



# SEGMENTACIÓN Y CARACTERIZACIÓN DE MASAS EN MAMOGRAFÍAS DIGITALES DE CAMPO COMPLETO

H. Peregrina-Barreto<sup>a</sup>, J.A. Sánchez Tiro<sup>a</sup>, G. López Armas<sup>b</sup>,  
<sup>a</sup>Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla  
<sup>b</sup>Centro de Enseñanza Técnica Industrial  
[hperegrina@ieee.org](mailto:hperegrina@ieee.org), [jsantonio@inaoep.mx](mailto:jsantonio@inaoep.mx), [glopez@ceti.mx](mailto:glopez@ceti.mx)



XI Congreso Nacional de Tecnología Aplicada a Ciencias de la Salud  
I Congreso Internacional de Tecnología Aplicada a Ciencias de la Salud  
"Generación de nuevas técnicas de diagnóstico y tratamiento"  
10, 11 y 12 de Junio de 2021  
Sede en línea.  
Facultad de Medicina, UNAM Facultad de Medicina, UNAM

**Resumen**— El análisis de mamografías es el instrumento que permite al especialista médico tomar decisiones de tratamiento oportunas evitando que el cáncer evolucione a etapas avanzadas con consecuencias fatales. En particular, la identificación de masas puede ser una tarea desafiante ya que depende de atributos que se ven afectados por la alta densidad mamaria. Sin embargo, la interpretación de mamografías es un trabajo demandante para el radiólogo e impone una gran carga de trabajo debido a la saturación de los servicios de salud. Por ello, se busca desarrollar herramientas tecnológicas que sirvan como apoyo en el análisis de mamografías, proveyendo información útil al radiólogo. Para la caracterización de masas, el estándar BIRADS usa tres descriptores: forma, margen y densidad. Este trabajo propone una metodología para detección y caracterización de masas en mamografías FFDM (Full-Field Digital Mammography). Las imágenes son procesadas automáticamente en varias etapas que incluyen extracción de la región mamaria y eliminación de artefactos, identificación de la zona que contiene a la masa y su segmentación final. Diversas características fueron extraídas a la masa segmentada para poder asociar un valor cuantitativo a cada descriptor BI-RADS. En el desarrollo de los experimentos, la caracterización representó un reto particular ya que, i) si bien las masas son descritas de manera general como formas redondas, frecuentemente también presentan espículas afectando significativamente los valores de las medidas estándar de redondez; ii) la alta densidad mamaria afecta significativamente a la segmentación de espículas. Además, iii) la densidad de la masa puede ser diferente a la densidad del tejido que la rodea, y fue necesario encontrar la forma de analizar esta relación. La solución desarrollada se basó en técnicas de procesamiento de imágenes digitales y los resultados mostraron que es posible alcanzar una coincidencia significativa con respecto al radiólogo en el 80% de los casos.

**Palabras clave**—masas, BI-RADS, segmentación, extracción de características

## I. INTRODUCCIÓN

Una detección temprana del cáncer de mama a través de las mamografías digitales puede garantizar un tratamiento eficaz para el paciente, previniendo el crecimiento y la propagación del cáncer, lo cual puede salvarle la vida. La incidencia del cáncer de mama se ha ido acrecentando, teniendo 2.1 millones casos nuevos en 2018 mientras que en 2020 hubo 2.3 millones [1]. La interpretación mamográfica se basa en los hallazgos observados y sus características, lo cual está establecido por el *Breast Imaging Report and Database System* (BI-RADS) [2]. También se establece el nivel de densidad que tiene un seno y esto a través del American College of Radiology (ACR) [3], en el que se clasifica de la A, B, C y D el nivel de tejido fibroglandular siendo la A (10%) bajamente denso y D (>90%) altamente denso. En el análisis de las mamografías es frecuente observar lesiones sospechosas de malignidad. Una de ellas son las masas, que se refieren a una lesión tridimensional circunscrita en una zona del tejido mamario cuyas características morfológicas que deben ser observadas por el radiólogo para determinar su benignidad o malignidad.

## II. OBJETIVO

Desarrollar una metodología automática para procesar y analizar imágenes digitales de mastografías FFDM a fin de localizar masas presentes en la región de la mama.

