

# IDENTIFICACIÓN AUTOMÁTICA DE POSIBLES LESIONES MAMARIAS EN MAMOGRAFÍA DIGITAL

## BREAST LESIONS DETECTION IN DIGITAL MAMMOGRAPHY AN AUTOMATED PRE-DIAGNOSIS

Any Estefany Ruiz Duque<sup>1</sup>, Diana Carolina Arboleda Gómez<sup>1</sup>, Jenny Kateryne Aristizábal Nieto<sup>1</sup>

<sup>1</sup> Grupo de Investigación en Bioinstrumentación e Ingeniería Clínica Programa de Bioingeniería  
Universidad de Antioquia UdeA.

Recibido: 1 Octubre de 2015

Aceptado: 15 Diciembre de 2015

\*Correspondencia del autor: Jenny Kateryne Aristizabal Nieto; E-mail: jenny.aristizabal@udea.edu.co

### RESUMEN

Una de las causas de muerte más comunes entre la población femenina a nivel mundial es el cáncer de seno, además de ser una de las formas de cáncer más prevalente entre otros tipos de cáncer. El diagnóstico temprano y oportuno es un factor determinante para un tratamiento adecuado de esta enfermedad, incrementando las probabilidades de supervivencia de quienes lo padecen. Con el fin de mejorar la eficiencia y la efectividad del diagnóstico del cáncer de seno, se implementó un sistema de análisis de imágenes, cuyo propósito es servir de apoyo para los radiólogos en la detección de lesiones en las mamografías.

Con el fin de encontrar e identificar lesiones mamarias en las mamografías, se implementaron técnicas de segmentación de imágenes en una Región de Interés (ROI), la cual está relacionada con el área donde está concentrada la densidad mamaria. La densidad mamaria se define como el área más brillante en la imagen mamográfica y que está compuesta por tejido fibroglandular y adiposo; es allí donde es probable que las lesiones mamarias sean expuestas. Este estudio provee una metodología dividida en dos etapas, usando dos técnicas de segmentación principales: 1- una técnica de crecimiento de regiones y 2- una técnica de división y fusión de regiones. Este estudio, además, provee una descripción completa del análisis de las imágenes y las herramientas utilizadas para el desarrollo de los algoritmos.

**Palabras claves:** Cáncer de seno; mamografías; lesiones mamarias; análisis de imágenes; segmentación basada en regiones.

## ABSTRACT

Breast cancer is one of the most common causes of death in the female population worldwide and one of the most prevalent cancers among other types of cancer. An early and adequate diagnosis is a key factor for an appropriate treatment, increasing the probability of survival. In order to enhance the efficiency and effectiveness of a diagnosis, an image analysis system was implemented; its purpose was to provide support for radiologists in detection of lesions from mammograms.

Image segmentation techniques were carried out to find breast lesions within the mammograms in the region of interest (ROI), which is related to the area where breast density is concentrated. Breast density is defined as the brightest part on the mammographic image and it is composed by glandular and adipose tissue where breast lesions are likely to be exposed. This study provides a methodology divided in two main segmentation techniques: 1) a region growing technique and 2) split and merge technique. This study also gives a complete description of image analysis and the tools used in it.

**Keywords:** Breast cancer; mammograms; breast lesions; image analysis; region-based segmentation.

## I. INTRODUCCIÓN

Se estima que el cáncer de seno es una de las causas más comunes de muerte en la población femenina a nivel mundial, y una de las formas de cáncer más prevalente entre otros tipos de cáncer. De acuerdo con la Organización Mundial de la Salud, para el año 2012, se presentaron alrededor de 1.38 millones de nuevos casos de cáncer y de estos casos, cerca de 458,000 son muertes causadas por cáncer de seno cada año (1). Un diagnóstico temprano y oportuno es un factor determinante para un tratamiento adecuado de esta enfermedad, incrementando la probabilidad de supervivencia de quienes lo padecen (2). Para este propósito, se han diseñado métodos de evaluación de riesgo de cáncer de mama, con el fin de generar conciencia sobre el impacto del cáncer en la comunidad femenina y además como una metodología de diagnóstico indirecto.

