

Detección de distorsión arquitectural en imágenes mamográficas en vinculación con el Instituto Estatal de Cancerología en el estado de Guerrero

Cruz David Cárdenas Clavel¹ Miriam Martínez Arroyo²,
José Antonio Montero Valverde³, José Francisco Gazga Portillo⁴

Resumen. El principal motivo por el cual el ser humano ha creado diferentes tecnologías es, facilitar y reducir costos ante una problemática. Según la organización mundial de la salud, en México, el cancer de mama es el cancer con mayor mortandad e incidencias en las mujeres. En este trabajo se muestran los resultados de una primera etapa para un sistema de detección automática de cáncer de mama. Se implementa una base de datos de imágenes de mamografías con lesiones de distorsión arquitectural proporcionados por el Instituto Estatal de Cancerología del estado de Guerrero “Dr. Arturo Beltrán Ortega”, en formato dicom, aplicando un algoritmo de umbralización y eliminación de elementos por su area centroide.

Palabras clave— umbralización, dicom, area centroide, Distorsión arquitectural

Introducción

La tecnología ha avanzado de manera exponencial los últimos años. Esto se puede observar a simple vista con los diferentes dispositivos que en la actualidad se están utilizando. El principal motivo por el cual el ser humano ha creado diferentes tecnologías es, facilitar y reducir costos ante una problemática. Según la organización mundial de la salud, en México, el cancer de mama es el cancer con mayor mortandad e incidencias en las mujeres.

Existen bases de datos internacionales como la MiniMIAS (The Mammographic Image Analysis Society) (Suckling, 2015) y la DDSM (*The Digital Database for Screening Mammography*) (Wang, 2015), que sirven para la prueba y validación de metodologías de procesamiento de mamografías a nivel mundial. Sin embargo, actualmente no hay muchos proyectos enfocados en la generación de nuevas bases de datos (Moya, 2014), que contengan registros de mamografías adquiridos con sistemas tecnológicos actualizados, además, a nivel local y nacional no existen o son escasas las bases de datos de registros de pacientes

El Instituto Estatal de Cancerología “Dr. Arturo Beltrán Ortega” (IECAN) ubicado en Acapulco, Guerrero. Es la máxima institución del estado en el tratamiento de los diferentes cancers que afectan a los pobladores de esta región.

Debido al bajo número de radiólogos (5 técnicos y 3 médicos) en el IECAN y a la gran cantidad de mastografías realizadas en diferentes localidades del estado, la detección oportuna de lesiones en la mama es afectada.

El IECAN en busca de un método auxiliar en la detección de estas lesiones, contribuye en este trabajo proporcionando imágenes de mastografías con diferentes tipos de lesiones en formato dicom, generando así una base de datos propia para la experimentación académica.

En este trabajo se implementó una base de datos de imágenes de mamografías las cuales tenían la lesión tipo distorsión arquitectural inicialmente se extrajeron de la DDSM, para posteriormente trabajar con el IECAN con un total de 200 imágenes en formato Dicom. La lesión es una lesión que por su naturaleza es difícil de detectar para muchos expertos del área debido a que se esconde fácilmente en el tejido mamario. La distorsión de la arquitectura glandular: es una alteración del tejido glandular con la producción de radiaciones finas, espiculadas (forma de estrella) que no se encuentran asociadas con una masa.

Descripción del Método

La metodología para la creación de la base de datos inicia con la solicitud del consentimiento informado de las pacientes, luego, se describe la adquisición y almacenamiento de los registros mamográficos obtenidos en cada

¹ Cruz David Cárdenas Clavel es alumno de la MSC en el Instituto Tecnológico de Acapulco (ITA) del TecNM, cdcc89@gmail.com (autor correspondiente)

² Miriam Martínez Arroyo es Profesora de la MSC en el ITA del TecNM miriamma_ds@hotmail.com

³ José Antonio Montero Valverde es Profesor de la MSC en el ITA del TecNM jamontero1@infinutmmail.com

⁴ Gazga Portillo José Francisco es Profesor de la MSC en el ITA del TecNM jfgazga@hotmail.com

estudio, en la siguiente etapa se realizó un análisis y validación de la información contenida en el reporte médico con el acompañamiento de un especialista y la última etapa corresponde al etiquetado de las imágenes acorde con la información de la etapa anterior y la confirmación del resultado de las biopsias en la metodología (Figura 1) que se empleará para la elaboración del clasificador de lesiones tipo distorsión arquitectural en imágenes de mastografías será las siguientes fases:

- Inserción de la imagen.
- Preprocesamiento de las imágenes de mastografías para la base de imágenes del modulo
- Segmentación de las imágenes
- Extracción y selección de características de las lesiones de distorsión arquitectural
- Clasificación de las regiones de interés (Maligno o benigno)



Figura 1: Metodología General

Inserción de Imágenes de Mamografía

Con un total de 200 imágenes en formato dicom de diversos pacientes, 100 con cancer de mama (50 con lesiones tipo masas, 30 con lesiones tipo microcalcificaciones y 20 con lesiones tipo distorsión arquitectural). y 100 con lesiones benignas. Estas imágenes de mastografías se almacenan en carpetas diversas de acuerdo a su diagnostico, esto quiere decir que no es necesario usar algún sistema gestor de base de datos.

Preprocesamiento

Para los métodos de pre-procesamiento se han utilizado técnicas que permitan analizar la distribución espacial de un pixel y los píxeles de su entorno, con el fin de incrementar la precisión en la y la severidad de la lesión (benigna o maligna), con lo cual disminuirían los índices de mortalidad por cáncer de mama y los episodios de estrés innecesarios en las pacientes. Esta preparación de las imágenes tras aplicar filtros para la eliminación de ruido es usada para su procesamiento.

Extracción y selección de Características

Para la extracción de características se utiliza la metodología mostrada la Figura 2.

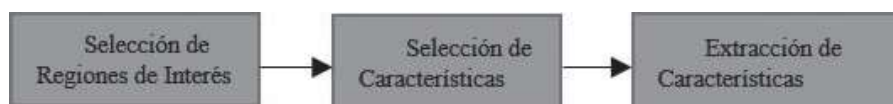


Figura 2: Metodología para extracción y selección de características

Clasificación de las regiones de interés

Una vez que la imagen se encuentra segmentada y área de interés se delimita, se clasifica el tipo de sesión. Para esto usaremos el algoritmo de Maquinas de Soporte Vectorial (MVS). La metodología para clasificación de las regiones de interés se puede observar en la Figura 3.



Figura 3: Metodología para clasificación

Para el módulo de clasificación de lesiones tipo distorsión arquitectural en imágenes de mastografías utilizaremos un separador no lineal de MVS. Hay casos donde los datos no pueden ser separados linealmente a través de un hiperplano óptimo en el espacio de entrada. En muchas situaciones, los datos, a través de una transformación no lineal del espacio de entradas, pueden ser separados linealmente, pero en un espacio de características y se pueden aplicar los mismos razonamientos que para las MVS lineal con margen máximo. La transformación de los datos de un espacio inicial a otro de mayor dimensión se logra mediante el uso de la función *kernel*. (Vapnik, 2000).

Resultados

De acuerdo a la metodología mencionada, el proyecto se encuentra terminando la tercera fase, el cual es segmentación. A continuación, se detallan los resultados parciales obtenidos en estas fases. Así como el diseño de la herramienta. La figura 4 muestra el diseño de la interfaz gráfica implementado en MATLAB en donde se desarrolla el proyecto, en la figura 4(a) se muestra la pantalla principal, la figura 4(b) muestra la selección de las imágenes.

Fase 1: Adquisición de la imagen.

En esta fase, se acordó una colaboración con el IECAN, el cual proporcionando imágenes de mamografías en formato dicom de diferentes tipos de lesiones, benignas y malignas. Para los resultados siguientes se usa una base de datos elaborado con las imágenes proporcionadas por el IECAN.

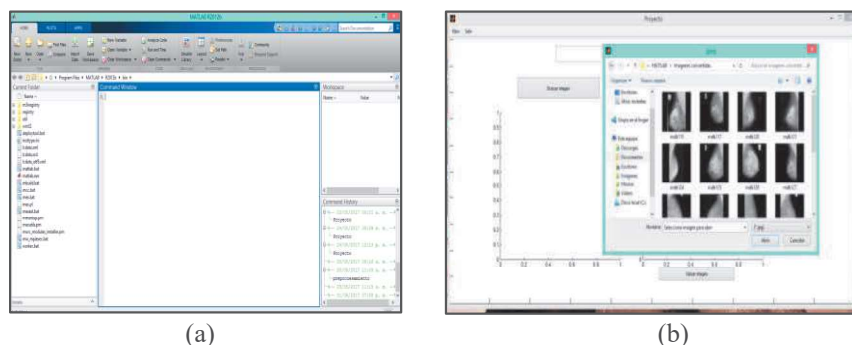
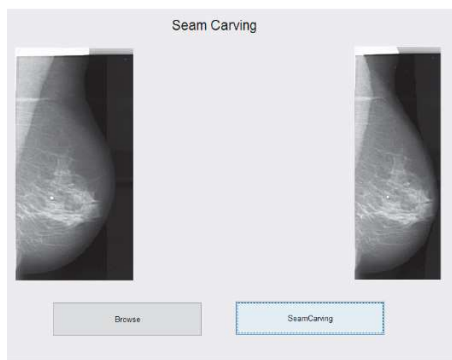


Figura 4: interfaz gráfica, 4(a) Pantalla principal, 4(b) selección de imágenes.

