

Artículo No. 06: Ruíz, J.; Taddei, J. y Rodríguez, R. (2014). Detección de potenciales afectaciones del seno mediante un modelo de regresión multivariada y técnicas informáticas. *Semilleros*, 1 (1), 63-76.

Detección de potenciales afectaciones del seno mediante un modelo de regresión multivariada y técnicas informáticas

Potential affectation detection via multivariate regression model and informatics techniques

Ruiz Duarte José Luis¹, Taddei Bringas Jorge Luis²,
Rodríguez Carvajal Ricardo Alberto³

Resumen. El presente documento muestra el desarrollo y aplicación de una metodología diseñada para apoyar la termografía infrarroja, como técnica de detección de cáncer de mama alternativa a la mastografía. Esta ha surgido como un complemento para la mastografía de rayos X, pero aún no es aceptado como su reemplazo. La problemática consiste en desarrollar una herramienta para analizar los termogramas obtenidos mediante cámaras térmicas especializadas. El objetivo de la presente investigación consiste en desarrollar un modelo matemático de las temperaturas que permita cuantificar el comportamiento de las temperaturas del seno y automatizarlo mediante software. Se revisó la bibliografía acerca de los métodos de detección actuales. Se plantea una metodología del tipo correlacional, la cual consiste en recopilar datos sobre termografías de personas clínicamente controladas, y desarrollar un modelo matemático, multivariado y univariado, para entender mejor la relación de los datos de los pacientes y el diagnóstico. Además, se desarrolla un software que apoya la aplicación de técnicas encontradas en la bibliografía, así como la aplicación del modelo del presente trabajo.

Palabras clave: *Cáncer de mama, Procesamiento de imágenes, Análisis multivariado.*

Abstract. This paper shows the development and application of a methodology designed to support infrared thermography as a breast cancer detection techniques different of mammography. This has emerged has a complement for X-Ray mammography, but not as its substitution. The actual problematic consists in developing a tool that analyses the thermograms obtained by thermic cameras specialized for study on humans. The objective of this research is to develop a mathematic model in order to quantify the behavior of the breast temperature, and to automatize it via software. A bibliographic review of actual detection methods and breast cancer detection via thermography was conducted. A correlational type methodology was presented, which consists of collecting thermography data of clinically controlled cases, and developing, proving and validating both univariate and multivariate mathematical models in order to better understand the relation between the cases and the diagnosis. Besides, a software that supports the application of the techniques found in the review and the application of the model was developed in this work.

Key Words: *Breast cancer, Image processing, Multivariate analysis.*

¹Ingeniería Industrial, M.I., Maestro de Asignatura, Departamento de Ingeniería Industrial, Universidad de Sonora, Hermosillo, Sonora, México. jduarte@industrial.uson.mx

²Ingeniería Industrial, Dr., Maestro de Tiempo Completo, Departamento de Ingeniería Industrial, Universidad de Sonora, Hermosillo, Sonora, México. jtaddei@industrial.uson.mx

³Ingeniería Industrial, Dr., Maestro de Tiempo Completo, Departamento de Ingeniería Industrial, Universidad de Sonora, Hermosillo, Sonora, México. ricardo@industrial.uson.mx

INTRODUCCIÓN

El cáncer de mama es una grave amenaza para la salud de la mujer a nivel mundial, siendo la segunda causa de muerte en mujeres de 30 a 54 años de edad en México (Knaul et al, 2009), y la segunda causa de muerte de mujeres en el mundo (Sree et al, 2011).

El estándar de oro para la detección del cáncer de mama es la mastografía, pero como se detallará más adelante en este artículo, cuenta con algunos inconvenientes al momento de la aplicación y de la emisión del diagnóstico. Por ello se han desarrollado alternativas a la mastografía, siendo una emergente la termografía infrarroja.

El objetivo de la presente investigación consiste en desarrollar una metodología que tenga como resultado un modelo multivariado para comprender las temperaturas obtenidas en los termogramas de una determinada población, así como un software que permita facilitar los cálculos relacionados con la aplicación de dicho modelo. El presente estudio se realiza en Hermosillo, Sonora, entre agosto de 2011 y agosto de 2013.

En un principio, se presenta el marco teórico que aborda el cáncer de mama y explica la necesidad de contar con un método de detección temprano, a la vez que sustenta las técnicas utilizadas para el desarrollo del modelo. Luego se muestra la metodología aplicada, se desglosan y describen los pasos de la misma. Después aparecen los resultados obtenidos. Finalmente, se detallan las conclusiones y recomendaciones para estudios posteriores.

Marco teórico*Cáncer de mama*

El cáncer puede ser categorizado como invasivo o no invasivo basándose en su

patrón de crecimiento (Ng y Kee, 2008). Una detección temprana permite más opciones clínicas en el tratamiento (Jiang et al, 2010), por lo cual es necesario contar con métodos que permitan detectar de una manera eficaz dicha enfermedad. La detección no siempre es correcta (Ng y Kee, 2008).

