrenamiento [37].

Existen otros enfoques en la literatura donde la densidad mamaria se clasifica en cuatro categorías mediante un método tradicional y se estima el porcentaje de densidad mamaria mediante reglas de umbral, los autores también aplican un modelo de CNN a las máscaras binarias generadas [38].

Otro enfoque útil y comparable es usar redes convolucionales neuronales profundas, con la diferencia de combinar los grupos A y B y los grupos C y D, de esta manera, los primeros serían considerados como tejido graso y los segundos como tejido denso. En este caso los autores concluyen que las dCNN permiten una clasificación precisa de la densidad mamaria según el sistema ACR BI-RADS y que la técnica propuesta podría permitir una evaluación precisa, estandarizada e independiente del observador de la densidad mamaria en mamografías [38].

La evaluación de la densidad mamaria mediante enfoques como LIBRA y como MONAI ofrecen ventajas clínicas significativas, destacando su capacidad de segmentación y proporcionar valores porcentuales y categóricos útiles para los radiólogos. Aunque los modelos de aprendizaje profundo muestran prometedores avances en el diagnóstico, la interpretación de sus resultados sigue siendo desafiante, el uso de reglas de umbral y modelos de CNN en clasificación, así como la combinación de categorías para una evaluación más precisa, amplían el espectro de opciones para la evaluación de la densidad mamaria. Todos estos enfoques abordan una visión más estandarizada al observador de la imagen, reduciendo así el sesgo intrínseco presente.

---

## VII. CONCLUSIONES

Los resultados de concordancia entre las categorías ACR muestran una aceptable concordancia de LIBRA. Esto resalta la variabilidad que puede existir en la interpretación de las mamografías, incluso con un sistema estandarizado como BI-RADS. La subjetividad y la necesidad de análisis cuantitativos en tiempo real se hacen evidentes para mejorar la precisión.

LIBRA presenta un buen desempeño, especialmente en equipos de mamografía de las marcas Hologic y Carestream. Sin embargo, su rendimiento puede variar según la tecnología de adquisición. Este hallazgo destaca la importancia de considerar la diversidad de equipos en la implementación de modelos automáticos.

MONAI mostró un buen rendimiento en las categorías C y D, pero enfrentó desafíos en la diferenciación entre las categorías B y C. Esto plantea la necesidad de modelos que aborden la complejidad del tejido mamario heterogéneo. A pesar de sus fortalezas, la elección entre LIBRA y MONAI dependerá de las prioridades clínicas y las categorías de interés.

La alta densidad mamaria, un factor de riesgo conocido, destaca la importancia de herramientas como LIBRA y MONAI. Aunque ambos modelos muestran asociaciones significativas con el riesgo de cáncer de mama, sus enfoques y desempeños varían. La elección entre LIBRA y MONAI debe considerar las necesidades específicas de la aplicación clínica y las categorías de interés.

Dada la posibilidad de enmascaramiento de tumores en mamas densas, se enfatiza la importancia del seguimiento cercano a pacientes con alta densidad. MONAI, al destacarse en categorías densas, puede ser preferido en escenarios donde la detección de malignidades en tejido denso es crucial, a pesar de su desafío en categorías menos densas.

LIBRA presenta ventajas en la segmentación del tejido mamario, proporcionando un valor cuantitativo del porcentaje de densidad. Estas características adicionales son valiosas para el radiólogo, mejorando la evaluación detallada del tejido mamario y potencialmente optimizando la toma de decisiones en el diagnóstico.

## VIII. RECOMENDACIONES

Se sugiere explorar la implementación de técnicas de transfer learning en el modelo MONAI con el objetivo de mejorar las predicciones en las categorías A y B. Esto podría lograrse mediante la búsqueda y adición de más imágenes a la base de datos de entrenamiento, seguido de un proceso de reentrenamiento del modelo. La transferencia de conocimientos de modelos previamente entrenados podría ser clave para potenciar la capacidad de MONAI en la detección de patrones específicos en estas categorías A y B.

