



VI CONGRESO NACIONAL DE TECNOLOGÍA APLICADA A CIENCIAS DE LA SALUD

4, 5 y 6 de junio de 2015

“Generación de Nuevas Técnicas
de Diagnóstico y Tratamiento”

DETECCIÓN DE MASAS Y ASIMETRÍAS EN MAMOGRAFÍAS CON DISTINTOS PATRONES PARENQUIMALES, UTILIZANDO ENTROPÍA DE TSALLIS-HAVRDA-CHARVÁT.

S. L. de la Fuente-Bermúdez ^a; F. J. Ornelas-Rodríguez ^a; J. J. González-Barbosa^a; J. B. Hurtado-Ramos ^a;
M. Romero del Valle ^b; y J. A. Salinas-Rincón^c,

^aCentro de Investigación en Ciencia Aplicada y Tecnología Avanzada, Instituto Politécnico Nacional, Santiago de Querétaro, Qro., salufub@ieee.org, ornelasr@ipn.mx, jgonzalezba@ipn.mx, jbautistah@ipn.mx

^bHospital General de Salamanca, Salamanca, Gto., marianoromero@hotmail.com

^cUnidad de Especialidades Médicas Dedicada a la Detección y Diagnóstico del Cáncer de Mama, Santiago de Querétaro, Qro., sara_bas_eleasar@hotmail.com

RESUMEN:

La alta incidencia de afección del cáncer de mama representa un gran impacto en la sociedad mundial, pues representa el 16% de todos los cánceres femeninos y las proyecciones indican que el número de mujeres diagnosticadas de cáncer de mama en América aumentará en un 46% en el 2030 (INEGI, 2014). Mientras que estudios han corroborado que la detección temprana por medio de mamografías ha reducido el índice de mortalidad de un 28% a un 65%, la detección de lesiones quísticas y microcalcificaciones en mamografías de patrones parenquimales homogéneos es, hasta la fecha, una tarea compleja incluso para radiólogos especializados. Sin embargo, el avance de tecnologías de la información ha permitido crear asistentes computarizados para la detección (CADE) de lesiones en mamografías como sustituto de segundas inspecciones de mamografías de médicos especializados. Los autores desarrollamos un CADE que implementa técnicas de procesamiento de imágenes en mamografías digitales para optimizar el hallazgo de lesiones, entre las que se encuentran masas y asimetrías. En el presente trabajo, reportamos la detección de asimetrías y masas sólidas y quísticas en patrones parenquimales diversos (I al V). Resaltamos las regiones de mayor densidad y segmentamos las zonas sospechosas mediante entropía de Tsallis-Havrda-Charvát en las sub-imágenes posterior, media, y anterior de proyecciones Medio Lateral Oblícuo (MLO) y Cráneo-Caudal (CC), combinando la técnica de inspección descrita en BIRADs y propuesta por Lázlo Tábár, y la técnica de detección de lesiones propuesta por Burcin Kurt en 2014. Corroboramos los resultados con el diagnóstico de una base de datos de 300+ estudios de mamografías con hallazgos Birads 3 a 5 del Hospital General de Salamanca, Guanajuato y de la Unidad de Especialidades Médicas Dedicada a la Detección y Diagnóstico del Cáncer de Mama (UNEME-DEDICAM), Querétaro; algunos de estos hallazgos han sido comprobados por inspección de ecografía y/o biopsia.

1. INTRODUCCIÓN

La detección temprana del cáncer de mama es un reto que requiere de segundas lecturas o asistentes computarizados para optimizar la detección de lesiones en mamografías, incrementar el hallazgo de lesiones pequeñas, reducir el índice de mortalidad por cáncer de mama, y producir ahorros en el sector salud. La manera



VI CONGRESO NACIONAL DE TECNOLOGÍA APLICADA A CIENCIAS DE LA SALUD

4, 5 y 6 de junio de 2015

“Generación de Nuevas Técnicas
de Diagnóstico y Tratamiento”

más confiable de detectar cáncer de mama en estadios tempranos es la revisión regular en mujeres, aparentemente sanas. Estudios comprueban que la detección temprana del cáncer de mama por mamografías reduce el índice de mortalidad de un 28% a un 65%. En México, la auto detección es muy alta, sin embargo, sólo entre el 5% y 10% de los tumores son identificados en estadio I.

