

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

**Segmentación de Masas de Cáncer de Mama Utilizando
Procesamiento Digital de Imágenes**

Camila Zambrano Campoverde

Carrera de Ingeniería en Ciencias de la Computación

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniería en Ciencias de la Computación

Quito, 13 de diciembre de 2023

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

**HOJA DE CALIFICACIÓN
DE TRABAJO DE FIN DE CARRERA**

**Segmentación de Masas de Cáncer de Mama Utilizando Procesamiento
Digital de Imágenes**

Camila Zambrano Campoverde

Nombre del profesor, Título académico

Noel Pérez Pérez, Ph.D.

Quito, 13 de diciembre de 2023

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Camila Zambrano Campoverde

Código: 000209260

Cédula de identidad: 1723822589

Lugar y fecha: Quito, 13 de diciembre de 2023

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETHeses>.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETHeses>.

RESUMEN

La detección, identificación temprana y clasificación de masas de cáncer de mama en mamografías son tareas vitales para reducir la mortalidad de los pacientes. Estos procesos permiten analizar las etapas iniciales del cáncer antes de la manifestación clínica. La segmentación efectiva, que separa la lesión del fondo, mejora la clasificación basada en la forma de la lesión. Sin embargo, este es un proceso complejo debido a la similitud entre la lesión y el tejido circundante. Por lo tanto, este proyecto propone explorar los modelos de contornos activos Geodésico y Chan-Vese para optimizar la segmentación de masas en imágenes de mamografía. Se utilizó para ajustar y validar ambos modelos un conjunto de datos experimental de 115 imágenes que contenían lesiones de masas, las cuales fueron obtenidas del conjunto de datos INbreast. El modelo seleccionado fue el mejor modelo Chan-Vese, con un radio de 50 píxeles y 436 iteraciones, el cual superó el mejor modelo Geodésico, obteniendo un promedio de Dice de 0.812 contra un 0.558. En general, el método propuesto segmentó con éxito masas con formas complejas e irregulares, un resultado crucial para la clasificación de masas malignas. Sin embargo, los resultados podrían mejorarse en un futuro aplicando técnicas adicionales de preprocesamiento de imágenes para mejorar la visualización de las lesiones de masa.

Palabras clave: Cáncer de Mama, Contornos Activos, Geodésico, Chan-Vase, segmentación de masas, Dice, IoU.

ABSTRACT

Breast cancer mass lesion detection, early identification, and classification in mammograms are vital for reducing patient mortality. These processes enable the analysis of cancer's initial stages before clinical manifestation. Effective segmentation, separating the lesion from the background, enhances shape-based classification. However, this is a challenging process due to tissue similarities between the lesion and the background. Therefore, this project proposes exploring the Geodesic and Chan-Vese active contour models to optimize mass segmentation in mammography images. Both models were fine-tuned and validated using an experimental dataset of 115 images with mass lesions, obtained from the INbreast Dataset. The selected model was the best Chan-Vese model, with a radius of 50 pixels and 436 iterations, which outperformed the best Geodesic model, achieving a mean Dice score of 0.812 versus 0.558. Overall, the proposed method successfully segmented complex and irregular mass shapes, a crucial outcome for malignancy classification. Nevertheless, the results could be further improved in the future by applying additional image preprocessing techniques to enhance the visualization of the mass lesions.

Keywords: Breast Cancer, Active Contours, Geodesic, Chan-Vese, mass lesions segmentation, Dice, IoU.

TABLA DE CONTENIDO

1. Introducción.....	10
2. Desarrollo del Tema.....	15
2.1 Materiales y Métodos.....	15
2.1.1. Bases de datos.....	15
2.1.2. Método propuesto.....	15
2.2. Setup Experimental.....	17
2.2.1. Creación del dataset experimental.....	17
2.2.2. Configuración de los modelos.....	18
2.2.3. Métricas de validación y criterio de selección.....	18
2.2.4. Plataformas de desarrollo.....	19
3. Resultados y Discusión.....	20
4. Conclusiones y Trabajo Futuro.....	25
Referencias Bibliográficas.....	28

ÍNDICE DE TABLAS

Tabla 1. Rendimiento promedio de segmentación del método propuesto en el conjunto de datos experimentales.....	20
--	----

ÍNDICE DE FIGURAS

Figura 1. Ejemplos aleatorios de imágenes de mamografías tomado de la base de datos INbreast con proyecciones MLO y CC.....	15
Figura 2. Flujo de trabajo experimental del método propuesto.....	16
Figura 3. Optimización de radios para el modelo Geodésico.....	21
Figura 4. Optimización de radios para el modelo Chan-Vase.....	22
Figura 5. Resultados de la segmentación de masas mediante el método propuesto.....	23

1. INTRODUCCIÓN

El cáncer es una enfermedad compleja impulsada por el crecimiento y la división descontrolados de las células, lo que interrumpe el proceso natural de renovación celular y da como resultado la formación de tumores, los cuales pueden ser malignos o benignos (National Institutes of Health (US), 2007), (National Cancer Institute, 2021). Los tumores malignos invaden los tejidos cercanos y pueden metastatizar a otras partes del cuerpo, mientras que los tumores benignos permanecen localizados y generalmente se pueden extraer exitosamente (National Cancer Institute, 2021). De entre todos los cánceres, el cáncer de mama es uno de los más comunes en mujeres alrededor del mundo, tan solo en 2020, la Organización Mundial de la Salud reportó 2.3 millones de nuevos diagnósticos a nivel global. Diagnósticos que resultaron en 685,000 muertes (WHO, 2023). Adicionalmente, este cáncer es particularmente prevalente, afectando alrededor de 7.8 millones de mujeres durante los últimos cinco años (WHO, 2023). Específicamente, en Ecuador, el cáncer de mama contribuye al 1.15% del total de muertes a nivel mundial (WHO, 2023), y es la 3era causa de muerte en mujeres (Ministerio de la Salud Pública (MSP), 2022). En base a estos datos, se puede determinar que el cáncer de mama es un problema muy importante actualmente.