Sin embargo, debido a la falta de conocimiento en la correcta aplicación de estas herramientas por parte de las mujeres, el examen de mamografía se ha convertido en la mejor herramienta de diagnóstico para la enfermedad (3). No obstante, debido al alto número de mamografías generadas cada año, como resultado de múltiples campañas de prevención y a la alta frecuencia de evaluaciones llevadas a cabo por los radiólogos, la tasa de precisión y la uniformidad del diagnóstico tienden a decrecer (2).

Además, debido a la naturaleza radiológica de la mamografía, se derivan durante la lectura por parte de los especialistas, algunos falsos positivos y/o nega-

tivos en el diagnóstico, que conllevan a disminuir la tasa de detección temprana de este cáncer (4).

Como solución a esta problemática, se propone el análisis de imágenes de mamografía digital como una herramienta necesaria de soporte a los radiólogos para detectar pequeñas anomalías en el tejido mamario, las cuales pueden estar relacionadas con masas inusuales como las microcalcificaciones (5). Para este propósito, se implementan técnicas de segmentación de imágenes y algunas operaciones morfológicas, con el fin de identificar la ubicación de posibles lesiones y mejorar la identificación visual de anomalías en las mamografías. Estas técnicas son algoritmos basados en regiones, como la segmentación por crecimiento de las mismas y la segmentación por división y fusión de regiones. Estos métodos son implementados en dos etapas: 1- la primera realiza el preprocesamiento de la imagen y la segmenta en regiones diferentes, tales como el borde del seno, el músculo pectoral y la zona del aire; como resultado se selecciona una región de interés (ROI) que está relacionada con el área donde la densidad mamaria se concentra en la imagen. La densidad mamaria se define como la parte más brillante en la imagen mamográfica y está compuesta por tejido fibroglandular y adiposo, y es allí donde es más probable hallar las lesiones mamarias (6). 2- La segunda etapa del algoritmo permite determinar áreas donde pueden encontrarse las lesiones. Los resultados muestran la efectividad de las técnicas de segmentación implementadas para el reconocimiento de patrones, en un diagnóstico asistido por computadora.

## II. REVISIÓN BIBLIOGRÁFICA

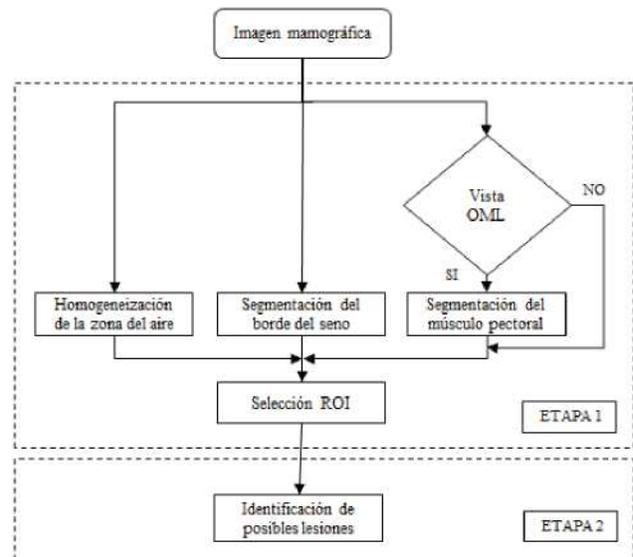
En los últimos años, algunos investigadores de diversas disciplinas han buscado la manera de proporcionar apoyo automático para el diagnóstico del cáncer de mama y con esto se han desarrollado varios métodos automatizados y semiautomatizados que se centran en la detección de lesiones en las imágenes de mamografía digital (7), (8). Estas investigaciones incluyen la extracción de las lesiones desde el fondo del pecho por binarización, para ser caracterizadas con base en su forma (9); otros autores proponen un esquema automático para la detección de masas anormales a través de segmentación anatómica de la región de interés mamaria (ROI) (10); también se propone un sistema de apoyo para el diagnóstico del cáncer de seno que consiste en superponer imágenes de mamografía y sugerir una anotación textual de las características del diagnóstico (11).