La demanda de trabajo no permite que la segunda lectura de especialistas sea una opción, por lo que los asistentes computarizados se han vuelto una herramienta indispensable para los especialistas que prefieren hacer uso de la tecnología para optimizar en tiempo y calidad su trabajo. En los últimos años, investigadores han mejorado los algoritmos de procesamiento de mamografías para la detección automática de calcificaciones, masas, y asimetrías; obteniendo una sensibilidad promedio de 80% a un 95% en los estudios reportados. La mayoría de las investigaciones utilizan la base de datos MIAS MiniMammographic Database y/o Digital Database for Screening Mammography (DDSM). En esta investigación, los autores reportamos los resultados CADe en una base de datos actual del Hospital General de Salamanca y la UNEME-DEDICAM Querétaro, comparando los resultados con el diagnóstico de un radiólogo especializado en cada una de las dos unidades médicas. El análisis se realiza en cuatro etapas. En el *pre-procesamiento* recortamos la etiqueta de la mamografía, segmentamos el músculo del pectoral, y preparamos la imagen para ser procesada; en el *mejoramiento* de la imagen destacamos las zonas de interés utilizando filtrado Top-hat, transformada bio-ortogonal de wavelet, CLAHE, y filtrado anisotrópico; en la *segmentación* analizamos la entropía de Tsallis-Havrda-Charvát de la imagen y clasificamos los píxeles en tres zonas (fondo de la imagen, tejido de fondo, y regiones de interés ROI), encontrando el umbral óptimo para segmentar los hallazgos sospechosos, volvemos a umbralizar de ser necesario (OTSU-N); para la extracción de características analizamos la intensidad, forma, textura, y ubicación de las lesiones; y la clasificación utiliza un algoritmo AdaBoost.

Las mamografías están tomadas en dos proyecciones, la Medio Lateral Oblicua o MLO y la Cráneo Caudal o CC y la técnica de inspección de mamografías es un punto clave en esta investigación. Basándonos en los reportes del Dr. Lázlo Tábár, procesamos las mamografías en zonas posterior, media, y anterior. Al software está adaptado a una interfaz gráfica para fines demostrativos, desarrollada en Matlab®.

2. TEORÍA

2.1 Pre-procesamiento

Segmentamos el músculo del pectoral y las etiquetas, utilizando un filtro medio de la imagen [16].

2.2 Mejoramiento de la imagen

El mejoramiento de la imagen y la segmentación de los objetos son altamente importantes para una correcta clasificación de lesiones en mamografías [16,20-21], más sin embargo, existen técnicas ampliamente costosas [1], es por esto, que en este proyecto se propone optimizar el proceso, utilizando imagen integral y una segmentación por pasos [19].

VI CONGRESO NACIONAL DE TECNOLOGÍA APLICADA A CIENCIAS DE LA SALUD

4, 5 y 6 de junio de 2015

“Generación de Nuevas Técnicas
de Diagnóstico y Tratamiento”

Transformada Top-hat.

La transformada Top-hat es un método útil para extraer objetos similares a los elementos de estructura. Basándose en esta propiedad, Zhang [5,9] deduce que se puede suprimir el fondo de la imagen, obteniendo así las regiones de interés (ROI). El elemento estructural es:

$$B_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (1)$$

El fondo menos la imagen suprimida, $F_2(x, y)$ es obtenida aplicando la transformada Top-hat:

$$F_2(x, y) = F(x, y) - (F(x, y) \circ B_3(x, y)) \quad (2)$$

Donde, “ \circ ” denota la operación morfológica “open.”

Transformada wavelet.