## III. METODOLOGÍA

### a. Pre-procesamiento y Detección

La base de datos de mamografías FFDM Inbreast [4] fue utilizada en el desarrollo de este trabajo; las imágenes tienen asociado el marcador generado por el experto radiólogo indicando la región de la masa (groundtruth). Se realizó la eliminación de información no relevante (etiquetas y el músculo pectoral) dejando solo el área de interés (mama). Ya que el interior de la mama está compuesto por tejido graso (oscuro) y fibroglandular (brillante), siendo este último el tipo de tejido donde aparece la gran mayoría de las masas, se eliminaron las regiones asociadas a tejido graso. Esto se realizó mediante un análisis dinámico de la distribución de niveles de gris, lo cual permite que el proceso se adapte a las condiciones individuales de cada imagen mejorando la extracción de las posibles regiones de interés.

b. Adicionalmente, el tejido fibroglandular puede encontrarse compactado en regiones de mayor área o disperso en pequeños elementos donde es menos probable que se encuentre la masa. Para eliminar las regiones dispersas se utilizó una apertura morfológica, que además tiene el efecto de generar regiones de menor valor respecto a la original. Sin embargo, este efecto es

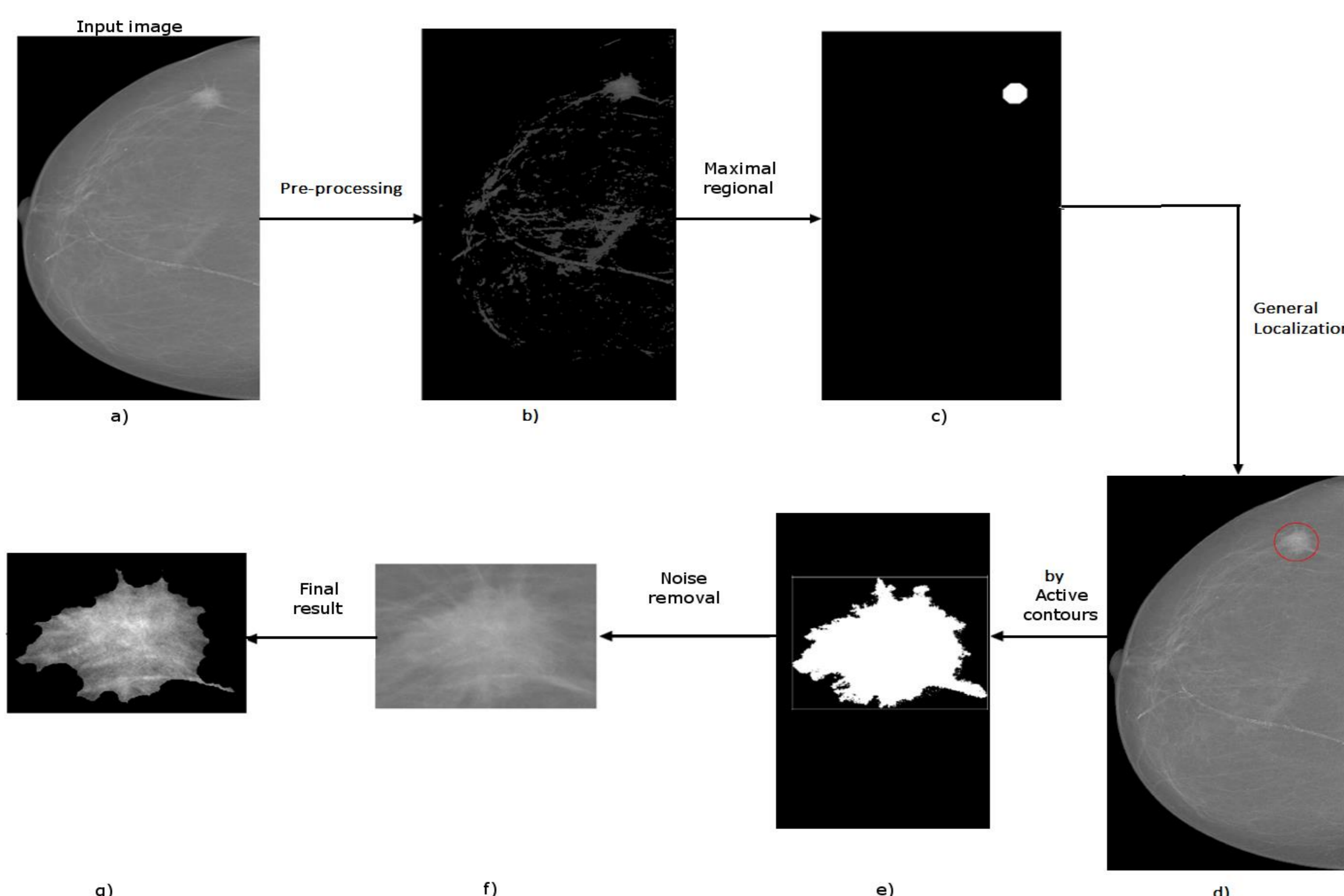


Fig. 1: Framework de la metodología propuesta.

menor en regiones brillantes como las masas, por lo cual es útil también para localizarlas. Mediante la localización del máximo regional que abarca mayor área se logró localizar la región general de la masa.