Continuando, dado a que el riesgo de cáncer de mama aumenta con la edad, la detección y el diagnóstico tempranos son fundamentales para reducir la tasa de mortalidad. Para la detección temprana los expertos suelen recomendar técnicas estándares como, autoexámenes regulares de la mama (Huang et al., 2022), exámenes clínicos de la mama (Huang et al., 2022), mamografías (Nounou et al., 2015), lectura doble de mamografías (Taylor-Phillips & Stinton, 2020), ultrasonidos de la mama (Ozmen et al., 2015), resonancias magnéticas (MRI) (Nounou et al., 2015), o biopsias (Nyström et al., 1993), siendo esta última un método invasivo. A pesar de tener tantas técnicas, las mamografías son una de las maneras más comunes de detectar

cáncer de mama, ya que es posible detectar la etapa inicial de una lesión maligna antes de que esta aparezca clínicamente. Sin embargo, a pesar de los esfuerzos de los métodos tradicionales por reducir las detecciones que dan falsos positivos o negativos, todavía existen diagnósticos erróneos. Esto se debe principalmente a errores humanos relacionados con la sobrecarga de trabajo, la fatiga y la naturaleza intrínseca del análisis de las características del tejido mamario, como la densidad mamaria, la variación en las ubicaciones de las lesiones, su forma, al igual que su complejidad textural (Pérez, 2015).

Varios enfoques informáticos basados en técnicas de procesamiento de imágenes digitales (Liu et al., 2015)(Niaz et al., 2020)(Radhi & Kamil, 2032)(Alhadidi et al., 2007) y modelos de aprendizaje profundo (Ekici & Jawzal, 2020)(Baccouche et al., 2021)(Su et al., 2022) se han centrado en la tarea de detección automática, segmentación y clasificación del cáncer de mama para ayudar a los expertos a analizar mamografías y proporcionarles una segunda opinión sobre posibles anomalías. Por ejemplo, se propuso un modelo de contorno activo geodésico local (LGAC) basado en información local de la imagen para segmentar masas en la Digital Database for Screening Mamography (DDSM) (Liu et al., 2015). Los resultados experimentales sugirieron que el modelo puede encontrar correctamente los bordes radiales y ambiguos de las masas, logrando puntuaciones en la métrica de superposición de área (AOM) de 0.85 y 0.87, para radios de 4 mm y 6 mm, respectivamente. En (Niaz et al., 2020), se propuso un método de contorno activo (AC) basado en una función combinada local y global reformada para abordar la segmentación de tumores de cáncer de mama. El modelo se validó en la base de datos mini Mammographic Image Analysis Society (MIAS), obteniendo una precisión (ACC) de 0.984, una sensibilidad del 0.9802 y una puntuación de Dice de 0.985. Además, en (Radhi & Kamil, 2021) se propusieron modelos de snakes y de level set basados en Chan-Vese para detectar el cáncer de mama en la base de datos mini MIAS. Los resultados resaltaron el modelo de snakes como el mejor método, alcanzando un puntaje promedio de IoU

(Intersección sobre Unión), Dice, media armónica de precisión y sensibilidad (PF-score o PFS), precisión (PRE) y sensibilidad del 90%, 95%, 98%, 97% y 97%, respectivamente.

Adicionalmente, se propuso un modelo de CNN optimizado por el algoritmo de Bayes para detectar y diagnosticar el cáncer de mama en (Ekici & Jawzal, 2020). El modelo se validó en un conjunto de datos en bruto que contenía 3895 imágenes térmicas de mama, produciendo un puntaje de ACC del 98.95%. En (Baccouche et al., 2021), se desarrollaron dos modelos U-NET (Atención y Residual UNet) integrados con la técnica de Atrous Spatial Pyramid Pooling (ASPP) para detectar y segmentar tumores de cáncer de mama en tres conjuntos de datos diferentes (una versión de la DDSM, el dataset INbreast y una base de datos privada). El modelo ResUNet (Unet residual) logró los mejores resultados de segmentación de masas con puntuaciones de Dice del 89.52%, 95.28% y 95.88% y puntuaciones de IoU del 80.02%, 91.03% y 92.27%, para los conjuntos de datos utilizados, respectivamente. Del mismo modo, en (Su et al., 2022), se propuso un modelo de detección y segmentación basado en la arquitectura YOLO. El modelo se validó en los conjuntos de datos DDSM e INbreast, alcanzando una tasa de verdaderos positivos (TPR) del 95.7% y una media de precisión (PRE) del 65.0% para la detección de masas. Además, alcanzó puntuaciones de F1-score y IoU del 74.5% y 64%, respectivamente, para la tarea de segmentación de masas.

Dado el contexto de los enfoques desarrollados, la mayoría de los métodos se centran en tareas de detección y clasificación del cáncer de mama, ya que son menos complejas que la segmentación. Por lo tanto, la segmentación de masas, ya sea mediante el procesamiento digital de imágenes o técnicas de aprendizaje profundo, sigue siendo un desafío debido a la similitud entre la lesión y el tejido circundante. Por lo tanto, se propone en este proyecto explorar dos modelos de contorno activo (Geodésico y Chan-Vase) para maximizar el rendimiento de segmentación de masas en imágenes de mamografías. La contribución de este trabajo se

considera un paso inicial esencial para integrarlo en un futuro a un detector de masas y ayudar a los expertos con una herramienta que sirva de segunda opinión para el diagnóstico.

Cabe destacar que también existen otras técnicas de procesamiento de imágenes digitales que comúnmente se usan en la segmentación, tales como, Watershed, Histogram-based o el Gradient Vector Flow. Sin embargo, en comparación a los contornos activos estos métodos no son una buena opción para resolver el problema planteado. En cuanto al Watershed, este crea un gráfico 3D de la imagen, tomando como altura la intensidad de los píxeles, e inunda las cuencas desde los marcadores, las cuales en su mayoría se eligen con los mínimos locales de la imagen, hasta que las cuencas atribuidas a diferentes marcadores se encuentran en las líneas de cuenca, segmentando así el borde del objeto de interés (Scikit-Image, s/f). El problema con este método es que al parar la segmentación al momento que se encuentran las cuencas atribuidas a diferentes marcadores la segmentación no es exacta (Scikit-Image, s/f). Dentro del contexto del problema a tratar en este trabajo, el contorno de la masa es esencial para determinar el grado de malignidad de la masa, por ende, debe ser lo más exacto posible.