En la obra de Elmore et al (1998) se ha demostrado que el 23.8% de los casos han obtenido un falso positivo en mastografía. En la investigación de Ohuchi et al, (1995) han demostrado que la proporción de falsos negativos es de 15.4% para la exploración física y 2.8% para la exploración basada en mastografía y exploración física.

Entre las alternativas para la detección, encontramos a la mastografía y el ultrasonido, los cuales dependen principalmente en la distinción estructural y la variación anatómica del tumor (Ng et al, 2001b). Otras alternativas que se han desarrollado para la detección del cáncer son la ultrasonografía, resonancia de imagen magnética, escáner de impedancia eléctrica, tomografía de emisión de positrones, imagenología óptica y la termografía infrarroja (Jiang et al, 2010).

Según Ng y Kee (2008), algunos indicadores para la evaluación de un método de detección son la Sensitividad, o la probabilidad que una prueba sea positiva en una población no saludable, y Especificidad, o la probabilidad que la prueba sea negativa en una población saludable.

Se ha documentado que una combinación de análisis clínico, mastografía y termografía da como resultado una sensibilidad del 98% (Ng y Sudharsan, 2004).

La mastografía ha logrado reducir la mortalidad debida a cáncer de mama de

las mujeres mayores a 50 años en 25 a 30% (Köşüş et al, 2010), y es el estándar de oro utilizado más ampliamente para su detección (Ng et al, 2001b). Pero la mamografía tiene algunos aspectos negativos, entre otros, que exhibe baja sensibilidad en la presencia de tejido glandular denso (Jiang et al, 2010), un promedio de 75% de falsos positivos, un rango de 4% a 34% de falsos negativos y requiere compresión del seno y exposición a radiación (Köşüş et al, 2010), la cual incrementa el riesgo de desarrollar cáncer de mama (Harris, 2009).

El ultrasonido se usa para detectar lesiones de seno y ha mejorado la detección en conjunto con la mastografía (Sree et al, 2011).

La imagenología infrarroja ha sido utilizada en aplicaciones como oncología, dolor, artritis, entre otras (Jiang et al, 2005). Fue aprobada por la Food and Drugs Administration de EEUU en 1982 como complemento a la mastografía para detectar cáncer de mama (Köşüş et al, 2010). Es un método de imagen no invasivo económico, rápido y no causa ningún daño sobre el paciente (Ng y Kee, 2008). Puede ser usada repetidas veces sin causar daño al paciente (Ng et al, 2001b).

La visión térmica es un método inofensivo, puede diagnosticar enfermedades en etapas tempranas y no requiere preparaciones especiales para tomar una imagen (Damnjanović et al, 2010).

La termografía por sí sola no es capaz de generar diagnóstico, pero es útil para brindar más información al diagnóstico obtenido por otros métodos (Ng y Kee, 2008). Con un enfoque multimodal puede detectar hasta el 95% de los casos de cáncer temprano (Jiang et al, 2005).

Si existe un tumor maligno en el seno, utilizará más nutrientes para soportar su crecimiento. Como resultado de este incremento en el metabolismo, el lugar donde crece el tumor tendrá una tendencia de aumento de temperatura y su transferencia en todas direcciones (Ng y Kee, 2008). Un tumor pequeño pero de gran temperatura puede indicar un rápido crecimiento (Köşüş et al, 2010).

La temperatura de la piel sobre un tumor maligno es frecuentemente entre 1 y 3° más caliente que la piel alrededor (Amri et al, 2011). Asimismo, un tumor puede causar que la distribución de temperatura se altere, causando asimetría térmica entre los senos (Ng et al, 2001b). Se asume que el cuerpo humano tiene simetría térmica, por lo que las variaciones en ambos senos deben ser similares (Ng et al, 2002).

La termografía ha ayudado a los científicos a diferenciar tejido benigno y maligno (Köşüş et al, 2010).

La utilización de termografía permite detectar cáncer hasta 10 años antes que otras modalidades, evaluar la eficiencia del tratamiento y pronosticar la condición del paciente. El 44% de los pacientes con termogramas anormales pueden desarrollar cáncer dentro de los siguientes 5 años. En el 70% de los casos se puede detectar cáncer de mama un año antes que por medio de mastografía (Moghbel y Mashohor, 2011).

Un termograma es un mapa térmico de la superficie de un cuerpo que irradia calor (Ibarra-Castaneda et al, 2007). Un termograma anormal ha demostrado ser un indicador confiable de un riesgo alto de cáncer de mama en etapas tempranas (Jiang et al, 2010).