### c. Extracción de la masa

La región general de la masa pasa a ser la semilla para una segmentación mediante contornos activos, el cual se encarga de encontrar los bordes que mejor delimitan la región de la masa. De esta forma se puede obtener un mayor detalle de la forma que tiene la masa, lo cual es relevante en la características que establece BI-RADS para su clasificación.

## III. RESULTADOS

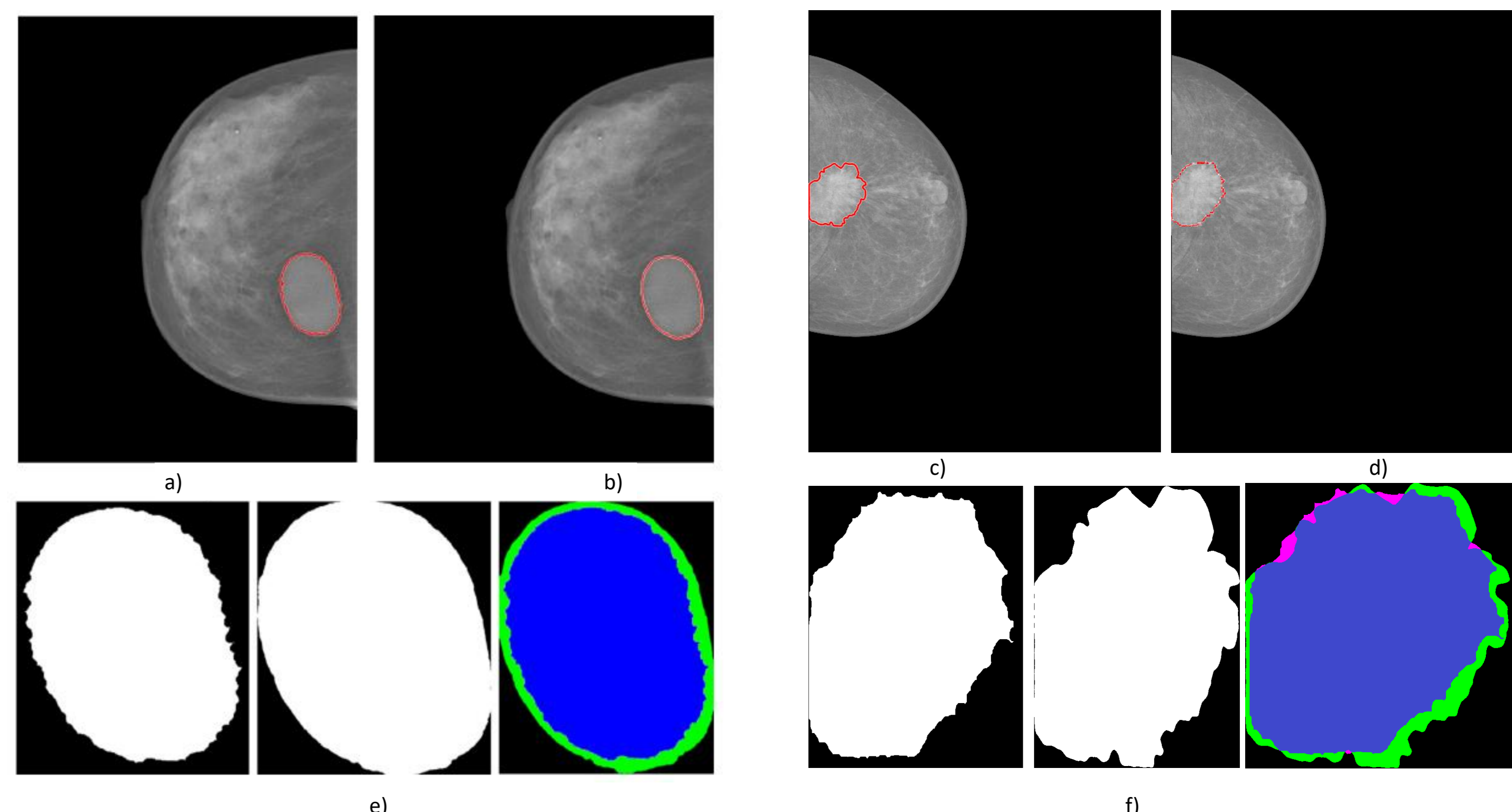


Fig. 2: Resultado de la localización y segmentación: a, c) mamografías tipo C y B indicando la localización de la masa (groundtruth), b,d) localización obtenida por la metodología propuesta; e,f) comparación de la segmentación obtenida (izq.) y el groundtruth (centro) mediante el coeficiente de Dice (der.)