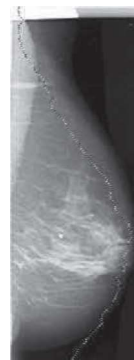
Fase 2: Preprocesamiento

Reducción del área de trabajo

En esta etapa se reduce la imagen para no procesar datos innecesarios como podría ser, nombre de la paciente, fecha o datos generales de la mamografía. Para la eliminación de los datos generales podríamos ocupar el método convencional de recortar la imagen. Esto sería un método eficaz la única problemática que habría después del corte sería la estandarización de las imágenes, ya que, en procedimientos posteriores es necesario que las imágenes tengan el mismo tamaño. Escalar la imagen no será una buena opción debido a la gran pérdida de la calidad de la misma. Para estandarizar un tamaño en las imágenes de mamografías se utiliza el algoritmo *Seam Carving*, el cual puede ampliar o reducir las imágenes de mamografías sin alterar la calidad de la misma. En la imagen 4(a) se observa la aplicación el algoritmo Seam Carving podremos ver los resultados de este procedimiento con la reducción del tamaño, donde no solamente se redujo la “parte oscura” de la imagen si no también parte de la representación de la mama como se puede observar en la línea blanco y negro que atraviesa la imagen en la figura 4(b). Esta línea representa los pixeles, *menos visibles*, eliminados. Esto ayudara en gran manera posteriormente puesto que puede reducir el costo computacional.



a) Aplicación del algoritmo Seam Carving



b) Eliminación de Pixeles

Figura 4 Algoritmo Seam Carving

Fase: Segmentación

En esta fase “se depura” información innecesaria y mostraremos solamente las lesiones de la mastografía. Esto ayuda en procesos posteriores como lo son la extracción de características. Como mencionado en la metodología se usa la técnica de eliminación mediante lumbarización y eliminación de objetos mediante el area centroide de ellos. A continuación, se mostrará la aplicación de esta técnica. El proceso comienza con la obtención de la imagen de la mamografía en su representación en escala de grises mostrada en la figura 5.

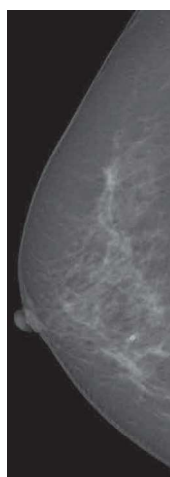


Figura 5 Imagen en Formato Dicom en escala de grises

Se analizan los niveles de umbralización como se muestra en la figura 6a y se determina que los valores de los umbrales menores a 9250 y mayores a 9800 se establezcan en 0 dando una *obscuridad* en los pixeles que se encuentren dentro de ese rango. El resto de la imagen se convierte en un formato binario el cual da por resultado la figura 6b.

