

## **Determinación de Tejido Glandular en Imágenes de Mastografía Mediante Métricas de Textura**

M. A. Vázquez<sup>1\*</sup>, A. R. Mejía Rodríguez<sup>1</sup>, J. P. Castillo López<sup>2</sup>, H. A. Galván Espinoza<sup>2</sup>

<sup>1</sup> Facultad de Ciencias, Universidad Autónoma de San Luis Potosí, S.L.P., México

<sup>2</sup> Departamento de Radiodiagnóstico, Instituto Nacional de Cancerología, CDMX, México

\*malejandro.vazmtz@mail.com

### **Resumen**

En el presente trabajo muestra el desarrollo de una estrategia para la determinación automática de tejido glandular en imágenes de mastografía a partir del uso de filtros de textura sobre las imágenes. Las métricas de textura permiten extraer información sobre la distribución de intensidades en una imagen en escala de grises, lo que minimiza la variación de intensidades inherentes al proceso de adquisición de una imagen de mastografía, permitiendo resaltar patrones específicos del tejido fibroglandular, los cuales regularmente son difíciles de ver en la imagen original. Se estudiaron particularmente las métricas de textura de contraste, correlación, energía y homogeneidad en un conjunto de datos de 40 mastografías (población de 20 pacientes). Los resultados encontrados sugieren que las métricas de textura pueden ser de utilidad para determinar el porcentaje de glandularidad de forma automática, pero es necesario evaluar una base de datos más extensa para poder determinar los umbrales de decisión para la asignación de cada pixel de la imagen de mastografía a las clases de tejido adiposo o tejido glandular.

**Palabras clave:** Cáncer de mama, Etiquetado automático, Métricas de textura, Tejido glandular.

## **1. Introducción**

El cáncer de mama es un problema de salud pública prioritario por su magnitud y trascendencia. En México a partir del año 2006, el carcinoma mamario se convirtió en la primera causa de muerte por cáncer en mujeres [1].

La mama comprende una mezcla de tejidos epiteliales y de tejido conectivo en conjunto con el tejido graso. La distribución de estos tejidos en los estudios de mamografía se conoce como densidad mamográfica. Radiográficamente la grasa puede observarse radiolúcida (gris oscuro) mientras que los tejidos epiteliales y conectivos (tejido fibroglandular) se observan radiopacos (blancos). La densidad mamaria puede describirse como un porcentaje, el cual consiste en el área de tejido fibroglandular visible en una mamografía en relación con el área total del seno que incluye al tejido adiposo. Esta área se define como porcentaje de área densa [2].

Los componentes epiteliales y estromales de la glándula mamaria son variables entre cada mujer, si existe mayor densidad, existe mayor componente epitelial. La mayoría de los cánceres de mama surgen del componente epitelial, por lo cual una mujer con mayor densidad puede tener mayor probabilidad de desarrollar cáncer de mama comparada con otra de menor densidad. Además, una mayor cantidad de tejido fibroglandular dificulta la identificación oportuna de lesiones. Por esta situación, el porcentaje de tejido fibroglandular se ha asociado con un factor de riesgo para desarrollar cáncer de mama [3].

La determinación de la densidad mamaria por parte del médico radiólogo se realiza utilizando las categorías descritas en el Breast Imaging Reporting And Data System (BI-RADS), las cuales se

encuentran divididas en 4 clases, utilizada de forma cualitativa a percepción del médico radiólogo. A partir del 2013, para mejorar la determinación el porcentaje de densidad utiliza escalas visuales para proporcionar una medida continua. Los programas de cómputo basados en segmentación de estructuras por área (cantidad de blancos en la imagen), se desarrollaron para mejorar la reproducibilidad en la determinación del porcentaje de tejido mamario. Sin embargo, estos programas requieren de personas altamente calificadas en la lectura de estudios de mastografía [4].

Existen varios métodos para medir la densidad mamaria. Muchas de estas medidas se basan en identificar tejido radiopaco (blanco) en una mamografía. Los programas semiautomáticos para calcular la densidad mamaria han mostrado una fuerte correlación con el riesgo de cáncer de mama. Varias medidas automáticas de textura también han demostrado estar relacionadas con el cáncer de mama; ejemplos de estas métricas incluyen filtrado de dominio de imagen, medidas de co-ocurrencia y momentos de orden superior (por encima de uno) derivados de la distribución de píxeles [5].