## III. PRUEBA EXPERIMENTAL

Para el desarrollo de las pruebas, se seleccionó un grupo de 40 exámenes mamográficos con el fin de evaluar el rendimiento global del algoritmo desarrollado. Cada examen mamográfico está compuesto por cuatro imágenes que fueron evaluadas individualmente, dos de ellas refieren una vista craneocaudal (CC) de cada seno, y las dos restantes refieren la vista oblicua mediolateral (OML) (12). Este conjunto de imágenes incluye diagnósticos normales y patológicos.

## IV. MÉTODOS PROPUESTOS

La metodología implementada fue realizada en dos etapas: 1) selección de la ROI y 2) detección de posibles lesiones. La primera de estas etapas permitió resaltar la región mamaria; para ello fue necesario realizar un filtrado que permitió homogeneizar la zona relacionada con el aire en la imagen, implementando operaciones morfológicas; además se realizó la segmentación del músculo pectoral utilizando técnicas de segmentación por regiones y finalmente se segmentó el borde del seno; estos tres procedimientos permitieron la identificación de la ROI, que contiene la densidad mamaria. En la segunda etapa, se detectaron posibles lesiones mediante la aplicación de la técnica de división y fusión. La metodología implementada se muestra en la Figura. 1.



**Figura. 1.** Algoritmo de detección de posibles lesiones. Incluye preprocesamiento y procesamiento de las mamografías.

### A. Selección de la ROI

Para seleccionar la ROI apropiada en cada imagen, fue necesario extraer de cada imagen todas las áreas presentes con niveles de intensidad más altos que el área de la densidad mamaria. Para lograrlo, se realizó un primer procedimiento que permitió identificar la orientación de la imagen.

1) *Orientación de la imagen:* La identificación de la orientación de la imagen (izquierda o derecha) se realizó mediante la comparación de la primera y la última columna de la matriz de la imagen, teniendo en cuenta sus valores de intensidad. La posición de la columna con el valor más alto de intensidad corresponde a la orientación y por lo tanto a la posición de la imagen mamaria.

2) *Homogeneización de la zona del aire:* Para la homogeneización del aire se utilizaron operaciones morfológicas para obtener una zona completamente homogénea dentro de la imagen, dicha zona es aquella en la cual no se encuentra ningún tejido en la imagen.

La dilatación y la erosión son definidas matemáticamente como operaciones entre conjuntos. Dada una imagen  $R$  definida como una matriz  $n \times m$  en el espacio  $Z^2$ ; y  $A, B$  como miembros del conjunto de un mismo espacio, la dilatación de  $A$  por  $B$ , denotado por  $A \oplus B$ , es el conjunto de todos los desplazamientos  $z$ , de tal manera que la reflexión del conjunto  $B$  (denotada por  $\widetilde{B}$ ) coincide por al menos un elemento

(ec. 1). La erosión de  $A$  por  $B$ , denotado  $A \ominus B$ , es el conjunto de todos los puntos  $z$  de tal manera que  $B$ , traducido por  $z$ , está contenido y se define como se muestra en (ec. 2). Se define  $B$  como una forma llamada elemento estructurante, que controla objetos gruesos o delgados. La dilatación es una operación que permite ampliar objetos en una imagen binaria; en cambio, la erosión es una operación que encoge o adelgaza objetos en una imagen binaria (13), (14), (15).

$$A \ominus B = \{z | [(B)_z \subseteq A]\} \quad (\text{ec. 1})$$

$$A \oplus B = \{z | (B)_z \cap A \neq \emptyset\} \quad (\text{ec. 2})$$

También se implementó otra operación morfológica llamada llenado de huecos. Esta es una reconstrucción morfológica basada en la formulación de que un hueco o agujero es un conjunto de píxeles del fondo de la imagen al que no se puede llegar llenando el fondo desde el borde de la imagen; es decir, un agujero es un área de píxeles oscuros rodeados de píxeles más claros (16).