La sensibilidad de la mastografía es menor en personas jóvenes con senos densos, mientras que la termografía no se ve afectada ello (Moghbel y Mashohor, 2011).

Cada pecho tiene un patrón termográfico en particular que no cambia con el tiempo. Un cambio en el termograma de una persona puede indicar que un algo está ocurriendo (Fok et al, 2002).

Una vez que la temperatura de cada pixel es compilada, se pueden aplicar técnicas bioestadísticas para analizarlos (Ng y Kee, 2008). Las imágenes del seno pueden seccionarse en múltiples áreas para evitar errores (Fok et al, 2002).

Otros métodos de detección, mencionados en la obra de Ng y Kee (2008), son: scintimamografía de rayos gamma, para evaluar la naturaleza de las lesiones de seno (Scopinaro et al, 1999); tomografía de emisión de positrones, una técnica de imagenología funcional sobre procesos fisiológicos, bioquímicos y farmacológicos (Marques De Miranda et al, 2012); imagenología de microondas, ya que existe un contraste dieléctrico entre tejido maligno y normal (Golnabi et al, 2011); imagenología de impedancia eléctrica, que mide la habilidad de los tejidos para conducir y almacenar electricidad (Oran e Ider, 2012), ya que se ha demostrado que los tumores difieren del tejido sano circundante (Laufer et al, 2012); mediciones de potencial eléctrico, que brindan información sobre crecimiento normal o canceroso (Marino et al, 1994); imagen de resonancia magnética, que utiliza núcleos de hidrógenos y sus propiedades magnéticas para propósitos de imágenes (Sree et al, 2011); espectroscopia de resonancia magnética, que identifican ciertos componentes presentes en lesiones malignas, pero no en lesiones benignas ni en tejido normal (Bolan et al, 2005); Elastografía, dado que un tejido canceroso puede ser de 5 a 20 veces más difícil de doblar que el tejido adiposo glandular (Mcknight et al, 2002).

Procesamiento de imágenes digitales

Una imagen es una función de dos dimensiones $f(x,y)$ donde x e y son coordenadas espaciales y f en cada pareja de coordenadas (x,y) representa la intensidad del nivel de gris en ese punto (Gonzalez et al, 2004).

Los valores de la intensidad de una imagen varían dependiendo del tipo de imagen: para imágenes a blanco y negro, $f(x,y)$ es binaria; para la escala de gris, $f(x,y)$ toma valores desde 0 (negro) hasta 255 (blanco); para imágenes de colores, $f(x,y)$ es un vector con tres componentes individuales para rojo, verde y azul (Megahed et al, 2011).

El procesamiento de imágenes infrarrojas consiste en la aplicación de técnicas como mejora de la resolución, remoción el ruido y medición de la temperatura (Jiang et al, 2005).

El software Matrix Laboratory® (MATLAB®) tiene la capacidad para analizar imágenes y convertirlas en arreglos numéricos, los cuales pueden someterse a operaciones matemáticas diversas (Gonzalez et al, 2004). Los procesos sobre una imagen se pueden denotar como en la Ecuación (1) (Gonzalez et al, 2004):

$$g(x,y)=T[f(x,y)] \quad (1)$$

Donde f es la imagen de entrada, g es la imagen de salida y T es un operador en f . Las operaciones de transformación de la intensidad, se pueden dividir en: Funciones de transformación de la intensidad, que se dan cuando se obtiene una imagen nueva a partir de una imagen existente que es alterada pixel por pixel; y procesamiento de histograma, el cual nos indica el número de pixeles que tienen un determinado nivel de gris de la imagen y que puede modificarse para parecerse a una distribución de probabilidad conocida (Gonzalez et al, 2004).

La segmentación de la imagen se refiere a clasificación de píxeles en Píxeles de Fondo y Píxeles Señal (Laza Dueñas et al, 2010). Después, sigue la etapa de extracción de características, en la que se calculan las medidas de interés para las regiones segmentadas en la imagen (Laza Dueñas et al, 2010).

Análisis de los datos

El objetivo de la inferencia estadística es obtener conclusiones de una población basándonos en una muestra de la misma (Montgomery, 2005).

En muchos problemas de la ingeniería se requiere que tomemos una decisión sobre aceptar o no alguna hipótesis de algún parámetro en una población de datos (Montgomery y Runger, 2003). Este procedimiento es conocido como prueba de hipótesis y es muy útil en el caso de experimentos comparativos.

Un caso especial de las pruebas de hipótesis ocurre cuando no se conoce la distribución subyacente de la población de estudio y queremos probar la hipótesis que una distribución es un modelo adecuado para una población. Para este caso, se utiliza un estadístico de prueba conocido como Chi-Cuadrada (χ^2) (Montgomery y Runger, 2003).