Descomponemos la imagen utilizando wavelet y se umbraliza para reducir el ruido. Después de suprimir el ruido de la imagen descompuesta, se reconstruye la imagen utilizando reconstrucción wavelet. Se propone utilizar la transformada bio-ortogonal de wavelet. En comparación con la transformada ortogonal de wavelet, las imágenes no tienen efecto de aliasing [16]. Después de la descomposición, la imagen dada es realizada por un coeficiente de aproximación y seis coeficientes detallados. Los últimos contienen coeficientes horizontales, verticales, y diagonales. Un umbral apropiado es utilizado para suprimir el ruido de los coeficientes [25]. El umbral es calculado mediante:

$$T = \frac{j}{2} \sqrt{\max(d_i)} \quad (3)$$

Donde j es el nivel al cual el umbral T es calculado. La imagen resultante después de implementar el wavelet, será una imagen mejorada en contraste con ruido suprimido [Sakellaropoulos, A wavelet based spatially].

Ecuación de contraste limitado del histograma (CLAHE).

CLAHE ha generado buenos resultados en el mejoramiento de imágenes médicas [20]. CLAHE opera en pequeñas regiones de la imagen en lugar de en toda la imagen. El mejoramiento de contraste es local, por lo que el histograma de la región de salida se ajusta aproximadamente al histograma especificado por el parámetro de distribución. Las vecindades de la zona local son entonces combinadas utilizando interpolación bilinear para eliminar artificialmente propiedades inducidas. El contraste, especialmente en áreas homogéneas, puede ser limitado para evitar incrementar cualquier ruido que se pueda presentar en la imagen.

Filtro anisotrópico difuso.

El objetivo principal de los algoritmos de difusión en un procesamiento de imágenes es remover el ruido por medio de ecuaciones diferenciales parciales (PDE), las cuales son utilizadas, principalmente, para propósitos de suavizado y restauración. En el enfoque introducido por Perona y Malik [20], es posible suavizar una región de preferencia sin afectar otras regiones, es decir, el suavizado no tiene interacción entre regiones.



VI CONGRESO NACIONAL DE TECNOLOGÍA APLICADA A CIENCIAS DE LA SALUD

4, 5 y 6 de junio de 2015
“Generación de Nuevas Técnicas
de Diagnóstico y Tratamiento”

El coeficiente utilizado para detener la difusión sobre los contornos de la imagen está dado por:

$$\frac{\partial I(x,y,t)}{\partial t} = \text{div}(c(|\nabla I|)\nabla I) \quad (4)$$

Donde, ∇ es el operador gradiente, div es el operador divergente, y c es el coeficiente de difusión, que puede ser una gaussiana.

2.3 Segmentación

Se ha demostrado que la entropía de Havrda y Charvát [16] es un parámetro útil para el procesamiento de mamografías, utilizándose para la detección de anomalías. Por lo tanto, se analiza la entropía mediante Havrda y Charvat [9] para identificar las regiones sospechosas.

La entropía Havrda y Charvát es calculada por:

$$HC = \frac{1}{1-\alpha} (\sum_{i=0}^{Ng-1} H_i^\alpha - 1) \quad (5)$$

$$\alpha \neq 1, \alpha > 0$$

$$H_i = \frac{h_i}{MN} \quad (6)$$

Donde, α es el parámetro real positivo, M y N representan las dimensiones del mamograma, h_i es el histograma con niveles de gris de 0 a 255.

Un mamograma es una combinación de tres objetos individuales [14, 16]: fondo del seno, tejido de fondo, y regiones sospechosas. Por lo anterior, los píxeles en el mamograma son divididos en tres clases: C_1 (fondo del seno), C_2 (tejido de fondo) y C_3 (regiones sospechosas), respectivamente, con un nivel de gris t . C_1 es el conjunto de píxeles con niveles de 0 a k , C_2 es el conjunto de píxeles con niveles de 0 a t con un incremento $k++$, y el resto de los píxeles pertenecen a la clase C_3 . La sección C_3 no provee ninguna información en el diagnóstico, por lo que puede ser ignorada. Khenhra propuso calcular el promedio de los niveles de gris de los tejidos del seno, excluyendo el fondo del seno, usando la siguiente fórmula:

$$k = \frac{1}{MN} \sum_{i,j \in R}^M \sum_{i,j \in R}^N I(i,j) \quad (7)$$

Donde, R es la región del seno e $I(i,j)$ es el valor en niveles de gris de la imagen en las coordenadas (i,j) .