Por otro lado, la técnica del Histogram-based utiliza el histograma de una imagen para agrupar los píxeles en regiones en base a los niveles de grises (Phillips, s/f), el objeto de interés va a tener unos niveles de grises y el fondo otros. El problema con este método es que no hace una segmentación semántica, esto se refiere a que le cuesta distinguir entre el objeto de interés y el fondo (Phillips, s/f). Esto se debe a que si bien trata de agrupar los píxeles que piensa que son fondo por un lado y los que son el objeto de interés por otro, tiende a confundirse si el objeto es similar al fondo (Phillips, s/f). Situación que ocurre en mamografías, ya que el tejido circundante es muy parecido a las lesiones de masas, como se mencionó previamente.

Finalmente, la técnica del Gradient Vector Flow consiste (Xu, 1997). Comúnmente se suelen utilizar en conjunto a los contornos activos para (Xu, 1997). Sin embargo, el principal problema de esta técnica es que los resultados son muy sensibles a la inicialización de los

vectores (Xu, 1997). Si estos no se inicializan correctamente el contorno puede que nunca converja al borde del objeto de interés o por el contrario converja más allá del borde el objeto de interés. Es por esto que se necesita un alto nivel de preprocesamiento de la imagen previo, aumentando significativamente el costo computacional del modelo (Xu, 1997). Debido a que la propuesta de este proyecto está enfocada en poder ayudar a los expertos con una segunda opinión, el presupuesto de los hospitales o centros médicos no es tan alto como para invertir en una infraestructura fuerte para correr un modelo muy complejo, descartando el uso de esta técnica.

En base a las desventajas de las técnicas explicadas previamente y tomando en cuenta las ventajas que tienen los modelos de contornos activos tales como: su robustez debido a que se pueden adaptar con facilidad a patrones homogéneos o irregulares, y su bajo poder computacional gracias a que implementan cálculos matemáticos sencillos, computacionalmente hablando (Menet et al., 2002), los contornos activos son los modelos más viables para la solución que se propone para el problema dado el contexto.

2. DESARROLLO DEL TEMA

2.1. Materiales y método

2.1.1. Base de datos.

En este trabajo se consideró la base de datos de acceso público INbreast, que contiene 410 mamografías digitales de vista completa y alta resolución, con proyecciones craneocaudales (CC) y medio laterales oblicuas (Moreira et al., 2012). Esta base de datos documenta varias observaciones, como asimetrías, calcificaciones, grupos de microcalcificaciones, distorsiones, regiones especulativas, músculo pectoral, y lesiones de masas basadas en anotaciones de la curva de la región verdadera y clasificaciones (BI-RADS¹ y binarias) realizadas por expertos. A continuación, se muestra un ejemplo de imágenes de mamografías con hallazgos de lesiones de masas y sus anotaciones de la región verdadera en la Figura #1.

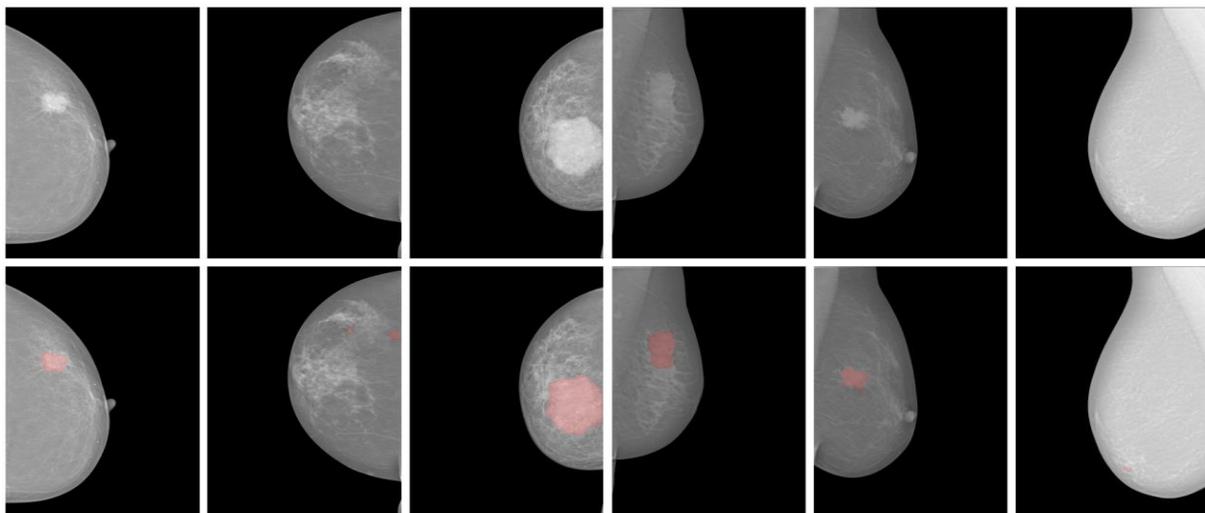


Figura 1. Ejemplos aleatorios de imágenes de mamografías tomado de la base de datos INbreast con proyecciones MLO y CC: imagen original (primera fila) y anotación de la región verdadera (área en rojo) realizada por expertos (segunda fila).

2.1.2. Método propuesto.

El método propuesto es un procedimiento de dos pasos, como se muestra en la Figura 2, e incluye: primero una etapa de filtrado utilizando la técnica de ecualización adaptativa del

El BI-RADS es una clasificación hecha por expertos en el campo de cáncer de mama para determinar el grado de malignidad de una masa en base a la forma del contorno de la misma.

histograma con límite de contraste (CLAHE) para reducir la amplificación del ruido limitando la amplificación del contraste. Y también, aplicando un filtro de mediana con una máscara $[5 \times 5]$ para minimizar el ruido aleatorio (por ejemplo, sal y pimienta) mientras se conserva el borde del objeto. La segunda etapa explora dos modelos de contorno activo, el Geodésico (Caselles et al., 1997) y el Chan-Vese (Chan & Vese, 2001), para refinar y maximizar el rendimiento de la segmentación de las lesiones de masas en imágenes de mamografías.