3) *Segmentación del músculo pectoral*: el alto contraste, gran tamaño y brillo del músculo pectoral pueden resultar en un falso positivo al momento de detectar las posibles lesiones; con el fin de extraer este músculo de la imagen OML se realizó una técnica de segmentación basada en regiones, la cual es una técnica que permite dividir una imagen en diferentes zonas (13), (17). Dada una imagen  $R$ , la segmentación puede ser vista como una partición de  $R$  en  $n$  subregiones, tal que:

- La conexión de todas las subregiones es la región completa:  $\bigcup_{i=1}^n R_i = R$
- $R_i$  es una región conectada  $i=1,2,\dots,n$ .
- $R_i \cap R_j = \emptyset$  para todo  $i$  y  $j$ ,  $i \neq j$ .
- $P(R_i) = \text{VERDADERO}$  para  $i=1,2,\dots,n$ .
- $P(R_i, R_j) = \text{FALSO}$  para  $i \neq j$ .

Donde,  $P(R_k)$  es un predicado lógico definido sobre los puntos en el conjunto  $R_k$  y  $\{\}$  es el conjunto vacío.

La segmentación por crecimiento de regiones se basa en el agrupamiento de píxeles o subregiones, a partir de una "semilla" que podría ser un punto o un conjun-

to de puntos con ciertos criterios predefinidos. A partir de esta semilla, las regiones pueden crecer mediante la unión de los píxeles vecinos con características similares a la semilla (tales como nivel de gris, textura, color y forma) (18). El criterio seleccionado para esta técnica de segmentación fue el nivel de gris de la semilla, porque los píxeles del músculo pectoral son muy similares entre ellos. La intensidad de cada pixel cercano a la región es fácilmente comparada con el valor promedio de la intensidad de la región completa.

El principal reto de esta etapa de la investigación fue determinar el lugar adecuado para seleccionar la semilla inicial y elegir los criterios para evaluar los posibles píxeles que serían incluidos en la región. El lugar para ubicar la semilla: derecha o izquierda, fue elegido dependiendo del tipo de vista de la mamografía. Por otra parte, como es particular en las imágenes de mamografía, los píxeles del músculo pectoral pueden tener valores de alta intensidad, por lo que se seleccionó un umbral basado en el valor de la intensidad del píxel donde se plantó la semilla inicial. Finalmente, se obtuvo una máscara de músculo pectoral, con la cual fue posible segmentar esta región.

4) *Segmentación del borde del seno*: el borde del seno es ligeramente brillante en comparación con el resto del tejido mamario y podría ser procesado por el algoritmo como posibles lesiones. Es necesario entonces segmentar este borde a través de una técnica llamada Extracción de bordes.

El borde de un conjunto  $A$ , denotado por  $\beta(A)$ , puede obtenerse por la erosión de  $A$  por  $B$  y realizar la diferencia de conjuntos entre  $A$  y su erosión como se muestra en (ec. 3), donde  $B$  es el elemento estructurante (13).

$$\beta(A) = A - (A \ominus B) \quad (\text{ec. 3})$$

#### B. Detección de posibles lesiones

La segunda etapa del método fue desarrollada con el fin de detectar posibles lesiones en imágenes de mamografía; y para esto se llevó a cabo una técnica de segmentación basada en regiones, conocida como División y Fusión (Fig. 1).

Dada  $R$ , la región de toda la imagen y  $P$ , los criterios de semejanza, se define el método para segmentar cualquier región  $R_i$  como la división en cuatro cuadrantes disjuntos si se cumple que  $P(R_i) = \text{Falso}$ . Cuando no

existe más división es posible unir regiones adyacentes  $R_j$  y  $R_k$  para el cual  $P(R_j \cup R_k) = Verdadero$ , y se detendrá cuando la fusión ya no sea posible (13), (19). A través de este procedimiento, se detectaron posibles lesiones en las mamografías como se muestra en los resultados.