Para enriquecer el modelo y mejorar la precisión de las predicciones, se recomienda considerar la inclusión de variables clínicas relevantes. Factores como antecedentes del paciente, la clasificación BI-RADS, la edad, hábitos como fumar o beber, entre otros, podrían ser incorporados como características adicionales durante el entrenamiento del modelo. La integración de información clínica puede proporcionar una perspectiva más completa y precisa para la predicción de la densidad mamaria usando dichas variables junto con las imágenes.

Una línea de estudio interesante consiste en evaluar el rendimiento de MONAI al utilizar imágenes preprocesadas de DeepLibra, específicamente aquellas sin el músculo pectoral y con el fondo eliminado. Este enfoque tiene el potencial de mejorar la calidad de las imágenes de entrada, permitiendo a MONAI hacer predicciones más precisas en cada categoría. La comparación del rendimiento antes y después de la aplicación de estas imágenes preprocesadas podría revelar mejoras significativas.

---

REFERENCIAS

- [1] "Cáncer de mama - OPS/OMS | Organización Panamericana de la Salud". <https://www.paho.org/es/temas/cancer-mama> (accedido 6 de diciembre de 2023).
- [2] "Cáncer de mama". <https://www.who.int/es/news-room/fact-sheets/detail/breast-cancer> (accedido 6 de diciembre de 2023).
- [3] "Políticas basadas en la evidencia científica: El caso del control del cáncer de mama en Colombia". [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0123-90152020000300103#:~:text=En%202018%2C%20la%20tasa%20estimada,84%2C8%20por%20100.000%20mujeres](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-90152020000300103#:~:text=En%202018%2C%20la%20tasa%20estimada,84%2C8%20por%20100.000%20mujeres) (accedido 6 de diciembre de 2023).
- [4] C. Wiesner et al., "Políticas basadas en la evidencia científica: El caso del control del cáncer de mama en Colombia", Revista Colombiana de Cancerología, vol. 24, n.o 3, pp. 103-112, sep. 2020, doi: 10.35509/01239015.261.
- [5] S. Sharma y P. Khanna, "Computer-Aided Diagnosis of Malignant Mammograms using Zernike Moments and SVM", J Digit Imaging, vol. 28, n.o 1, pp. 77-90, feb. 2015, doi: 10.1007/s10278-014-9719-7.
- [6] V. Paulina Neira, "Densidad mamaria y riesgo de cáncer mamario", Revista Médica Clínica Las Condes, vol. 24, n.o 1, pp. 122-130, ene. 2013, doi: 10.1016/S0716-8640(13)70137-8.
- [7] S.-K. Lee et al., "A computer-aided design mammography screening system for detection and classification o Medical Informatics, vol. 60, n.o 1, pp. 29-57, oct. 2000, doi: 10.1016/S1386-5056(00)00067-8.
- [8] B. M. Keller et al., "Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation", Med Phys, vol. 39, n.o 8, pp. 4903-4917, ago. 2012, doi: 10.1118/1.4736530.
- [9] V. Paulina Neira, "Densidad mamaria y riesgo de cáncer mamario", Revista Médica Clínica Las Condes, vol. 24, n.o 1, pp. 122-130, ene. 2013, doi: 10.1016/S0716-8640(13)70137-8.
- [10] H. A. Martínez, D. N. P. Jiménez, L. M. M. Quesada, Y. R. Quintian, C. A. P. Suárez, y C. W. Ceballos, "ANÁLISIS DE CONCORDANCIA EN LA INTERPRETACIÓN DE IMÁGENES MAMOGRÁFICAS ENTRE MÉDICOS RADIÓLOGOS DE BOGOTÁ, D. C.", Medicina, vol. 40, n.o 2, Art. n.o 2, jun. 2018.
- [11] O. Haji Maghsoudi et al., "Deep-LIBRA: An artificial-intelligence method for robust