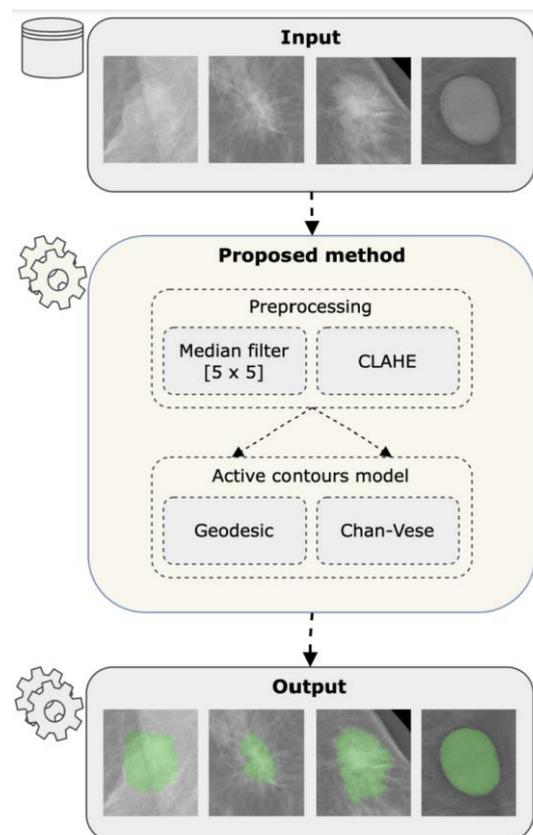


Figura 2. Flujo de trabajo experimental del método propuesto

Un modelo de contorno activo es una curva continua y deformable cuya forma está controlada por fuerzas internas (el modelo implícito) y fuerzas externas (los datos) (Menet et al., 2002). Las fuerzas internas actúan como una restricción de suavidad. Por otro lado, las fuerzas externas guían el contorno activo hacia las características de los objetos presentes en las imágenes, como líneas, bordes y contornos subjetivos, ubicando el modelo de contorno

activo cerca del mínimo local deseado (Kass et al., 1988). El modelo geodésico mejora los contornos activos al incorporar información geométrica intrínseca de la imagen que está bajo análisis. El principio fundamental detrás de este modelo es deformar el contorno de interés tomando en cuenta la distancia más corta entre dos puntos a lo largo de la superficie curva de la imagen (Caselles et al., 1997). El cálculo de la distancia considera el espacio de Riemann subyacente, es decir, considera una construcción matemática que define la distancia teniendo en cuenta el contenido de la imagen (Caselles et al., 1997).

Por otro lado, el modelo Chan-Vese introduce un enfoque innovador que aprovecha la evolución de curvas, la funcionalidad Mumford-Shah para la segmentación y las técnicas de nivel de conjuntos, las cuales no dependen de información basada en gradientes (Chan & Vese, 2001). La funcionalidad Mumford-Shah es una formulación matemática utilizada para la segmentación de imágenes. Esta se centra en minimizar la energía asociada con la detección de los contornos de los objetos (Chan & Vese, 2001). La formulación del nivel de conjuntos implica definir la evolución de los contornos como un flujo de curvatura media (Chan & Vese, 2001). Esto significa que el contorno activo evoluciona gracias a un proceso basado en la curvatura, el cual se ajusta a lo largo de la misma en base a la posición del contorno en cada punto (Chan & Vese, 2001).

2.2. Setup de los experimentos

2.2.1. Creación de la dataset experimental.

Dado a que este trabajo tiene como objetivo maximizar la segmentación de lesiones de masas, consideramos solo 115 lesiones de masa extraídas de 106 imágenes de mamografías de la base de datos. Posteriormente, estas imágenes se recortaron utilizando un bounding box que contiene la región verdadera, generando así regiones de interés (ROIs) con la lesión. De esta manera, se creó un conjunto de datos experimentales de 115 imágenes con los ROIs.

2.2.2. Configuración de los modelos de contornos activos.

Los modelos de contorno activo utilizados requieren un punto de inicial de coordenadas, un radio y un número de iteraciones para comenzar la deformación de la curva. El punto inicial de coordenadas se inicializó en el centro del ROI de entrada, asegurando un punto dentro de la masa. El radio se optimizó en el intervalo de 10 a 100, con pasos de 10 píxeles, y el número de iteraciones, el cual indica cuantas veces se va a deformar la curva, se estableció de manera empírica en 500 unidades.

2.2.3. Métrica de validación y criterio de selección.

Se calculó las métricas de IoU (Rosebrock, 2023) y Dice (OECD.AI, s.f.) para evaluar la calidad de la segmentación del método propuesto. Ambas métricas miden la proporción de precisión entre la región segmentada por el método y la región verdadera anotada por los expertos. Los resultados obtenidos se expresan en un rango de 0 a 1. Cuanto más cercana sea la puntuación al 1, mejor será el rendimiento de la segmentación. La formulación matemática de estas métricas es la siguiente:

$$IOU = \frac{(A \cap B)}{(A \cup B)} \quad (1)$$

$$Dice = \frac{2 \cdot (A \cap B)}{(A + B)} \quad (2)$$

donde A y B son, el área predicha y el área verdadera, respectivamente.

Ambas métricas se utilizarán para describir los resultados obtenidos. Sin embargo, el modelo de contornos activos que maximiza la métrica de Dice con menos iteraciones será el modelo seleccionado para la propuesta. Se escogió Dice como la métrica principal en lugar de IoU porque este es más sensible a cambios muy pequeños en la segmentación, una ventaja al analizar lesiones de cáncer de mama. Adicionalmente, esta métrica trata los falsos positivos y negativos de manera equitativa, evitando cualquier tipo de sesgo. En contraste, IoU está

ligeramente sesgado hacia los falsos negativos, por la operación de intersección que se hace en el denominador (Wang et al., 2020).