De acuerdo a las definiciones del método de entropía de Havrda y Charvat, la entropía de los píxeles de tejido de fondo y la entropía de los píxeles de las regiones sospechosas pueden ser definidas como:

$$HC_{TB}(t) = \frac{1}{1-\alpha} \left[\sum_{i=k}^t \left(\frac{p_i}{p_t - p_{k-1}} \right)^\alpha - 1 \right] \quad (8)$$

$$p_t = \sum_{j=1}^t \frac{h_j}{MN} \quad (9)$$

$$p_{k-1} = \sum_{i=1}^{k-1} \frac{h_i}{MN} \quad (10)$$

$$HC_{SR}(t) = \frac{1}{1-\alpha} \left[\sum_{i=t+1}^L \left(\frac{p_i}{1-p_t} \right)^\alpha - 1 \right] \quad (11)$$

Donde, h es el histograma de niveles de gris en el mamograma, t es el valor del umbral del mamograma, L es el valor máximo de niveles de gris en el mamograma, que es igual a 255.



Para descomponer los dos subsistemas estadísticamente independientes, el tejido de fondo y las regiones sospechosas, se utiliza la siguiente fórmula [16]:

$$HC_{TB}(t) + HC_{SR}(t) + (1 - \alpha)HC_{TB}(t)HC_{SR}(t) \quad (12)$$

Este método se basa en obtener el umbral óptimo t_{opt}^{HC} mediante:

$$t_{opt}^{HC} = Argmax_{t=k}^L [HC_{TB}(t) + HC_{SR}(t) + (1 - \alpha)HC_{TB}(t)HC_{SR}(t)] \quad (13)$$

En función de $t \in [k, \dots, L]$, el umbral óptimo el valor máximo. En este sentido, las regiones sospechosas pueden ser separadas del tejido de fondo. Definir un valor óptimo de α es muy importante para la segmentación debido a los distintos patrones parenquimales y anomalías, por lo que se define un valor dinámico de α [9] como:

$$\alpha = 1 - \left(\frac{k}{255}\right) \quad (14)$$

El promedio del valor de nivel de gris de los píxeles sospechosos deben ser mayores al valor de k, por lo que se normaliza el valor de k y se sustrae del valor máximo, 1. Obteniendo así el valor de α de la imagen de estudio. Estudios han demostrado que utilizar un valor dinámico de α es más efectivo que utilizar un valor predeterminado de α [16].

2.4 Obtención de características.

Después de segmentar los hallazgos sospechosos, obtenemos características internas y externas de los objetos detectados: medición del área, medición de circularidad, medición de densidad, varianza de la intensidad, entropía, forma, y ubicación de la lesión.

2.5 Clasificación.

AdaBoost es un método particular de entrenamiento de un clasificador tipo boost y utiliza múltiples iteraciones para generar un aprendizaje robusto, es decir, genera un aprendizaje robusto implementando aprendizajes débiles en cascada. En cada iteración, los objetos con las características más débiles se van descartando, dejando solamente los objetos que pertenecen a la clasificación. Un clasificador boost de la forma:

$$F_T(x) = \sum_{t=1}^T f_t(x) \quad (15)$$

Donde, cada f_t es un aprendizaje débil que tiene un objeto x de entrada y un indicador de clasificador de salida [19].

3. PARTE EXPERIMENTAL

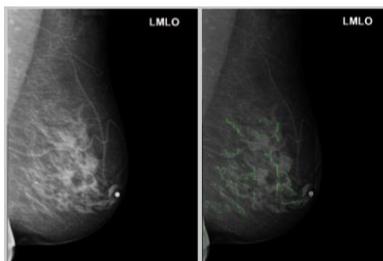


Figura 1. Izq. Imagen original de estudio. Der. Imagen con detección de lesiones.