- 
- quantification of breast density with independent validation in breast cancer risk assessment", *Med Image Anal*, vol. 73, p. 102138, oct. 2021, doi: 10.1016/j.media.2021.102138.
- [12] "LIBRA | Section for Biomedical Image Analysis (SBIA) | Perelman School of Medicine at the University of Pennsylvania". [https://www.med.upenn.edu/sbia/libra.html#:~:text=The%20Laboratory%20for%20Individualized%20Breast,either%20raw%20\(i.e.%2C%20%E2%80%9CFOR](https://www.med.upenn.edu/sbia/libra.html#:~:text=The%20Laboratory%20for%20Individualized%20Breast,either%20raw%20(i.e.%2C%20%E2%80%9CFOR) (accedido 6 de diciembre de 2023)
- [13] "Mamografías - NCI", 9 de enero de 2017. <https://www.cancer.gov/espanol/tipos/seno/hoja-informativa-mamografias> (accedido 6 de diciembre de 2023).
- [14] "Sistema de Información de la Investigación - HERMES". <http://www.hermes.unal.edu.co/pages/Consultas/Proyecto.xhtml?idProyecto=85755> (accedido 6 de diciembre de 2023).
- [15] "Estudio de Casos y Controles: Cáncer de Mama y su Asociación con la Densidad Mamaria Incrementada Visualizada por Mamografía en el Instituto del Cáncer SOLCA Cuenca – Ecuador". Accedido: 18 de julio de 2023. [En línea]. Disponible en: <https://docs.bvsalud.org/biblioref/2022/11/1016050/85-texto-del-articulo-143-1-10-20200122.pdf#:~:text=La%20mamograf%C3%ADa%20es%20el%20principal,visuali%2D%20zaci%C3%B3n%20de%20c%C3%A1ncer%20subyacente>
- [16] "SciELO - Salud Pública - Errores de medición en la interpretación mamográfica por radiólogos Errores de medición en la interpretación mamográfica por radiólogos". <https://www.scielosp.org/article/rsap/2018.v20n4/518-522/> (accedido 6 de diciembre de 2023).
- [17] "Fortalecer detección temprana de cáncer de mama, un desafío". <https://www.minsalud.gov.co/Paginas/Fortalecer-deteccion-temprana-de-cancer-de-mama-un-desafio.aspx> (accedido 6 de diciembre de 2023).
- [18] "Tiempo para ti". <https://www.epssura.com/tiempoparati/> (accedido 6 de diciembre de 2023).
- [19] "Por qué importa la densidad mamaria | Breastcancer.org". <https://www.breastcancer.org/es/riesgo/factores-riesgo/mamas-densas> (accedido 6 de diciembre de 2023).
- [20] "Fully Automated Breast Density Segmentation and Classification Using Deep Learning - PMC". <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7700286/> (accedido 6 de diciembre de 2023).
- [21] "Clasificación radiológica y manejo de las lesiones mamarias | Clínica e Investigación en Ginecología y Obstetricia". <https://www.elsevier.es/es-revista-clinica-e-investigacion-ginecologia-obstetricia-7-articulo-clasificacion-radiologica-manejo-lesiones-mamarias-S0210573X10001231> (accedido 6 de diciembre de 2023).
- [22] "Prevalência de tejido mamario denso en una población de Bucaramanga, Colombia".