Para describir algunas situaciones reales se requiere considerar simultáneamente varias variables. El análisis de datos multivariantes estudia p variables medidas en n elementos de una población. (Peña, 2002). Esta información se presenta generalmente en una matriz de datos $\mathbf{X}$ de dimensiones $n \times p$. El vector de medias es de dimensión p , y contiene las medias de cada variable (Peña, 2002). La matriz de varianzas y covarianzas $\mathbf{S}$ contiene la información sobre la varianza de cada variable y la relación lineal entre dos variables.

La matriz de precisión $\mathbf{S}^{-1}$ es la inversa de la matriz de varianzas y covarianzas y

contiene la información sobre la relación multivariante entre cada una de las variables y el resto. Si se llaman s^{ij} a los elementos de la matriz de precisión, tenemos la Ecuación (2).

$$\hat{\beta}_{ij} = -s^{ij}s_r^2(i) \quad (2)$$

Donde $\hat{\beta}_{ij}$ representa el coeficiente de regresión de la variable i explicado por la variable j y $s_r^2(i) = 1/s^{ii}$ es la varianza residual (Peña, 2002).

Un vector $\mathbf{X}$ sigue una distribución normal p -dimensional si su función de densidad es aquella que aparece en la Ecuación (3) (Peña, 2002).

$$f(\mathbf{X}) = |\mathbf{V}|^{-1/2} (2\pi)^{-p/2} \exp \{-(1/2)(\mathbf{X}-\boldsymbol{\mu})' \mathbf{V}^{-1}(\mathbf{X}-\boldsymbol{\mu})\} \quad (3)$$

Algunas propiedades son: La distribución es simétrica alrededor de $\boldsymbol{\mu}$; la media del vector aleatorio normal es $\boldsymbol{\mu}$ y su matriz de varianzas y covarianzas es $\mathbf{V}$; si se cortara con hiperplanos paralelos al definido por las variables, se obtienen curvas de nivel definidas por la Ecuación (4).

$$D^2 = (\mathbf{x}-\boldsymbol{\mu})' \mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \quad (4)$$

La variable D^2 , conocida como distancia de Mahalanobis, se distribuye como una χ^2 con p grados de libertad.

El muestreo es un conjunto de técnicas estadísticas que estudian la forma de seleccionar una muestra representativa de una población que permita inferir sobre los parámetros de la misma (Pérez López, 2005). En el caso de las distribuciones multivariadas se recomienda utilizar un tamaño de muestra tal que, en combinación con la correlación entre las variables existentes y el nivel de significancia α , logre una potencia por lo menos del 80%. Con un tamaño de muestra 100 o superior se puede lograr una potencia aceptable para una correlación de 0.5 o superior (Hair et al, 2009).

Estudio de casos previos

En el estudio de Ng et al (2001b) se realizó un análisis estadístico de las temperaturas de ambos senos, analizando imágenes de 82 pacientes: 30 pacientes asintomáticos, 48 pacientes con una anomalía benigna en la mama y 4 pacientes con cáncer. Afirman que las diferencias entre el seno derecho y el izquierdo entre personas sanas, con tumores benignos o malignos, es insignificante.

En el trabajo de Ng et al (2002), se muestra la utilización de redes neuronales artificiales. Se obtiene una red donde sus resultados son correctos en un 61.54% de las ocasiones.

En la obra de Tan et al (2009), se logró la clasificación de diferentes diagnósticos (normal, benigno, cáncer o sospechoso) con más del 80% de precisión mediante la aplicación de redes neuronales.

En el trabajo de Moghbel y Mashohor (2011) se hace una revisión de la aplicación de software de asistencia para el diagnóstico basado en redes neuronales y lógica difusa, encontrando que la más grande mejora se encuentra con una combinación de ambas técnicas, obteniendo más de un 90% de certeza global en el diagnóstico.

Además, en el estudio de Ng y Kee (2008) aplican redes neuronales artificiales y métodos bio-estadísticos. Esta técnica obtuvo un 80.95% de certeza global, con 100% de sensibilidad y 70.6% de especificidad.

Por su parte, Köşüş et al (2010) hacen una comparación entre la mamografía convencional, mamografía digital y la imagen térmica digital. Su estudio indica que la mamografía de filme de pantalla

tiene un rango entre 4 y 34% de falsos negativos, y de los casos considerados como positivos, el 75% fueron descartados por la biopsia.

En la obra de Fok et al (2002) se utilizó una combinación de termogramas y redes neuronales artificiales en las cuales aplicaron técnicas de procesamiento de imágenes antes de ingresarlas a la red, y una segunda neurona en la cual se utilizaban datos fisiológicos en lugar de termogramas. La red basada en termogramas tienen buena capacidad para predecir casos con cáncer, pero mala para identificar casos sin cáncer. Ocurrió lo contrario con la otra red.