2.2.4. Plataforma de desarrollo.

La etapa de preprocesamiento del método propuesto se realizó con MATLAB versión R2023a, y los modelos de contorno activo se implementaron y ejecutaron en Python versión 3.9 utilizando varias bibliotecas distribuidas en (Álvarez et al., 2010) y (Márquez-Neila et al., 2014).

3. RESULTADOS Y DISCUSIÓN

En base a la sección anterior, se evaluó el método propuesto en un conjunto de datos experimental de 106 mamografías con 115 regiones de interés (ROIs) de lesiones de masas. Los resultados obtenidos, basados en el promedio de las métricas de Dice y IoU, se presentan a continuación en la Tabla #1.

Tabla #1. Mejor rendimiento promedio de la segmentación del método propuesto

Radio	Modelos					
	Geodésico			Chan-Vase		
	Iteración	Dice	IoU	Iteración	Dice	IoU
10	153	0.539	0.395	468	0.810	0.695
20	142	0.553	0.406	458	0.810	0.695
30	114	0.552	0.407	430	0.810	0.695
40	175	0.558	0.410	436	0.812	0.698
50	88	0.558	0.415	410	0.811	0.696
60	91	0.557	0.412	398	0.808	0.693
70	75	0.556	0.413	384	0.803	0.688
80	63	0.555	0.412	450	0.788	0.671
90	55	0.554	0.412	435	0.744	0.655
100	43	0.553	0.410	334	0.768	0.649

Todos los modelos basados en los contornos activos Geodésicos proporcionaron desempeños de segmentación que oscilaron entre puntuaciones promedio de 0.55 y 0.56 en las métricas de Dice e IoU, respectivamente (consultar Tabla #1). El modelo alcanzó la puntuación promedio de Dice más alta utilizando un radio de 50 píxeles con 88 iteraciones, obteniendo 0.558 y un IoU de 0.415 (Tabla #1). Estos resultados indican que las segmentaciones de las masas en su mayoría fueron imprecisas. Algo interesante a descartar en este modelo es que mientras la longitud del radio y las iteraciones avanzan hacia el centro del intervalo del radio analizado (50 píxeles), el rendimiento de las métricas también aumenta. Sin embargo, al pasar este centro (50 píxeles) y hasta el final del intervalo (100 píxeles), el rendimiento de la segmentación comienza a disminuir. Este efecto está vinculado a la naturaleza de la

deformación del modelo Geodésico, el cual expande la deformación de la curva en cada iteración. Por esto, las longitudes iniciales del radio (antes del centro del intervalo) con más iteraciones tienden a proporcionar mejores resultados. Esto se puede corroborar al analizar la Fig. 3, la cual muestra el rendimiento del modelo geodésico a lo largo de la variación del radio para ambas métricas.

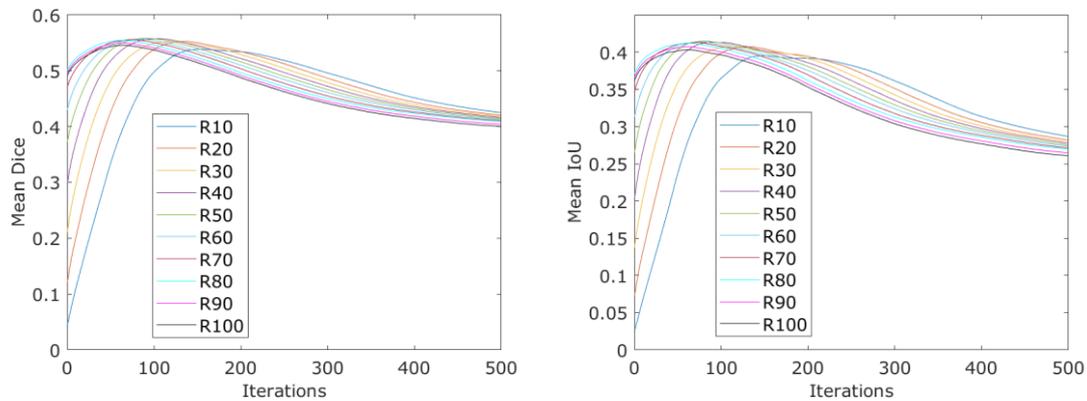


Figura 3. Optimización de radios para el modelo Geodésico: con los promedios de las métricas Dice (izquierda) y IoU (derecha)

Por otro lado, los modelos basados en los contornos activos Chan-Vese lograron realizar la segmentación de masas con resultados promedios de Dice superiores a 0.80 (consultar Tabla #1), resultados que se pueden considerar razonables para el problema en cuestión. El mejor modelo utilizó un radio de 40 píxeles y 436 iteraciones, alcanzando la puntuación promedio de Dice más alta, siendo 0.812 y un IoU de 0.698. Estos resultados demuestran que las segmentaciones de las masas en su mayoría fueron exitosas. Cabe destacar que este modelo tuvo un buen rendimiento de las métricas con un alto número de iteraciones. Estos resultados se explican por el proceso de deformación que implementa el modelo Chan-Vese. Este contrae la deformación de la curva en cada iteración, buscando el límite de la lesión de masa. Este comportamiento se puede observar para ambas métricas en la Fig. 4.