- 
- [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0123-90152020000300124](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-90152020000300124) (accedido 6 de diciembre de 2023).
- [23] "Mammographic Breast Density as a General Marker of Breast Cancer Risk | Cancer Epidemiology, Biomarkers & Prevention | American Association for Cancer Research". <https://aacrjournals.org/cebpa/article/16/1/43/276591/Mammographic-Breast-Density-as-a-General-Marker-of> (accedido 6 de diciembre de 2023).
- [24] "¿Son sus senos densos?", George Washington University Hospital, 30 de septiembre de 2015. <https://es.gwhospital.com/services/breast-center/cancer-breast-density> (accedido 6 de diciembre de 2023).
- [25] D. R. M. Rojas, "¿Qué significa ACR C en una mamografía?", Diagnóstico Rojas, 18 de febrero de 2020. <https://www.diagnosticorojas.com.ar/blog/salud/que-significa-acr-c/> (accedido 6 de diciembre de 2023).
- [26] G. W. Haggstrom, "Logistic Regression and Discriminant Analysis by Ordinary Least Squares", *J. Bus. & Econ. Statist.*, vol. 1, n.º 3, p. 229, julio de 1983. Accedido el 1 de enero de 2024. [En línea]. Disponible: <https://doi.org/10.2307/1391344>
- [27] "GitHub - CBICA/Deep-LIBRA". GitHub. Accedido el 6 de diciembre de 2023. [En línea]. Disponible: <https://github.com/CBICA/Deep-LIBRA>
- [28] "ArXiv", *Choice Rev. Online*, vol. 45, n.º 02, pp. 45–0602–45–0602, octubre de 2007. Accedido el 6 de diciembre de 2023. [En línea]. Disponible: <https://doi.org/10.5860/choice.45-0602>
- [29] "MONAI Model Zoo". MONAI - Home. Accedido el 6 de diciembre de 2023. [En línea]. Disponible: <https://monai.io/model-zoo.html>
- [30] J. J. Diamond, "Cohen's kappa", *J. Clin. Epidemiol.*, vol. 44, n.º 6, p. 609, enero de 1991. Accedido el 6 de diciembre de 2023. [En línea]. Disponible: [https://doi.org/10.1016/0895-4356\(91\)90224-w](https://doi.org/10.1016/0895-4356(91)90224-w)
- [31] M. Calas, R. Almeida, B. Gutfilen y W. Pereira, "Interobserver concordance in the BI-RADS classification of breast ultrasound exams", *Clinics*, vol. 67, n.º 2, pp. 185–189, febrero de 2012. Accedido el 6 de diciembre de 2023. [En línea]. Disponible: [https://doi.org/10.6061/clinics/2012\(02\)16](https://doi.org/10.6061/clinics/2012(02)16)
- [32] J. S. Uebersax, "A Generalized Kappa Coefficient", *Educational Psycholog. Meas.*, vol. 42, n.º 1, pp. 181–183, marzo de 1982. Accedido el 6 de diciembre de 2023. [En línea]. Disponible: <https://doi.org/10.1177/0013164482421018>
- [33] E. Badawy, R. ElNaggar, S. A. M. Soliman y D. S. Elmesidy, "Performance of AI-aided mammography in breast cancer diagnosis: Does breast density matter?", *Egyptian J. Radiol. Nucl. Medicine*, vol. 54, n.º 1, octubre de 2023. Accedido el 6 de diciembre de 2023. [En línea]. Disponible: <https://doi.org/10.1186/s43055-023-01129-3>
- [34] L. A. Habel et al., "Examination of fully automated mammographic density measures using

- 
- LIBRA and breast cancer risk in a cohort of 21,000 non-Hispanic white women”, *Breast Cancer Res.*, vol. 25, n.º 1, agosto de 2023. Accedido el 6 de diciembre de 2023. [En línea]. Disponible: <https://doi.org/10.1186/s13058-023-01685-6>
- [35] Z. Y. Lee, Y. L. E. Goh y C. Lai, “Classification of mammographic breast density and its correlation with BI-RADS in elder women using machine learning approach”, *J. Med. Imag. Radiat. Sci.*, vol. 53, n.º 1, pp. 28–34, marzo de 2022. Accedido el 6 de diciembre de 2023. [En línea]. Disponible: <https://doi.org/10.1016/j.jmir.2021.10.004>
- [36] “RSNA Screening Mammography Breast Cancer Detection | Kaggle”. Kaggle: Your Machine Learning and Data Science Community. Accedido el 6 de diciembre de 2023. [En línea]. Disponible: <https://www.kaggle.com/competitions/rsna-breast-cancer-detection/discussion/392449>
- [37] “Weakly-supervised High-resolution Segmentation of Mammography Images for Breast Cancer Diagnosis”. PubMed Central (PMC). Accedido el 6 de diciembre de 2023. [En línea]. Disponible: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8791642/>
- [38] N. Saffari et al., “Fully Automated Breast Density Segmentation and Classification Using Deep Learning”, *Diagnostics*, vol. 10, n.º 11, p. 988, noviembre de 2020. Accedido el 6 de diciembre de 2023. [En línea]. Disponible: <https://doi.org/10.3390/diagnostics10110988>
- [39] A. Ciritsis et al., “Determination of mammographic breast density using a deep convolutional neural network”, *Brit. J. Radiol.*, p. 20180691, octubre de 2018. Accedido el 6 de diciembre de 2023. [En línea]. Disponible: <https://doi.org/10.1259/bjr.20180691>