Considerando lo anterior, la finalidad de este trabajo es establecer una estrategia para determinar de forma automática el porcentaje de tejido glandular (área densa) presente en imágenes de mastografía a través del uso de métricas de textura.

## **2. Metodología**

### **2.1 Imágenes de Mastografía**

Para este estudio se utilizaron datos de una población de 20 individuos (40 mastografías considerando mama derecha e izquierda de cada paciente) donde el 50% de la población de estudio tuvo un diagnóstico negativo para cáncer de mama y el otro 50% habían recibido un diagnóstico confirmado de cáncer. Todas las imágenes fueron proporcionadas por el Instituto Nacional de Cancerología (INCan) de la Ciudad de México, en formato DICOM, con dimensiones de la imagen en el rango de 2864x1262 hasta 5257x3052 píxeles con tamaño de pixel de 50x50  $\mu\text{m}$ . Adicionalmente, se contó con una imagen de cada mastografía donde un especialista clínico había etiquetado los tejidos adiposo y glandular usados como regiones de referencia, para evaluar el desempeño de la estrategia propuesta; estas regiones de referencia fueron obtenidas mediante el uso de un software en donde el especialista clínico manualmente modifica el histograma de la imagen para resaltar el contraste entre tejido fibroglandular (glandular) y adiposo.

### **2.2 Métricas de Textura**

Para homogeneizar la detección de tejido glandular presente en las mastografías se propone el uso de filtros de textura, que al no depender directamente de la intensidad de cada pixel brindan información confiable y permiten generalizar el método de detección automática a un mayor número de casos.

Tradicionalmente la segmentación de imágenes se realiza directamente sobre los niveles de intensidad de los píxeles; sin embargo, existen diferentes operadores que permiten obtener información diferente para evidenciar las diferencias presentes en diversas regiones de interés de la imagen. Particularmente, las métricas de textura en una imagen dan información de la distribución espacial de las intensidades de grises, lo que permite diferenciar entre regiones con una disposición o patrón diferente sin importar el nivel de intensidad de los píxeles. Es posible realizar un análisis de textura local por vecindarios dentro de la imagen, lo que permite generar una nueva imagen transformada con diferente información a la imagen original.

Las métricas de textura utilizadas en este trabajo son calculadas a partir de la estimación de una matriz de co-ocurrencia (MCO), la cual se basa en distribuciones de probabilidad conjunta de pares de píxeles. Una MCO es una matriz donde el número de renglones y columnas es igual al número de niveles de intensidad (tonalidades de gris) presentes en la imagen. Cada elemento de la matriz está

definido por  $p(i, j|d, \theta)$  que contiene valores de probabilidad para los cambios entre los niveles de gris  $i$  y  $j$  a una distancia de desplazamiento particular  $d$  a un cierto ángulo  $\theta$  (que puede ser a  $0^\circ, 45^\circ, 90^\circ$  o  $135^\circ$ ) [6]. El valor de cada pixel  $MCO_d(i, j)$  es calculado a partir de:

$$MCO_d(i, j) = \text{card}\{(r, c) | I(r, c) = i, I(r + dr, c + dc) = j\} \tag{1}$$

Un ejemplo de cómo se estima una MCO se puede observar en la Fig. 1 donde el valor de  $C_{0,1}(1,2)$  corresponde al número de veces que se puede encontrar los valores consecutivos 1,2 en el vector de desplazamiento  $[0,1]$ , correspondiente a  $0^\circ$ , que en este caso corresponde con un valor de 2.