Enfocándose a la parte de procesamiento de imágenes, Ng et al (2001a) se han desarrollado varios algoritmos que permiten minimizar los resultados no reales en los termogramas.

En la investigación de Amri et al (2011) se enfocó a investigar el estado térmico de un modelo de seno regular con tumor incrustado de 3 dimensiones utilizando la matriz de líneas de transmisión. Encontraron que la presencia de un tumor afecta la variación de la temperatura en la superficie del seno.

En la obra de Kapoor et al (2012) se menciona que el análisis de los termogramas requiere inspección cuidadosa de diferencias de temperatura pequeñas y patrones vasculares anormales. Además, presenta un enfoque computacional para la detección de tumores basados en imágenes térmicas, utilizando un método automatizado para identificar la RI, desarrollando una metodología al respecto. Basado en eso, hace una distinción entre un paciente “normal” y uno “anormal”.

MÉTODOS

Se aplicó la metodología expuesta en la Figura 1, la cual combina técnicas de estadística multivariada, univariada y procesamiento de imágenes.

En síntesis, la metodología consistió en obtener una muestra de imágenes térmicas, a las cuales se les extrajeron sus propiedades de interés. Con dichas propiedades, se desarrolló un modelo matemático multivariado y uno univariado, con la finalidad de explicar la relación entre la temperatura del paciente y las conclusiones del oncólogo experto. Además se desarrolló un software y una interfaz gráfica de usuario (GUI, por sus siglas en inglés) que permitiera facilitar la aplicación de las herramientas de procesamiento de imágenes necesarias para hacer el análisis y la aplicación de la información de los modelos. Cada aspecto de la metodología citada se explica a continuación.

Obtención de las imágenes térmicas

Desde 2009, el oncólogo Enrique Martín del Campo, ha trabajado durante años en el estudio del cáncer de mama obteniendo imágenes térmicas de a las cuales se les ha determinado un diagnóstico, basado en los datos térmicos y el historial clínico y su experiencia como oncólogo, señalando la región de interés para cada caso.

El oncólogo facilitó el uso de una base de datos de un total de 149 imágenes térmicas, acompañadas por su diagnóstico, de las cuales se seleccionaron aleatoriamente 130 para el análisis de los datos y las 19 restantes para su validación.

Definir las propiedades de interés

Para cada imagen se obtuvo la temperatura máxima en todo el pecho, a partir de la imagen, y mediante un código

que convertía la imagen en binario, se obtuvo la temperatura umbral a la cual se detectaba la región señalada por el oncólogo en su diagnóstico. Además, esta región debía cumplir con las características de ser un único máximo local y encontrarse únicamente en un seno, es decir, causar asimetría térmica.

Desarrollo de algoritmo para análisis

La fase de desarrollo del algoritmo, se dividió en tres etapas: algoritmo de pre procesamiento, de detección de bordes y de detección de región de interés. La fase de pre procesamiento busca facilitar la fase de detección de los bordes en el caso de imágenes en que la diferencia no sea lo suficientemente clara para ser detectada directamente por la computadora.

El algoritmo de pre procesamiento de imágenes cuenta con funciones de transformación de la intensidad y ecualización del histograma. El software MATLAB® cuenta con funciones específicas para estas operaciones. Para el caso de la transformación de la intensidad, se utiliza la función *imadjust*, la cual requiere la imagen que será ajustada, el rango de entrada y el rango de salida deseado de dicha imagen. En la GUI, esto se llevó a cabo mediante el uso de sliders que ajustaran dichos rangos. Para la ecualización del histograma, se utilizó la función *histeq*, y se adaptó para que el histograma de las imágenes se ajustara a las distribuciones de probabilidad uniforme, normal y exponencial.

Para la detección de los bordes se utilizó la función *edge*, adecuando el algoritmo para funcionar correctamente con los detectores *Sobel*, *Prewitt*, *Roberts*, *Laplaciano de Gaussiano* y *Canny*.