Respecto a la localización de la masa dentro de la mama, se logró localizar correctamente el 75% de los casos tomando como referencia el groundtruth de la base de datos (Fig. 2a-b y Fig. 3a-b). Para evaluar los resultados de la segmentación, se utilizó el coeficiente Dice, que toma en cuenta el traslape entre dos conjuntos (segmentación obtenida y groundtruth) [5]. En la Fig. 2e, correspondiente a una mamografía tipo C, la masa segmentada obtenida por los contornos activos (marcador rosa) se comparó con la referencia (marcador verde). La superposición entre ambos marcadores (azul) ocurre en la mayor parte del área de masa mientras que su diferencia se encuentra principalmente en el borde superior. La medida de Dice obtuvo un valor de 0.87, lo que indica una alta similitud de los resultados. Otro resultado se muestra en la Fig. 2f donde la masa fue segmentada en una mamografía tipo B. Como se observa, existe una mayor similitud entre la segmentación obtenida y la segmentación de referencia, obteniendo una medida de Dice de 0.92. Se encontró una mayor dificultad en la localización de masas dentro de mamas con alta densidad, ya que existe una alta similitud en niveles de gris y algunas partes de la masa no logran ser delimitados de manera más precisa. La medida de Dice tuvo un valor >0.6 en el 80% de las mamografías tipo A, B y C, y 70% para mamografías tipo D. Para la clasificación se evaluaron características de margen y densidad de 81 imágenes segmentadas con la etiqueta asignada por BI-RADS. Utilizando el clasificador SVM se obtuvo una precisión del 72%, lo cual es un resultado interesante tomando en cuenta que se utilizaron sólo 2 características BIRADS que describen la masa. Se espera que al incluir la tercera característica (forma) el porcentaje de clasificación se incremente.

## IV. CONCLUSIONES

La identificación y la segmentación de masas en mastografías es vital para poder realizar su análisis en un sistema de apoyo al diagnóstico. Los resultados obtenidos en este trabajo muestran que es posible alcanzar una coincidencia significativa con el marcador del médico en el 80% de los casos. Sin embargo, aunque fue posible localizar la masa en las mamografías tipo D, el alto nivel de densidad en la mama afecta los resultados de segmentación alcanzando una medida de Dice menor respecto a los otros tipos de densidad. Se pudo observar además que la medida de Dice es sensible a pequeñas diferencias en los límites de la masa, principalmente en masas espiculadas, lo cual genera que no se obtengan valores más altos aún cuando la segmentación alcanzada es altamente similar a la referencia del experto. Como trabajo futuro se propone mejorar la localización y segmentación de masas en mamografías de alta densidad así como utilizar otras medidas de similitud que reflejen una comparación más adecuada de los resultados. Además proponer características que reflejen la forma, el margen y la densidad de manera que discriminen mejor las masas.

## V. REFERENCIAS

- [1] WHO. Breast cancer: prevention and control. [Online]. Available: <https://www.who.int/topics/cancer/breastcancer/es/index1.html>
- [2] A. C. of Radiology, C. J. D'Orsiet et al., ACR BI-RADS Atlas: Breast Imaging Reporting and Data System; Mammography, Ultrasound, Magnetic Resonance Imaging, Follow-up and Outcome Monitoring, Data Dictionary. ACR, American College of Radiology, 2013.
- [3] Balleyguier, C., Ayadi, S., Van Nguyen, K., Vanel, D., Dromain, C., & Sigal, R. (2007). BIRADS™ classification in mammography. *European journal of radiology*, 61(2), 192-194.
- [4] Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., & Cardoso, J. S. (2012). Inbreast: toward a full-field digital mammographic database. *Academic radiology*, 19(2), 236-248.
- [5] Thada, V., & Jaglan, V. (2013). Comparison of jaccard, dice, cosine similarity coefficient to find best fitness value for web retrieved documents using genetic algorithm. *International Journal of Innovations in Engineering and Technology*, 2(4), 202-205.