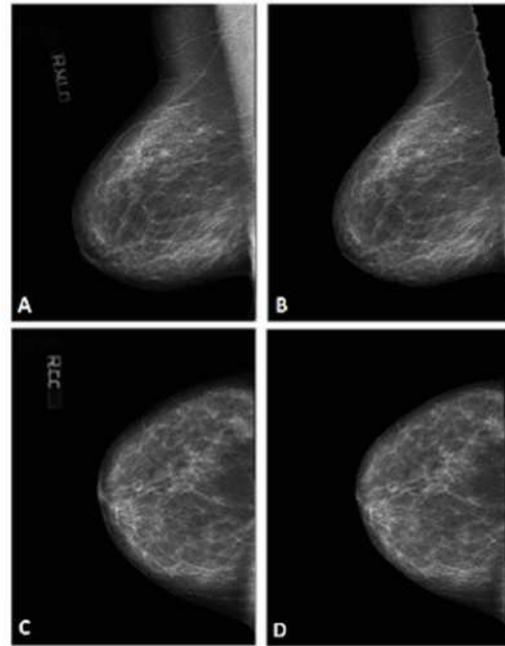
## V. RESULTADOS

Los resultados para cada etapa del algoritmo diseñado se muestran en las Figura. 2, Figura. 3 y Figura. 4. La ROI seleccionada se delimitó exitosamente. La zona del aire fue definida y se removieron las marcas de identificación de la imagen. Además se removió completamente el borde del seno y finalmente se segmentó el músculo pectoral (ver Figura. 2-A y 2-B). A diferencia de la imagen OML, en las imágenes CC, el músculo pectoral no fue removido; por lo tanto solamente se llevó a cabo la homogeneización del aire y la remoción de las marcas de identificación (ver Figura. 2-C y 2-D).

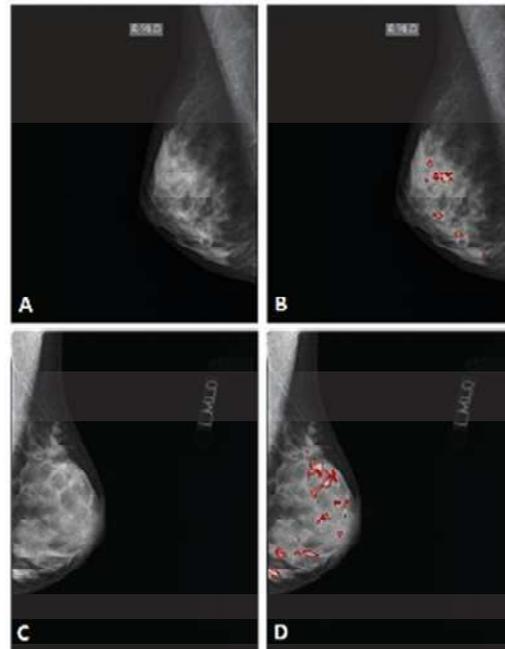
La segmentación del músculo pectoral es importante debido a su alto brillo en comparación con el resto del tejido mamario, además, la segmentación de este permite aislar la región de interés que se limita a la densidad de la mama y por lo tanto, puede evitar posibles falsos diagnósticos.

La detección de posibles lesiones en la ROI se logró mediante la técnica de segmentación de división y fusión de regiones. Los parámetros estadísticos con los que fue construido el algoritmo fueron la media y la desviación estándar. La identificación de las posibles áreas de lesiones se realizó con una precisión apropiada. Esta metodología tuvo una efectividad satisfactoria; las regiones de las posibles lesiones fueron identificadas y resaltadas con líneas rojas como se muestra en las Figura. 3 y Figura. 4.

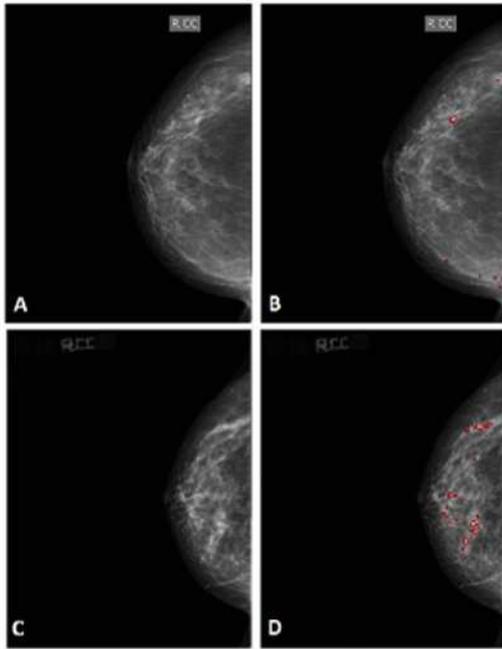
El algoritmo propuesto en este trabajo presentó los resultados esperados en gran parte de las imágenes evaluadas. De acuerdo a los resultados, aquellas áreas con posibles lesiones fueron detectadas satisfactoriamente. Sin embargo, todos los resultados deben ser analizados y evaluados por un especialista, análisis que se realizó en una etapa posterior a este proyecto, con el fin de validar clínicamente los resultados de los algoritmos implementados.