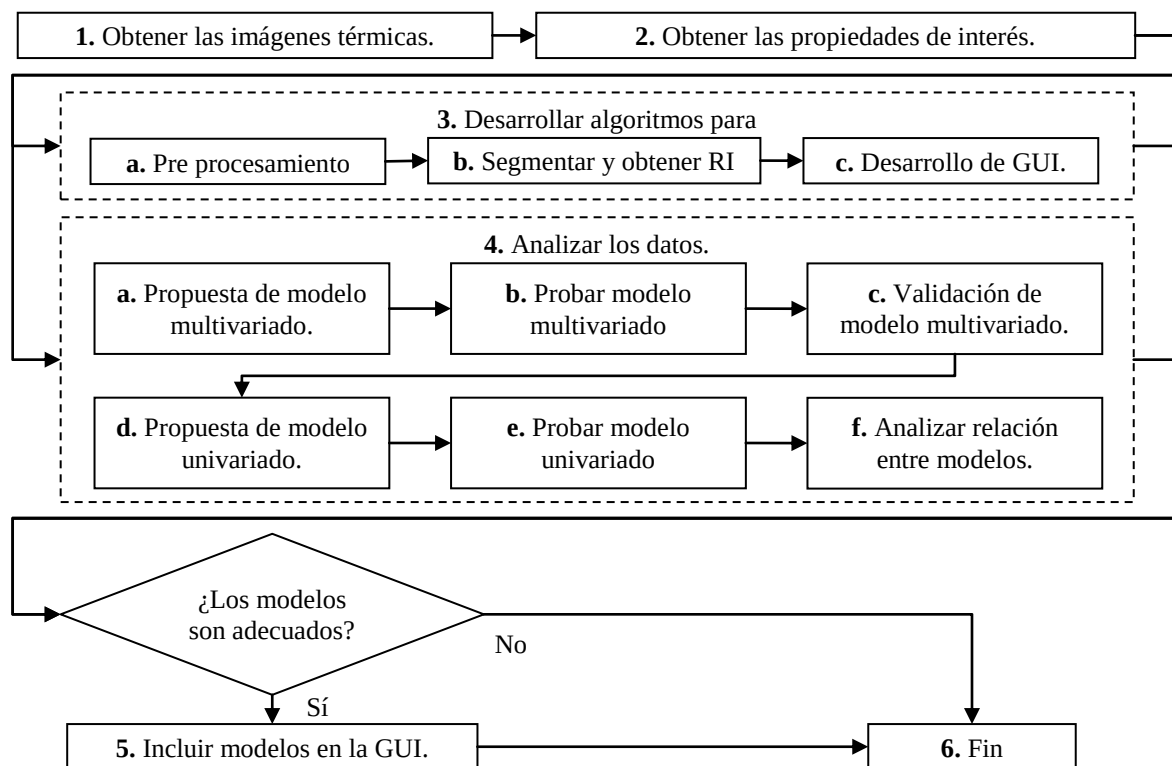


Figura 1. Metodología empleada

Para la detección de la región de interés se busca trazar un círculo cuyo centro sea el pezón del paciente (ubicado mediante la detección de los bordes) y el radio del círculo sea ubicado en la parte final del borde del seno, cerca del centro del pecho. Luego de ello se calculan las posiciones de cada punto del círculo, mediante la Ecuación (5).

$$Y = Y_0 \pm \sqrt{r^2 - (X - X_0)^2} \quad (5)$$

Donde r corresponde al radio del círculo especificado por las dos coordenadas mencionadas anteriormente y X_0 e Y_0 corresponden al centro del círculo.

Además, se utilizó la función *roypoly* para seccionar manualmente todas las áreas dentro del círculo que no nos sean de interés. La fase de detección de región de interés tiene vital importancia debido a

que en la región inferior del seno, así como en otras partes del pecho del paciente, se registran temperaturas altas que pueden afectar la información relativa a la temperatura máxima del paciente, por lo cual es necesario excluirlas del análisis. Es necesario entonces que todo pixel que se encuentre fuera de nuestra región de interés tenga un valor térmico de 0, y solo los que se encuentren dentro de la región conserven su valor original. Al final, el algoritmo permite obtener un archivo con las temperaturas de cada pixel dentro de la región de interés, a los cuales se les puede obtener parámetros bio estadísticos, como la temperatura máxima o la temperatura umbral. En resumen, el algoritmo para el análisis de las imágenes se puede expresar mediante la Figura 2.

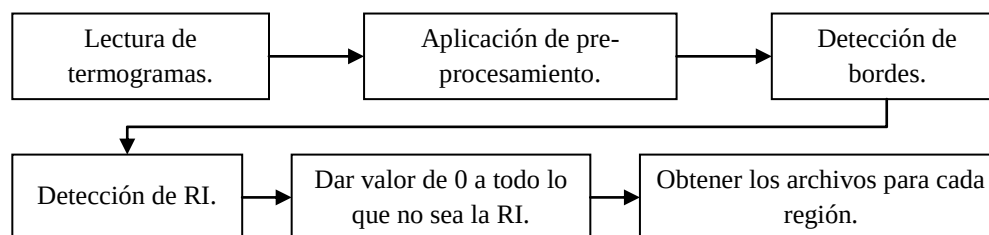


Figura 2. Algoritmo de análisis de imágenes

Analizar los datos

Al obtener las propiedades mencionadas con anterioridad de la muestra de 130 imágenes, se obtuvo que el promedio de las temperaturas máximas correspondió a 35.7446 y el de las temperaturas umbrales a 34.9046, es decir se obtuvo el vector de medias μ :

$$\mu = \begin{bmatrix} 35.7446 \\ 34.9046 \end{bmatrix}$$

La matriz de varianzas y covarianzas S , así como la matriz de precisión S^{-1} son:

$$S = \begin{bmatrix} 1.2233 & 1.0413 \\ 1.0413 & 1.0541 \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} 5.1370 & -5.0746 \\ -5.0746 & 5.9615 \end{bmatrix}$$

Los datos muestran una correlación del 91.70%, por lo que el tamaño de muestra de 130 fue más que suficiente para obtener la potencia deseada.