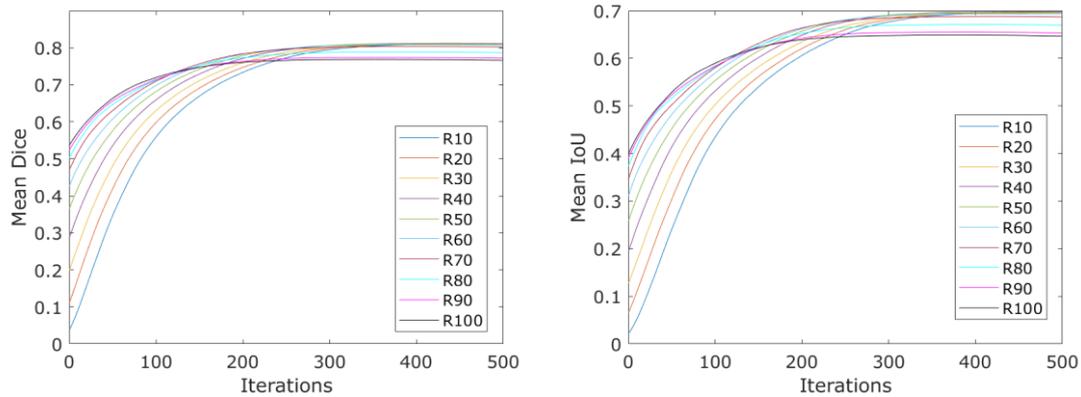


Figura 4. Optimización de radios para el modelo Chan-Vese: con los promedios de las métricas Dice (izquierda) y IoU (derecha)

En cuanto a la comparación entre ambos modelos, el modelo Chan-Vese superó, en términos de resultados, al modelo Geodésico para todas las configuraciones de radio. Esto ocurre principalmente debido a la implementación del modelo Chan-Vese de la minimización de la energía funcional Mumford-Shah, lo que contribuye a su capacidad para manejar variaciones en la apariencia de los objetos, haciéndolo más robusto para tareas de segmentación en escenarios complejos (Chan & Vese, 2001, Menet et al., 2002). Además, el manejo de diversas apariencias de los objetos a través del enfoque de minimización de energía incorpora información de región y contorno del objeto lo cual mejora la segmentación y supera los mínimos locales (Chan & Vese, 2001). Por el contrario, el modelo Geodésico tiende a ser sensible a los mínimos locales (Caselles et al., 1997), por ende, se le dificulta la tarea de segmentación en ciertas condiciones texturales de una imagen, como, por ejemplo, el tejido circundante de las lesiones de masas. Adicionalmente, lo previamente mencionado sobre los modelos se puede corroborarse a través de los resultados de segmentación obtenidos por los mejores modelos Geodésico (radio 50, 88 iteraciones) y Chan-Vese (radio 40, 436 iteraciones) ilustrados en la Fig. 5.

A partir de esta figura, es posible notar que el modelo Geodésico desborda el área segmentada fuera de la lesión en algunos casos (consultar Fig. 5, segunda fila, columnas 1 y 3)

y queda dentro de la lesión sin llegar al borde en otros (consultar Fig. 5, segunda fila, columnas 2, 4, 5 y 6). En todos los casos, el modelo Geodésico segmenta lejos del verdadero borde de la masa. En contraste, el modelo Chan-Vese alcanza el verdadero contorno en casi todos los casos, demostrando su superioridad en términos de segmentación. Por lo tanto, basándose en los resultados expuesto previamente, el modelo Chan-Vese con un radio de 40 píxeles y 436 iteraciones es seleccionado como el mejor modelo para la propuesta de este proyecto.

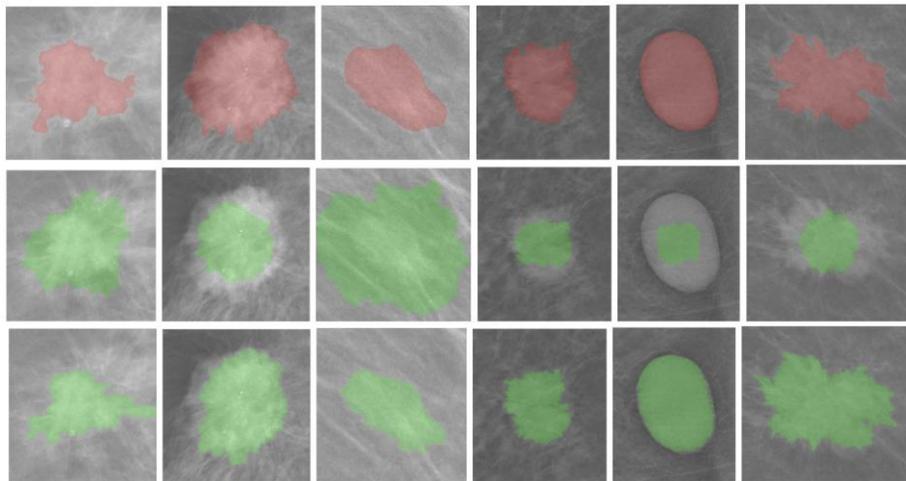


Figura 5. Resultados de la segmentación de masas mediante el método propuesto: imagen original con, la región verdadera (primera fila) y la lesión de masa segmentada por los modelos de contornos activos Geodésico (segunda fila) y Chan-Vese (tercera fila).

Adicionalmente, en cuanto a la comparación con el estado del arte, el modelo basado en U-Net desarrollado en (Baccouche et al., 2021) fue entrenado y validado en un conjunto de datos experimental similar al de este proyecto, extraído de la misma base de datos INbreast. Los puntajes de IoU y Dice obtenidos fueron del 91.03% y 95.28%, respectivamente. Estos resultados fueron superiores a los obtenidos por el método propuesto en este proyecto. Sin embargo, la complejidad del algoritmo y los requisitos de hardware también fueron superiores. Siendo una desventaja en términos de aplicabilidad y despliegue práctico. Dado a que los modelos U-Net integran las tareas de detección y segmentación en un único proceso proporcionado por una espina dorsal de CNN profunda, son computacionalmente intensivos y

propensos al sobreajuste, especialmente si la representación de datos es pequeña. Lo que, también hace que la comparación no pueda ser directa como tal, ya que el método propuesto no hace detección. Además, el rendimiento de la segmentación siempre depende de la calidad de los datos de entrenamiento, que en este caso dependen de la región verdadera creada por los expertos. En contraste, el método propuesto para este proyecto aprovecha las ROIs extraídos para iterar un modelo de contorno activo el cual maneja la segmentación de masas con formas complejas y desconocidas, las cuales tienden a tener un mayor grado de malignidad de la masa, lo cual es un buen paso antes de la clasificación del cáncer. Y que, de igual manera, no dependen de la anotación de experto, sino preferiblemente de un modelo de detección, el cual se pensó como paso previo y como una posible integración en un futuro al modelo de segmentación seleccionado para este proyecto. Además, los modelos de contorno activos no requieren de un hardware complejo o costoso, ya que no son tan demandantes computacionalmente, siendo mucho más prácticos y adaptables en términos de implementación.