**Figura. 2.** Mamografías originales (A,C) y la ROI seleccionada (B,D), de la vista oblicua mediolateral (A,B) y de la vista craneocaudal (C,D).



**Figura. 3.** Vista mamográfica oblicua mediolateral después de ser procesada. Imagen original (A,C). Imagen procesada (B,D), posibles lesiones resaltadas con líneas rojas.



**Figura. 4.** Vista mamográfica craneocaudal después de ser procesada. Imagen original (A,C). Imagen procesada (B,D), posibles lesiones resaltadas con líneas rojas.

**VI. TEST DE VALIDACIÓN**

Con el fin de verificar la eficiencia de la metodología propuesta, se llevó a cabo un procedimiento de validación en el cual se realizó la comparación del número de píxeles con alta intensidad en la imagen de

mamografía, antes y después del procesamiento de la imagen.

Debido a que las lesiones mamarias presentan una alta intensidad respecto al resto del seno, se seleccionó un umbral correspondiente al 70% de la distribución del histograma. Para la etapa 1, se realizó una evaluación preliminar del número de píxeles de la imagen procesada con base en el umbral, es decir, se obtuvo el número de píxeles cuya intensidad fue mayor que el valor del umbral, obteniendo así la región más brillante de la muestra. Luego, en la etapa 2, nuevamente se calculó la cantidad de píxeles con las mayores intensidades; debido a que el propósito es determinar las posibles lesiones mamarias, se esperaba conservar la cantidad de píxeles de alta intensidad antes y después de la aplicación del algoritmo, confirmando los resultados preliminares obtenidos en la etapa 1. Sin embargo, debido a las diferencias entre los resultados en ambas etapas y a la cantidad de regiones que falsifican la detección, el número de píxeles de alta intensidad varía de una etapa a otra, haciéndose necesario calcular la eficiencia de cada una de las etapas.

La clasificación de los resultados de la validación se muestra en la tabla 1. La clase 1 representa alta eficiencia (> 85%); la clase 2 corresponde a aquellas imágenes para las cuales se obtuvo una eficiencia entre 70% y 85%, y finalmente, la clase 3 representa una eficiencia menor a 70%.

**TABLA 1.** TEST DE VALIDACIÓN. NÚMERO DE IMÁGENES POR CADA CLASE DE ACUERDO A LA EFICIENCIA DEL ALGORITMO DE SEGMENTACIÓN PARA LA ETAPA 1 Y ETAPA 2.

Tipo de Imagen	ETAPA 1			ETAPA 2		
	Clase 1	Clase 2	Clase 3	Clase 1	Clase 2	Clase 3
CCD	40	0	0	29	4	7
CCI	40	0	0	27	6	7
OMLD	28	4	8	27	8	5
OMLI	29	2	9	31	5	4

\* CCD, OMLD: derecha. CCI, OMLI: Izquierda

**VII. CONCLUSIONES**

Los resultados obtenidos al aplicar en conjunto todo el algoritmo desarrollado, en varias mamografías digitales de diferentes pacientes, demostraron la efectividad de las técnicas propuestas. Las zonas donde posiblemente se encuentra una lesión mamaria, fueron identificadas satisfactoriamente en la mayoría de las mamografías evaluadas.

El algoritmo propuesto promete ser una herramienta de soporte diagnóstico a los radiólogos y de esta manera aumentar los casos de detección temprana de cáncer de seno. El análisis de mamografía digital, en conjunto con el procesamiento digital de imágenes, demostraron ser un método efectivo para la prevención del cáncer de seno.