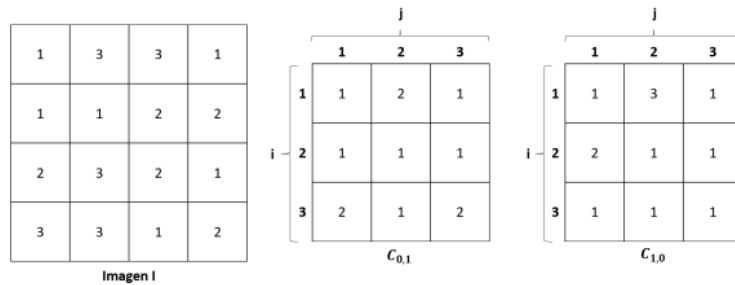


Fig 1. Ejemplo de dos Matrices de Co-ocurrencia para vectores de desplazamiento a  $0^\circ$  y  $90^\circ$ , de una imagen I [6].

Al definir la MCO normalizada como  $p(i, j) = C_d(i, j) / \sum_{i,j} C_d(i, j)$  se pueden calcular diferentes medidas estadísticas que darán la información sobre la textura de la imagen. Algunas de las métricas más utilizadas para la clasificación de imágenes médicas son [6]: Contraste (2), Correlación (3), Energía (4), Homogeneidad (5), las cuales se definen a continuación:

$$CONTRASTE = \sum_{i,j} |i - j|^2 p(i, j) \tag{2}$$

$$CORRELACIÓN = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j) p(i, j)}{\sigma_i \sigma_j} \tag{3}$$

$$ENERGÍA = \sum_{i,j} p(i, j)^2 \tag{4}$$

$$HOMOGENEIDAD = \sum_{i,j} \frac{p(i, j)}{1 + |i - j|} \tag{5}$$

En este trabajo, cada filtro fue aplicado a matrices calculadas en 3 direcciones (horizontal –  $0^\circ$ , vertical –  $90^\circ$  y diagonal –  $135^\circ$ ) con ventanas de  $3 \times 3, 5 \times 5$  y  $7 \times 7$  pixeles, y 8, 16, 32, 64 y 128 niveles de gris, por medio de un algoritmo desarrollado en Matlab el cual aplicaba las operaciones aritméticas correspondientes a cada métrica, en el momento que se calculaba la MCO correspondiente a cada ventana, asignando de esta manera un valor de textura al pixel central y devolviendo la imagen filtrada, obteniendo como resultado 2400 imágenes filtradas con las que se definieron métodos algebraicos para la clasificación de tejido glandular.

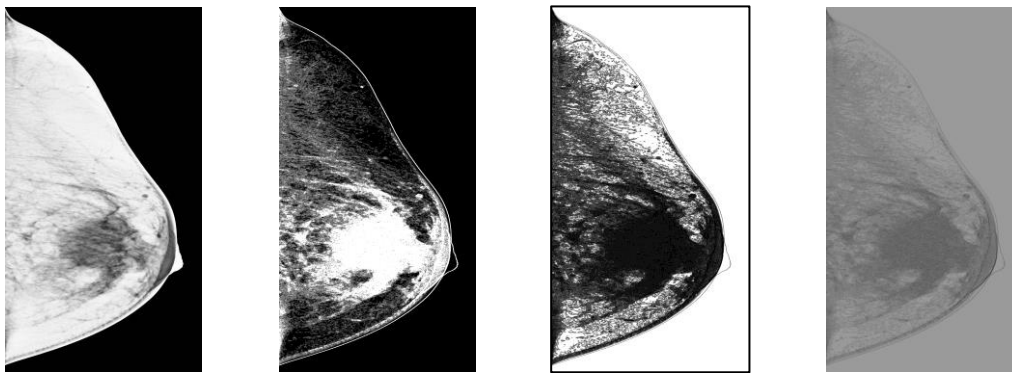
Para la clasificación de tejidos se propuso analizar la derivada del histograma de la imagen de textura, considerando que la mayoría de los histogramas presentan 2 picos pronunciados que corresponden al umbral para diferenciar tejido glandular y adiposo, tomando en cuenta que el valor de fondo permanece constante en todas las imágenes.

Por último, se presentaron los resultados a especialistas, y se estimó el porcentaje de píxeles clasificados como tejido glandular y aquellos como tejido adiposo para la estimación del porcentaje de área densa.