3.1 Pre-procesamiento.

Segmentamos el pectoral y dividimos la imagen en tres segmentos: posterior, medio, y anterior.

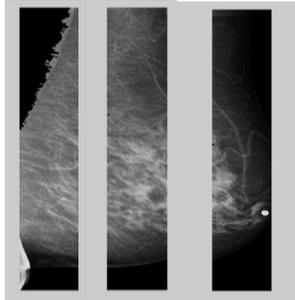


Figura 2. Segmentación del músculo del pectoral, supresión de etiqueta, e implementación de zonas posterior, media, y anterior.

3.2 Mejora de la imagen.

La mejora consiste en cuatro etapas. Los resultados experimentales demuestran la necesidad de utilizar transformada Top-hat y Bottom-hat.

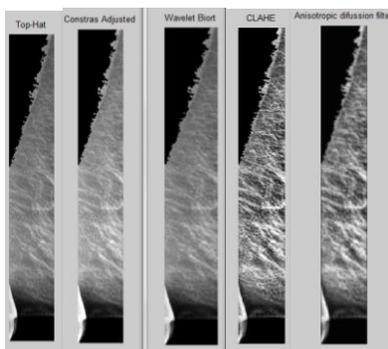


Figura 3. Mejoramiento de la imagen: a) Transformada Top-hat, b) Ajuste de contraste, c) Transformada bio-ortogonal de wavelet, d) CLAHE, y e) Filtrado anisotrópico difuso.

3.3 Segmentación.

La segmentación es adecuada para una etapa de mejora adecuada.



Figura 4. Segmentación de hallazgos sospechosos por umbral óptimo encontrado con entropía de Tsallis-Havrda-Charvát.



VI CONGRESO NACIONAL DE TECNOLOGÍA APLICADA A CIENCIAS DE LA SALUD

4, 5 y 6 de junio de 2015

“Generación de Nuevas Técnicas
de Diagnóstico y Tratamiento”

3.4 Obtención de características.

Las características analizadas dan un panorama general para determinar si se trata de una asimetría o masa.

4. CONCLUSIONES

La primera etapa de mejora de la imagen necesita de operadores morfológicos de apertura y cierre.

La detección de lesiones mejoró con la inspección por zonas posterior, media, y anterior de la mamografía. Es decir, cuando con la implementación detecta tres objetos, sin esta implementación, el software detecta uno, quedando generalizadas como un solo objeto.

Es recomendable utilizar otro clasificador, pues aunque AdaBoost es más resistente a overfitting que muchos algoritmos de machine learning, también es sensible a outliers y datos ruidosos.

BIBLIOGRAFÍA

- [1] INEGI, "Estadísticas a propósito del día internacional contra el cáncer de mama," 2013.
- [2] American College of Radiology, "BI-RADS- Mammography. Guidance Chapter," 2003.
- [3] L. Tabár, "Viewing technique," Chapter C, Multimodality Approach to Detection and Diagnosis of Breast Diseases, 2014.
- [4] R. C. González and R. E. Woods, "Digital image processing," Ed. Pearson. 2011.
- [5] J. Tang, R. Rangayyan, J. Xu, I. El Naqa, and Y. Yang "Computer – Aided Detection and Diagnosis of Breast Cancer with Mammography: Recent Advances," IEEE Transactions on Information Technology in Biomedicine. Vol. 13, No. 2. IEEE, 2009.
- [6] K. Ganesan, U. Acharya, C. Chua, L. Min, K. Abraham, and Ng Kwan- H. "Computer –Aided Breast Cancer Detection Using Mammograms: a review," IEEE Reviews in Biomedical Engineering, Vol., 6. IEEE, 2013.
- [7] S. M. A. Beheshti, H. AhmadiNoubari, E. Fatemizadeh, and M. Khalili, "An efficient fractal method for detection and diagnosis of breast masses in mammograms," Society for imaging informatics in medicine, 2014.
- [8] L. Liu, Q. Liu, and W. Lu, "Pectoral muscle detection in mammograms using local statistical features," Society for imaging informatics in medicine, 2014.
- [9] D. Pereira, R. Pereira, and Do Nascimento M. Zanchetta, "Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm," Computer Methods and Programs in Biomedicine 114: 88-101. Elsevier, 2014.
- [10] R. Nandi, A. Nandi, R. Rangayyan, y D. Scutt, "Classifications of breast masses in mammograms using genetic programming and feature selection," International Federation for Medical and biological engineering, UK, 2006.
- [11] E. Pietka, J. Kawa, Badura P., and D. Spinczyk, "Open architecture computeraided diagnosis system," Expert Systems, Blackwell Publishing Ltd., Vol. 27, No. 1: pp. 17-39, 2010.
- [12] C. García, "Revisión del sistema BIRADS en los informes mamográficos," Universidad de La Plata, Facultad de Ciencias Médicas, Brazil, 2010.