Los resultados derivados de este proyecto fueron eva-

luados y comparados por un equipo de especialistas en análisis de mamografía, en una fase posterior a este. Este proyecto hace parte de un macroproyecto titulado “Identificación de Lesiones en Mamografía Digital”.

## AGRADECIMIENTOS

Este trabajo ha sido financiado parcialmente por la Universidad de Antioquia, proyecto CODI SOS11-2-01 Estrategia de Sostenibilidad 2011-2012. Las imágenes fueron facilitadas por el centro de diagnóstico CEDIMED.

---

## BIBLIOGRAFÍA

1. Ferlay J, Soerjomataram, I. Ervik M, Dikshit R, Eser S, Mathers C, Robelo M, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: <http://globocan.iarc.fr>
2. Fernández McCann DS. Ayuda al Diagnóstico Médico de la Mamografía Mediante Técnicas de Procesamiento Digital de Imágenes. *Rev Fac Ing Univ Antioquia*. 2001;(22):116–25.
3. Bandyopadhyay SK. Pre-processing of Mammogram Images. *Int J Eng Sci Technol*. 2010;2(11):6753–8.
4. Hadjiiski L, Sahiner B, Chan H-P, Helvie MA. Automated Registration and Classification Techniques for Interval Change Analysis in Mammography. *Fuzzy Inf Process Soc 2005 NAFIPS 2005 Annu Meet North Am*. 2005;55–60.
5. Freitas da Cruz C. Automatic analysis of Mammography Images. Universidade do Porto; 2011.
6. Yaffe MJ. Mammographic density. Measurement of mammographic density. *Breast Cancer Res* [Internet]. 2008 Jan [cited 2014 Nov 1];10(3):209. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2481498&tool=pmcentrez&rendertype=abstract>
7. Sampat MP, Bovik AC. Detection of spiculated lesions in mammograms. *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*. 2003. p. 810–813 Vol.1.
8. Guerra A, Rivas J. Detección de microcalcificaciones en imágenes mamográficas usando redes neuronales. *Rev la Fac Ing Univ Cent Venez. scielo\_ven*; 2011;26:7–14.
9. Szekely N, Pataki B. Detecting lesions in a mammogram. *Video/Image Processing and Multimedia Communications, 2003. 4th EURASIP Conference focused on*. 2003. p. 113–118 vol.1.
10. Maitra IK, Nag S, Bandyopadhyay SK. Automated Digital Mammogram Segmentation for Detection of Abnormal Masses Using Binary Homogeneity Enhancement. *Indian J Comput Sci Eng*. 2011;2(3):416–27.
11. Narváez Espinoza FR. Sistema de Anotación para Apoyo en el Seguimiento y Diagnóstico de Cáncer de Seno. Universidad Nacional de Colombia; 2010. p. 85.
12. Yam M, Brady M, Highnam R, Behrenbruch C, English R, Kita Y. Three-dimensional reconstruction of microcalcification clusters from two mammographic views. *Medical Imaging, IEEE Transactions on*. 2001. p. 479–89.
13. Gonzalez RC, Woods RE, Eddins SL. *Digital Image Processing Using MATLAB*. 2nd ed. New Jersey: Gatesmark Publishing; 2009.
14. Gonzalez RC, Woods RE. *Digital Image Processing*. Pearson/entice Hall; 2008.
15. Yu Z, Zhao Y, Wang X. Research advances and prospects of mathematical morphology in image processing. *Cybernetics and Intelligent Systems, 2008 IEEE Conference on*. 2008. p. 1242–7.
16. Mathworks, Inc., MATLAB® Image Processing Toolbox, The MathWorks, Inc, USA. 2010.
17. Yim PJ, Choyke PL, Summers RM. Gray-scale skeletonization of small vessels in magnetic resonance angiography. *Medical Imaging, IEEE Transactions on*. 2000. p. 568–76.
18. Almi’ani MM, Barkana BD. Automatic segmentation algorithm for brain MRA images. *Systems, Applications and Technology Conference (LISAT), 2012 IEEE Long Island*. 2012. p. 1–5.
19. Karmakar GC, Dooley LS, Murshed M. Image segmentation using modified extended fuzzy rules. *Signal Processing, 2002 6th International Conference on*. 2002. p. 941–944 vol.2.