### 3. Resultados y Discusión

A continuación, se presentan los resultados obtenidos de aplicar los diferentes filtros de textura a las imágenes de mastografía disponibles, y se muestran particularmente ejemplos de 2 imágenes correspondientes a una paciente sana y otra con diagnóstico positivo de cáncer de mama para poder realizar una comparación visual. Es importante mencionar que no se muestran resultados del filtro de correlación debido a que no se observaron detalles relevantes ya que para todas las mastografías se obtuvieron imágenes completamente blancas. Adicionalmente, por inspección visual avalada por especialistas clínicos, se acotaron los parámetros estudiados de las métricas a textura a aquellos con mejor desempeño, los cuales corresponden escalamiento de las imágenes a 32 niveles de intensidad de gris, y una ventana de  $7 \times 7$  píxeles para la estimación de la MCO. De esta forma se tiene una buena relación en cuanto a calidad de la imagen, calidad del filtrado de textura y tiempo de procesamiento, siendo este último de aproximadamente 20 segundos para una imagen de aproximadamente  $4000 \times 1500$  píxeles.

La Fig. 2 muestra los resultados en una imagen de la mama izquierda de una paciente diagnosticada con cáncer, en donde se puede observar en la imagen original de intensidades en escalas de grises (Fig. 2.a) que el tejido glandular corresponde a los píxeles en tonos de gris oscuros (radiolúcidos) y el tejido adiposo son aquellos píxeles en tonalidades cercanas al blanco (radiopacos). En esta imagen es posible apreciar que determinar la proporción de tejido glandular es particularmente difícil debido a que hay una gran variación de tonalidades de gris correspondientes al tejido glandular. Por otro lado, las Fig. 2.b, 2.c y 2.d corresponden a las imágenes obtenidas al aplicar los filtros de contraste, energía y homogeneidad respectivamente, en las que es posible apreciar que es más evidente la separación de tejidos glandular y adiposo respecto de la imagen original. Sin embargo, se decidió trabajar con el resultado de la métrica de homogeneidad debido a que visualmente permite una separación casi binaria de los tejidos de interés, es decir, que tiene una forma más evidente de realizar la diferenciación de tejidos mamarios respecto de las otras dos métricas de textura y la imagen de intensidades.



a)

b)

c)

d)

Fig 2. a) Imagen original (intensidades) de la mama izquierda de una persona con diagnóstico positivo de cáncer de mama. b) Imagen resultante de contraste. c) Imagen resultante de energía. d) Imagen resultante de homogeneidad.

En la Fig. 3.a podemos observar el resultado proporcionado por los profesionales de la salud utilizando un ajuste manual de los valores del histograma para evidenciar la presencia de tejido glandular al oscurecer el tejido adiposo. La Fig. 3.b muestra el resultado de la segmentación de las tres clases (tejido glandular, tejido adiposo y fondo) obtenido por el método derivativo aplicado sobre el histograma de la imagen con la métrica de homogeneidad. En esta imagen se puede observar que cualitativamente el método propuesto es comparable al ajuste manual del histograma de la imagen original realizado por el especialista clínico, y que, aunque las etiquetas pueden variar en algunos píxeles, las regiones conservan su morfología y tamaño de manera general. Es importante mencionar que, para todas las mastografías de pacientes con diagnóstico positivo de cáncer de mama, el especialista clínico corroboró el buen desempeño cualitativo de la propuesta.

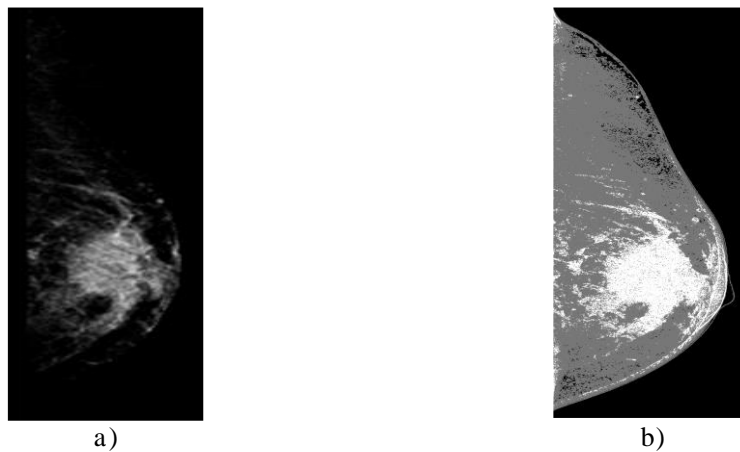


Fig 3 a) Imagen con ajuste manual del histograma de mastografía con diagnóstico positivo de cáncer de mama. b) Correspondiente resultado de segmentación (etiquetado) por el método derivativo del tejido glandular (color blanco), tejido adiposo (color gris) y fondo (color negro).