VI CONGRESO NACIONAL DE TECNOLOGÍA APLICADA A CIENCIAS DE LA SALUD

4, 5 y 6 de junio de 2015

“Generación de Nuevas Técnicas
de Diagnóstico y Tratamiento”

- [13] N. Gedik y A. Atasoy “A computer-aided diagnosis system for breast cancer detection by using a curvelet transform,” Turkish Journal of Electrical Engineering & Computer Sciences. 21: pp. 1002-1014, 2013.
- [14] E. Alberdi, A. Povyakalo, L. Strigini, and P. Ayton, “Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography,” Academic Radiology, 11(8): 909-918, 2004.
- [15] P. Taylor and R. Given, “Evaluation of computer-aided detection (CAD) devices,” The British Journal of Radiology, 78:S26-S30, 2005.
- [16] B. Kurt, V. Nabiyev, and K. Turhan, “A novel automation suspicious mass regions identification using Havrda & Charvat entropy and Otsu’s N thresholding,” Computer methods and programs in biomedicine II4:349-360, 2014.
- [17] S. Beheshti, H. AhmadiNoubari, E. Fatermizadeh, and M. Khalili, “An efficient fractal method for detection and diagnosis of breast masses in mammograms,” J. digit. Imaging, 2014.
- [18] P. Shanmugavadivu and V. Sivakumar, “A novel technique for mammogram mass segmentation using fractal adaptive thresholding,” The 8th. International Conference on Robotic, Vision, Signal Processing & Power applications, 2014.
- [19] P. Viola and M. Jones, “Rapid object detection using aboosted cascade of simple features,” In Computer Vision and Pattern Recognition, Proceedings of the 2001 IEEE Computer Society Conference. 1: 1–511. IEEE, 2001.
- [20] B. Kurt, K. Turhan, and V. Nabíyev, “Medical images enhancement by using anisotropic filter and CLAHE,” IEEE, 2012.
- [21] N. Otsu, “A threshold selection method from gray-level histogram,” IEEE Transactions on systems, man, and cybernetics, vol. SMC-9, No. 1, 1979.
- [22] K. Marias, C. Behrenbruch, R. Highnam, S. Parbhoo, A. Seifalian, and M. Brady, “A mammographic image analysis method to detect and measure changes in breast density,” European Journal of Radiology. 52(2004) 276-282. Elsevier Ireland Ltd.
- [23] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, “Pedestrian detection using wavelet templates,” 1063-6919/97. IEEE, 1997.
- [24] Z. Wang and A. C. Bovik, “A universal image quality index,” IEEE signal processing letters, vol. 9, no. 3, 2002.
- [25] V. Georgieva, R. Kountchev, and I. Draganov, “An adaptive enhancement of X-Ray images,” Springer-Verlag Berlin Heidelberg. 2013.
- [26] R. Lienhart and J. Maydt, “An extended set of Harr-like features for rapid object detection,” IEEE, 2002. 0-7803-7622-6/02.
- [27] H. D. Cheng, X. J. Shi, R. Min, L. M. Hu, X. P. Cai, and H. N. Du, “Approaches for automated detection and classification of masses in mammograms,” Pattern recognition society. Elsevier Ltd, 2005.