Por la dispersión de los datos, se propuso que estos seguían una distribución normal

bivariada. Tal como se muestra en la Figura 3, la mayoría de los datos caen dentro de la curva especificada como una χ^2 con 2 grados de libertad y una significancia del 5%, razón por la que se considera que esta distribución es apropiada para los datos.

Dado lo anterior, se obtuvo, a partir de la matriz de precisión y el vector de medias, un modelo de regresión para una distribución normal bivariada, tal como se muestra en las Ecuaciones 6, 7 y 8.

$$y - \mu_y = \beta_{yx}(x - \mu_x) \quad (6)$$

$$\hat{\beta}_{yx} = -(-5.0745) \times (1/5.9615) = 0.8512 \quad (7)$$

$$y = 34.9046 + 0.8512(x - 35.7446) \quad (8)$$

Dicho modelo nos permite pronosticar una temperatura umbral (y) para una temperatura máxima dada (x) para casos que no hayan sido previamente analizados.

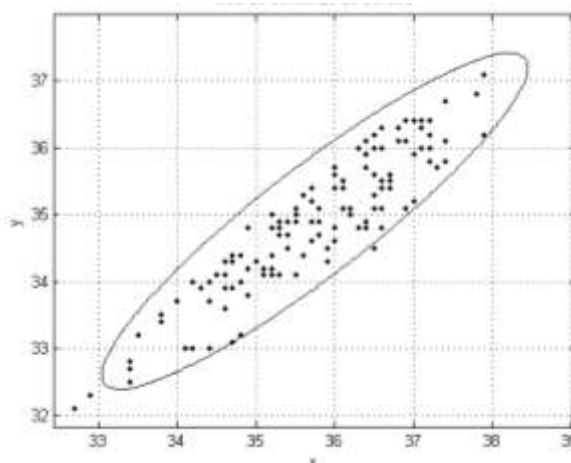


Figura 3. Prueba de χ^2 para la distribución normal bivariada

Además, se propuso que la diferencia de los datos, mostrada en la Figura 4, podría ajustarse a alguna distribución de probabilidad teórica conocida. Se aplicó

la prueba de bondad de ajuste para las distribuciones Normal, Exponencial, Lognormal y Gamma.

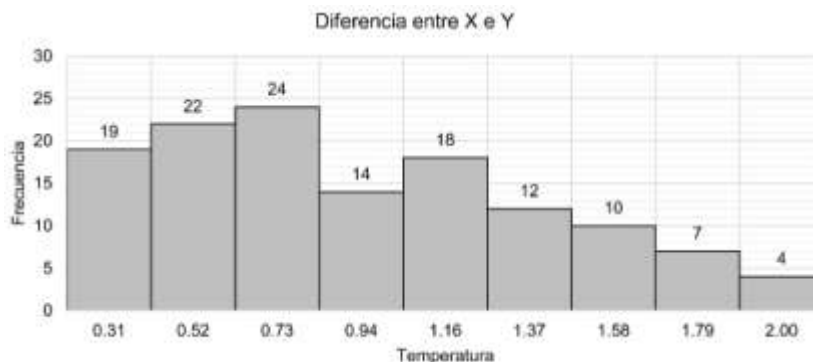


Figura 4. Diferencia de los datos de temperatura máxima y umbral

RESULTADOS

Se obtuvo un modelo basado en una distribución normal bivariada para predecir la temperatura umbral a la cual se detectan anomalías en el seno a partir de la temperatura máxima, mediante una ecuación de regresión.

Al graficar el modelo de regresión multivariada, junto con los datos de los 130 pacientes, obtenemos una línea que divide a los datos por debajo y por encima de la línea de regresión, tal como se muestra en la Figura 5. Además, se tiene que si la temperatura umbral es más parecida a la temperatura máxima, se sabe que la diferencia causada por el tumor es pequeña, por lo que se puede pensar que aquellos datos encontrados por encima de esa línea de regresión (y cercana a la línea

punteada “ideal” de la figura, la cual corresponde a los casos en que la temperatura umbral es igual a la temperatura máxima) pueden tener menor probabilidad de que la anomalía térmica se pueda relacionar con alguna afectación fisiológica.

Asimismo, se obtuvo la evidencia necesaria para afirmar que la diferencia entre ambas temperaturas no se ajusta a las distribuciones de probabilidad teóricas previamente mencionadas.