De forma similar, las imágenes de la Fig. 4 presentan los resultados obtenidos a partir de una mastografía de mama izquierda de una persona con diagnóstico negativo de cáncer, en donde se puede observar en la imagen original (Fig. 4.a) una menor presencia de tejido glandular, que igualmente corresponde con los píxeles en tono de gris más oscuro. Sin embargo, también se aprecia en este caso que parte del tejido adiposo se encuentra en intensidades de gris similares al tejido glandular por lo que la correcta identificación de cada tipo de tejido se vuelve una tarea complicada. Por otra parte, las Fig. 4.b, 4.c y 4.d corresponden a las imágenes obtenidas al aplicar los filtros de contraste, energía y homogeneidad respectivamente, y de nuevo es posible evidenciar las regiones correspondientes a ambos tipos de tejidos (glandular y adiposo) de mejor manera que con la imagen de intensidad; no obstante, bajo los mismos argumentos expuestos en la Fig. 2, se decidió trabajar con el resultado de la métrica de homogeneidad.

La Fig. 5 muestra la comparación de los resultados proporcionados por los profesionales de la salud, utilizando un ajuste manual de los valores del histograma para evidenciar la presencia de tejido glandular al superponer las regiones de interés con color rojo (Fig. 5.a), y la Fig. 5.b muestra el resultado de la segmentación de las tres clases (tejido glandular, tejido adiposo y fondo) por el método propuesto. Al igual que en el caso anterior se puede observar que el resultado es cualitativamente comparable a la estimación con el ajuste manual sobre el histograma de la imagen, por lo que de nuevo se corroboró el buen desempeño cualitativo de la propuesta.

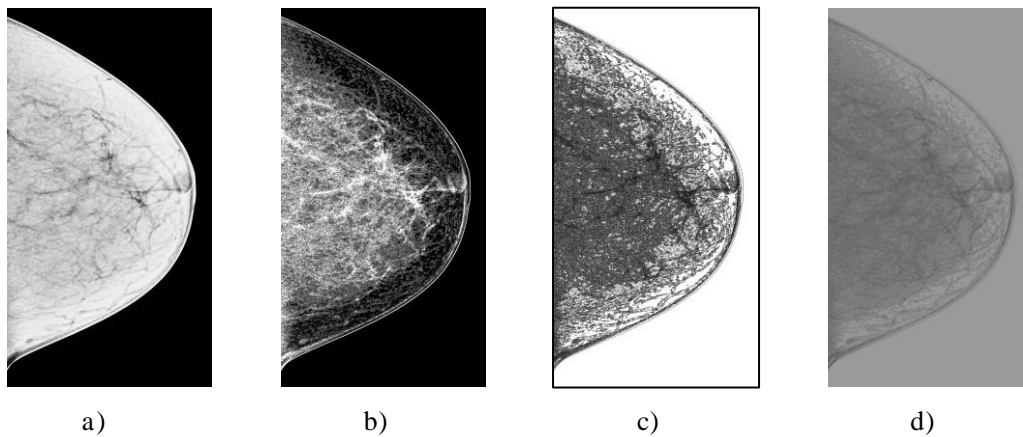


Fig 4 a) Imagen Original (intensidades) de la mama izquierda de una persona con diagnóstico negativo de cáncer de mama. b) Imagen resultante de contraste. c) Imagen resultante de energía. d) Imagen resultante de homogeneidad.