4. CONCLUSIONES Y TRABAJO FUTURO

En este proyecto se ha hecho una exploración y comparación de dos modelos de contornos activos, Geodésico y Chan-Vese, para mejorar la segmentación de masas en imágenes de mamografía. En conclusión, el mejor modelo seleccionado fue el Chan-Vese con un radio de 50 píxeles y 436 iteraciones, el cual superó al mejor modelo Geodésico, logrando un puntaje promedio de Dice de 0.812 frente a 0.558. Este resultado resaltó el rendimiento exitoso del modelo Chan-Vese en la segmentación de lesiones de masas en diferentes imágenes. Los filtros de Mediana y CLAHE fueron cruciales para mejorar la calidad del borde de la lesión antes del paso de segmentación. Además, el método propuesto logró segmentar con éxito masas con formas complejas e irregulares, lo cual se considera un resultado esencial para la clasificación del cáncer en términos del grado de malignidad.

Además, al comparar los enfoques del modelo Geodésico y el modelo Chan-Vese, se evidenciaron diferencias sustanciales, destacando la tendencia del modelo Geodésico a quedar atrapado en mínimos locales, no llegando así a segmentar correctamente las masas. Por el contrario, el modelo Chan-Vese demostró una mayor capacidad para superar estos desafíos. Este análisis proporciona una comprensión crítica de los puntos fuertes y limitaciones de cada modelo, información valiosa para la selección adecuada de técnicas en entornos clínicos específicos.

En el contexto nacional e internacional, la relevancia de esta investigación radica en su contribución directa a la mejora de las herramientas de diagnóstico temprano del cáncer de mama, una enfermedad de alta incidencia y significativa importancia en la salud pública a nivel mundial. La precisión alcanzada por el modelo Chan-Vese en la segmentación de lesiones de masas, resalta su potencial para mejorar la eficiencia diagnóstica y, en última instancia, la atención al paciente.

En la continuación de esta investigación, se abren diversas oportunidades para ampliar y perfeccionar el enfoque propuesto. En primer lugar, se sugiere aumentar la cantidad de datos en el conjunto de datos experimental, para así obtener una validación más robusta y exhaustiva del método. Esto permitirá evaluar la generalización del modelo en una gama más amplia de casos clínicos y fortalecer la confianza en su aplicabilidad práctica. Además, se propone explorar más técnicas de preprocesamiento de imágenes con el objetivo de mejorar la visibilidad de las masas en las mamografías y así sean más fácilmente distinguidas del tejido circundante. Esto podría incrementar considerablemente los resultados de las métricas obtenidas por los modelos de contornos activos propuestos. Finalmente, se propone integrar el método desarrollado con un modelo de detección de masas, proporcionando a los profesionales una segunda opinión en el diagnóstico de cáncer de mama. Este trabajo sienta las bases para futuras investigaciones que podrían profundizar en la interpretación de resultados, considerar otros modelos de segmentación, y abordar desafíos específicos asociados con la variabilidad en las imágenes de mamografías.

En términos de aprendizaje, en este proyecto se han explorado los parámetros que se pueden optimizar dentro de los modelos de contornos activos Geodésico y Chan-Vase, tales como el radio de inicialización y el número de iteraciones. Parámetros que se usan para maximizar el rendimiento de los modelos de contornos activos. Por ende, se puede decir que se ha obtenido un conocimiento valioso y profundo en el funcionamiento, implementación y aplicabilidad de estos modelos. Además, la aplicación de filtros de Mediana y CLAHE fueron una estrategia esencial para mejorar la calidad del borde de las lesiones, contribuyendo así a una segmentación más precisa. La aplicación de este preprocesamiento de imágenes contribuyó a construir una base, en términos de conocimiento, de algunas técnicas comunes que existen para destacar un área de interés dentro de una imagen. Un área de conocimiento que sería interesante seguir explorando e investigando. Finalmente, el enfocarse en un problema

relacionado con el cáncer de mamá se profundizo el conocimiento y la concientización sobre la importancia y el impacto que este tipo de cáncer tiene en el mundo, pero especialmente en Ecuador. De igual manera, la importancia de la detección temprana del cáncer y los chequeos contantes. Factores que generaron una nueva perspectiva sobre como poder seguir contribuyendo a una posible solución a este problema a traves de la carrera de ingeniería en ciencias de la computación.

REFERENCIAS BIBLIOGRÁFICAS

- Alhadidi, B., Zu`bi, H. M., & Suleiman, N. (2007). *Mammogram breast cancer image detection using image processing functions*. *Information Technology Journal*, 6(2), 217-221. <https://doi.org/10.3923/itj.2007.217.221>
- Álvarez, L., Baumela, L., Henríquez, P., & Márquez-Neila, P. (2010). *Morphological snakes*. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2197-2202. <https://doi.org/10.1109/cvpr.2010.5539900>
- Baccouche, A., García-Zapirain, B., Cristian, C., & Elmaghraby, A. (2021). *Connected-UNETs: a deep learning architecture for breast mass segmentation*. *NPJ Breast Cancer*, 7(1). <https://doi.org/10.1038/s41523-021-00358-x>
- Caselles, V., Kimmel, R., & Sapiro, G. (1997). *Geodesic active contours*. *International Journal of Computer Vision*, 1(22), 61-79. <https://doi.org/10.1109/iccv.1995.466871>
- Chan, T. F., & Vese, L. A. (2001). *Active contours without edges*. *IEEE transactions on image processing*, 10(2), 266-277. <https://doi.org/10.1109/83.902291>
- Ekici, S., & Jawzal, H. (2020). *Breast cancer diagnosis using thermography and convolutional neural networks*. *Medical Hypotheses*, 137, 109542. <https://doi.org/10.1016/j.mehy.2019.109542>
- Huang, N., Chen, L., He, J., & Nguyen, Q. D. (2022). *The efficacy of clinical breast exams and Breast Self-Exams in detecting malignancy or positive ultrasound findings*. *Cureus*. <https://doi.org/10.7759/cureus.22464>
- Kass, M., Witkin, A., & Terzopoulos, D. (1988). *Snakes: Active contour models*. *International Journal of Computer Vision*, 1(4), 321-331. <https://doi.org/10.1007/bf00133570>
- Liu, F., Zhang, G., Chen, Y., & Gu, Y. (2015). *Segmentation of mass in mammograms by a novel integrated active contour method*. *International Journal of Computational Science and Engineering*, 11(2), 207. <https://doi.org/10.1504/ijcse.2015.071883>
- Márquez-Neila, P., Baumela, L., & Álvarez, L. (2014). *A morphological approach to Curvature-Based evolution of curves and surfaces*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1), 2-17. <https://doi.org/10.1109/tpami.2013.106>
- Menet, S., Saint-Marc, P., & Medioni, G. (2002). *Active contour models: overview, implementation and applications*. *IEEE International Conference on Systems, Man,*