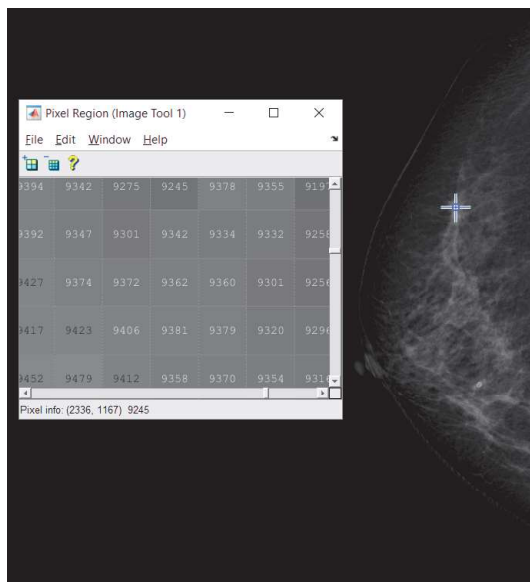


Figura 6a Eliminación de de Pixeles por Umbral

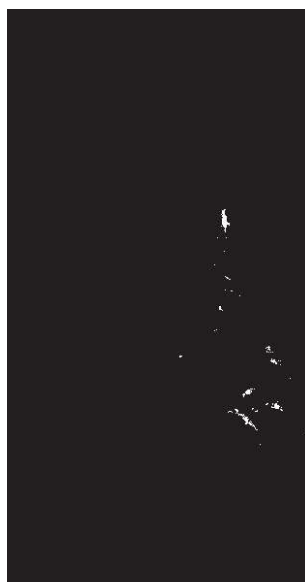


Figura 6b Imagen en Formato Binario



Figura 6c Imagen Binaria con Filtro

De la imagen resultante se aplica una matriz de uno para eliminación de algunos objetos dimituos como se puede observar en la imagen 6c. De esta imagen resultante se empieza a determinar cual es la lesión que se quiere aislar. De la imagen 6c se determina el area de cada objeto. De los objetos que tenga un area mayor a 1500 y menor que 10000 se marca con un rectangulo en color azul alrededor del area seleccionada. Se sobre pone la imagen original en la escala de grises y se obtiene como resultado la figura 7.



Figura 7 Imagen con lesión resaltada

Conclusión

En la actualidad, en México y en varias partes del mundo, el cáncer de mama es una de las enfermedades causantes de los más altos índices de muerte por cáncer, principalmente por ser diagnosticados en etapas avanzadas, es por eso que diversas investigaciones y trabajos enfocados a este padecimiento se han realizados en diferentes países del mundo, en busca de una solución para su diagnóstico temprano, haciendo uso de las tecnologías que existen y del conocimiento de expertos en esta área de la medicina con el objetivo de obtener resultados más rápidos y eficientes a la hora de emitir un diagnóstico. En México las herramientas para este fin que se han realizado son escasas debido a la complejidad del problema y la falta de aporte económico para investigación y desarrollo de este tipo de proyectos. En los países desarrollados existen sistemas de computadoras basados en técnicas de inteligencia artificial enfocados al diagnóstico médico, su aportación ha sido realmente notable, sin embargo, en México y la mayoría de los países de la región de Latinoamérica se carece de este tipo de sistemas.

En este proyecto se desarrolló la construcción de una base de datos con registros mamográficos del IECAN, donde se aplicó el preprocesamiento de las imágenes. El algoritmo *Seam Carving* fue utilizado debido a su conveniencia de aplicarse sin importar cuál sea el tamaño de la imagen, ya que este algoritmo puede modificar el tamaño de la imagen sin alterar la calidad de la misma. Esto significa que además de estandarizar el tamaño de las imágenes, podrá reducir costo computacional.

La técnica de umbralización se usó para eliminar objetos de la imagen mediante su brillo de cada pixel. Dado que esta técnica por si sola no elimina por completo los ruidos que tiene la imagen, se ocupó una eliminación por una matriz de unos el cual eliminaba pequeños objetos los cuales no representaban alguna lesión. A pesar de dicho *filtro por matriz de uno* aun así había suficientes objetos que no representan lesiones. Para eso se ocupa la eliminación por áreas centroide, esto es sacar el area de cada objeto a partir de su centro. Los objetos que tenían un area mayor a 1500 pero menor a 10000 se consideran objetos la cual pueden tener una lesión.

Finalmente, después de realizar el preprocesamiento con la técnica de Seam Carving reduciendo el tamaño de las imágenes proporcionadas por el IECAN para así poder realizar nuestra propia base de datos. Se realizó la segmentación de la lesión con la umbralización y eliminación de objetos por el area centroide, se obtuvieron datos satisfactorios.

Como trabajos futuros y recomendaciones se plantea el continuo mejoramiento de la base de datos, además del registro de la información y el desarrollo de una plataforma para el libre acceso a la comunidad científica. Además, se plantea elaborar una matriz de confusión como se muestra en la tabla 8 para medir la efectividad de la segmentación.

	P	N
Lesiones con Distorsión	VP (Señala la lesión de distorsión)	VN (No señala la lesión de distorsión)
Otras Lesiones	FP (Señala como distorsión otras lesiones masas, micro)	NN (No señala como distorsión otras lesiones, masas, distorsión, microcalcificaciones)

Tabla 8 Tabla de Confusión

Referencias bibliográficas.

- Moya, J. S. (2014). Construcción de una base de datos de imágenes.
 Shamir, S. A. (2007). Seam carving for content-aware image resizing. *ACM Trans. Graph*, 10.
 Suckling, J. (2015). Mammographic Image Analysis Society Database. *Mias*.
 Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.
 Wang, Z. (2015). A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. 6. 54.