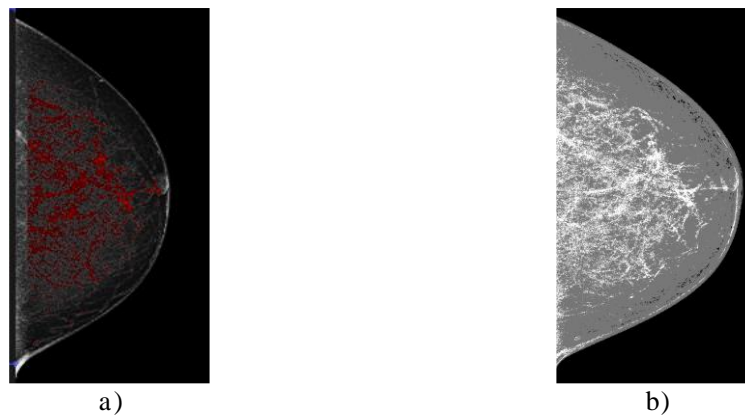


Fig 5 a) Imagen con ajuste manual del histograma de mastografía con diagnóstico negativo de cáncer de mama. b) Correspondiente resultado de segmentación (etiquetado) por el método derivativo del tejido glandular (color blanco), tejido adiposo (color gris) y fondo (color negro).

Finalmente, la Fig. 6 presenta de manera gráfica el resultado cuantitativo del error entre el porcentaje de área densa reportado por el médico radiólogo y el porcentaje de tejido glandular, %TG, (equivalente al área densa) calculado por medio del etiquetado derivativo de las 40 imágenes proporcionadas para este proyecto. El %TG fue calculado mediante (6), usando las imágenes de homogeneidad con ventanas de 7x7 y 32 niveles de intensidad (ver Fig. 2.d y 4.d):

$$\%TG = \frac{\text{TejidoGlandular (Blanco)}}{\text{TejidoGlandular (Blanco)} + \text{TejidoAdiposo (Gris)}} \quad (6)$$

en donde *TejidoGlandular* representa el número total de pixeles etiquetados en color blanco y *TejidoAdiposo* representa el número total de pixeles etiquetados en color gris. Las cuatro gráficas

corresponden al error obtenido en imágenes de mama derecha e izquierda de pacientes con cáncer, e imágenes de mama derecha e izquierda de pacientes sin cáncer (10 imágenes de cada grupo). En esta figura, podemos observar que el promedio de errores de los 4 grupos tiene un rango de 24% a 28%, con presencia de casos atípicos en donde el error es menor a 5%, o mayor a 50%. En este caso es importante destacar que se encontraron algunas inconsistencias en el valor numérico reportado del %TG respecto de las proporciones observables de las imágenes obtenidas por el ajuste manual del histograma, lo que causa el error elevado que se reporta cuantitativamente en este trabajo.

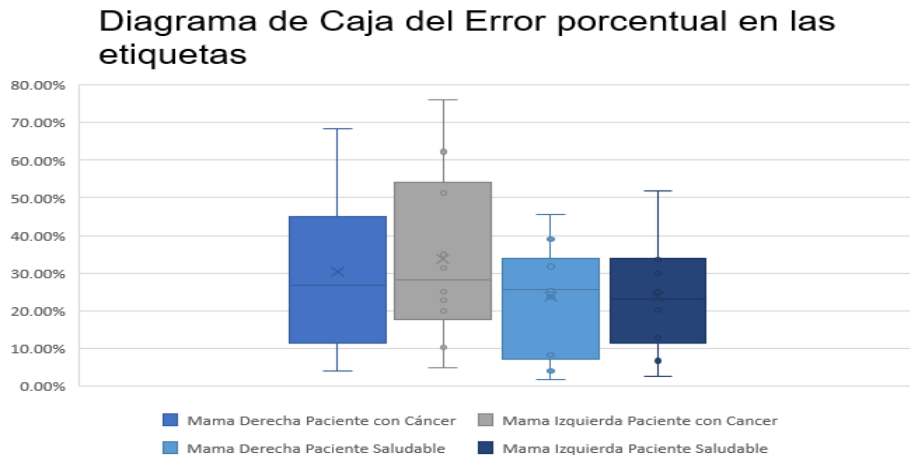


Fig 6. Error porcentual entre estimaciones de porcentaje de tejido glandular realizados con la estrategia propuesta y los valores reportados usando el ajuste manual del histograma.