- and Cybernetics Conference Proceedings, 194-199.
<https://doi.org/10.1109/icsmc.1990.142091>
- Ministerio de la Salud Pública (MSP). (2022). *Cáncer de mama: riesgos, síntomas y tratamiento oportuno*. Ministerio de la Salud Pública.
<https://www.salud.gob.ec/cancer-de-mama-riesgos-sintomas-y-tratamiento-oportuno/#:~:text=Seg%C3%BAn%20datos%20de%20la%20Organizaci%C3%B3n,muerte%20en%20mujeres%20de%20Ecuador.>
- Moreira, I., Amaral, I. B. C., Domingues, I., Cardoso, A. N. L., Cardoso, M. J., & Cardoso, J. S. (2012). *INbreast*. *Academic Radiology*, 19(2), 236-248.
<https://doi.org/10.1016/j.acra.2011.09.014>
- National Cancer Institute. (2021). *What is cancer?*. National Cancer Institute.
<https://www.cancer.gov/about-cancer/understanding/what-is-cancer>
- National Institutes of Health (US). (2007). *Understanding cancer*. NIH Curriculum Supplement Series - NCBI Bookshelf.
<https://www.ncbi.nlm.nih.gov/books/NBK20362/>
- Niaz, A., Memon, A. A., Rana, K., Joshi, A., Soomro, S., Kang, J. S., & Choi, K. N. (2020). *Inhomogeneous image segmentation using hybrid active contours model with application to breast tumor detection*. *IEEE Access*, 8, 186851-186861.
<https://doi.org/10.1109/access.2020.3029333>
- Nounou, M. I., ElAmrawy, F., Ahmed, N. A., Abdelraouf, K., Goda, S., & Syed-Sha-Qhattal, H. (2015). *Breast cancer: conventional diagnosis and treatment modalities and recent patents and technologies*. *Breast Cancer: Basic and Clinical Research*, 9s2, BCBCR.S29420. <https://doi.org/10.4137/bcbr.s29420>
- Nyström, L., Wall, S., Rutqvist, L. E., Lindgren, A., Lindqvist, M., Rydén, S., Andersson, J., Bjurstam, N., Fagerberg, G., Frisell, J., Tabár, L., & Larsson, L. (1993). *Breast cancer screening with mammography: Overview of Swedish randomised trials*. *The Lancet*, 341(8851), 973-978. [https://doi.org/10.1016/0140-6736\(93\)91067-v](https://doi.org/10.1016/0140-6736(93)91067-v)
- OECD.AI. (s. f.). Dice score. OECD.AI. <https://oecd.ai/en/catalogue/metrics/dice-score>
- Ozmen, N., Dapp, R., Zapf, M., Gemmeke, H., Ruitter, N. V., & Van Dongen, K. W. A. (2015). *Comparing different ultrasound imaging methods for breast cancer detection*. *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, 62(4), 637-646. <https://doi.org/10.1109/tuffc.2014.006707>

- Perez, N. P. (2015). *Improving variable selection and mammography-based machine learning classifiers for breast Cancer CADX*. Ph.D. dissertation, Universidade do Porto (Portugal). <https://hdl.handle.net/10216/129876>
- Phillips, D. (s/f). *Image Processing, Part 9: Histogram-Based Image Segmentation*. Pennelynn. <http://www.pennelynn.com/Documents/CUJ/HTML/93HTML/19930064.HTM>
- Radhi, E. A., & Kamil, M. Y. (2021). *Breast tumor detection via active contour technique*. *International Journal of Intelligent Engineering and Systems*, 14(4), 561-570. <https://doi.org/10.22266/ijies2021.0831.49>
- Rosebrock, A. (2023). *Intersection over Union (IOU) for object detection - PyImageSearch*. PyImageSearch. <https://pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>
- Scikit-Image. (s/f). *Watershed Segmentation*. Scikit-Image. https://scikit-image.org/docs/stable/auto_examples/segmentation/plot_watershed.html
- Su, Y., Liu, Q., Xie, W., & Hu, P. (2022). *YOLO-LOGO: a transformer-based YOLO segmentation model for breast mass detection and segmentation in digital mammograms*. *Computer Methods and Programs in Biomedicine*, 221, 106903. <https://doi.org/10.1016/j.cmpb.2022.106903>
- Taylor-Phillips, S., & Stinton, C. (2020). *Double reading in breast cancer screening: Considerations for policy-making*. *British Journal of Radiology*, 93(1106), 20190610. <https://doi.org/10.1259/bjr.20190610>
- Wang, Z., Wang, E., & Zhu, Y. (2020). *Image Segmentation Evaluation: A Survey of methods*. *Artificial Intelligence Review*, 53(8), 5637-5674. <https://doi.org/10.1007/s10462-020-09830-9>
- World Health Organization (WHO). (2023). *Breast cancer*. <https://www.who.int/news-room/fact-sheets/detail/Breast-Cancer>
- Xu, C Xu & Prince, J. L. (1997). *Gradient vector flow: a new external force for snakes*. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 66-71, doi: 10.1109/CVPR.1997.609299.