Por tal motivo, se tiene contemplado a corto plazo, realizar una comparación de la estrategia automática propuesta, con segmentaciones manuales del tejido adiposo y glandular realizadas por especialistas clínicos y de esta manera estimar índices de superposición que permitan disipar dudas del buen desempeño de la estrategia propuesta, ya que se considera que los resultados cuantitativos obtenidos al momento no corresponden a los resultados cualitativos validados por los especialistas. Sin embargo, los resultados encontrados en este trabajo sugieren que la estrategia propuesta puede ser una herramienta útil en apoyo del diagnóstico de cáncer de mama, recordando que el análisis de la proporción y distribución del tejido glandular es un factor importante para la determinación de presencia o ausencia de la enfermedad.

## 4. Conclusiones

El presente trabajo mostró un método para estimar el porcentaje de tejido glandular en imágenes de mastografía de forma automática a partir de la estimación de métricas de textura sobre la imagen, permitiendo realizar una mejor diferenciación del tejido glandular y adiposo respecto de la imagen original de intensidades. Se realizaron pruebas con una base de datos de 40 imágenes de mastografía en donde se tenían pacientes con diagnósticos positivos y negativos de cáncer de mama, y para ambos casos los resultados obtenidos con la estrategia propuesta fueron prometedores de forma cualitativa, particularmente usando la métrica de Homogeneidad.

Al analizar los resultados obtenidos por los algoritmos propuestos de filtrado por textura y clasificación por el método de derivación del histograma podemos concluir que dichos procedimientos pueden representar una herramienta que brinde información útil al médico especialista para el

diagnóstico de cáncer de mama, ya que estos brindan una perspectiva homogeneizada de diferentes patrones e intensidades presentes en cada mastografía que son independientes de la intensidad de los rayos X de la imagen. Si bien es difícil establecer un umbral específico generalizado para la detección automática de cáncer, al momento la información cualitativa obtenida que brinda nuestra propuesta puede representar un ahorro significativo de tiempo al momento de diferenciar regiones de interés en las imágenes de mastografía lo que podría facilitar el diagnóstico oportuno de la enfermedad. Debido a que evita el consumo de tiempo del especialista en el ajuste manual de intensidades para el realce de dichas zonas, el cual puede llegar a ser muy variado entre pacientes.

Como trabajo futuro se plantea realizar una comparación de la estrategia propuesta con segmentaciones manuales realizadas por especialistas clínicos, buscar realizar un preprocesamiento para la eliminación del borde (piel) del seno y el área del pezón de tal manera que se tengan porcentajes más precisos, buscar una forma de optimizar la búsqueda de los umbrales para el etiquetado de los tejidos glandular y adiposo, y explorar la radiómica para los diferentes tipos de cáncer de mama.

## Declaración de conflictos de interés

Los autores declaran no tener ningún conflicto de interés para este trabajo

## Referencias

- [1] Consenso Mexicano sobre diagnóstico y tratamiento del cáncer mamario. Octava revisión, [online]. Colima 2019. Disponible en: <https://consensocancermamario.com>.
- [2] G. Torres-Mejia, B. De Stavola, D. Allen, *et al*, “Mammographic features and subsequent risk of breast cancer: a comparison of qualitative and quantitative evaluations in the Guernsey prospective studies”, *Cancer Epidemiol Biomarkers Prev.*, vol. 14, no. 5, pp. 1052-1059, May 2005.
- [3] A. Gastouniotti, E. F. Conant, D. Kontos, “Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast cancer risk assessment,” *Breast Cancer Res.*, vol. 18, no 1, p 91, Sep 2016.
- [4] S. M. Astley, *et al*, “A comparison of five methods of measuring mammographic density: a case-control study”, *Breast Cancer Res.*, vol. 20, no.1, p. 10, 2018.
- [5] E. E. E. Fowler, *et al*, “Generalized Breast Density Metrics”, *Phys. Med. Biol.*, vol. 64, no. 1, p. 015006, 2018.
- [6] D. E. Rodríguez Obregón, “Cuantificación y representación volumétrica de fibrosis pulmonar idiopática mediante segmentación por contornos activos y métricas de textura en imágenes de TAC,” M.S. tesis, F.C., UASLP., San Luis Potosí, S.L.P., 2018.