Los resultados, expresados en la Tabla 1, demuestran que las distribuciones Normal, Exponencial, Lognormal y Gamma no son adecuadas para la diferencia de las temperaturas, para una significancia del 5%.

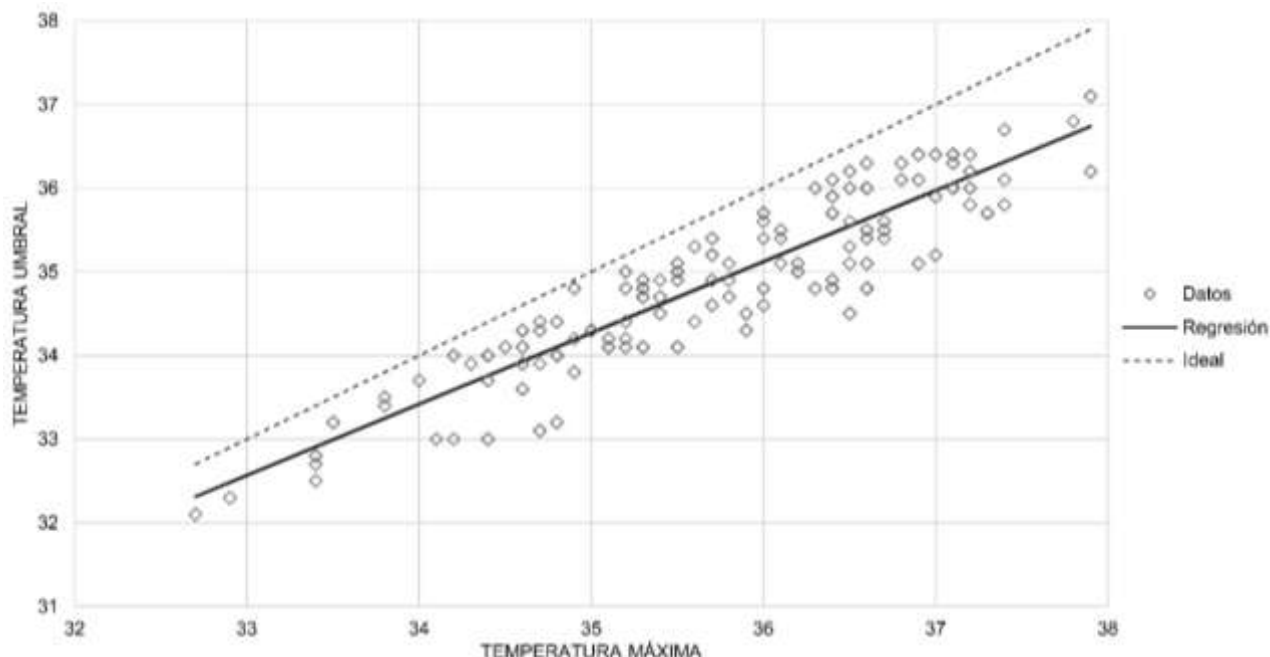


Figura 5. Línea de regresión multivariada.

Tabla 1. Prueba de bondad de ajuste para la diferencia de las temperaturas.

Distribución	Valor P	Decisión
Normal	0.0161	Rechazar Hipótesis.
Exponencial	0.00005	Rechazar Hipótesis.
Lognormal	0.0092	Rechazar Hipótesis.
Gamma	0.0084	Rechazar Hipótesis.

Se desarrolló un software que permite realizar operaciones de procesamiento de imágenes para detectar la región de interés y obtener los datos de dicha región. Además, se incluyó un módulo especial al software para incluir el modelo de regresión bivariado y obtener la temperatura máxima y la temperatura umbral pronosticada por el modelo para

cada paciente de manera automática, una vez encontrada la región de interés.

En contraste con los estudios previos citados, en esta investigación no se obtuvo el diagnóstico de los pacientes, sino el desarrollo de una base para una futura herramienta. Por lo tanto, este modelo aún no es capaz de mostrar niveles de sensibilidad ni especificidad.

CONCLUSIONES

El presente estudio nos permite conocer la relación que existe entre el diagnóstico otorgado por el oncólogo y el comportamiento de las temperaturas del paciente. Existe una relación entre la temperatura máxima y la temperatura a la cual se detecta un tumor, sin detectar el resto del tejido circundante

Se define, para nuestro estudio, una anomalía térmica como aquel pico de temperatura que se encuentre en áreas específicas de un seno bajo estudio y que puedan ser detectadas a partir de cierto umbral de temperatura, pero que dicho umbral no detecte actividad en el seno contrario; es decir, que exista asimetría térmica.

Un modelo de regresión multivariado de temperatura que permite pronosticar las temperaturas umbrales para la detección de anomalías. Se considera que una temperatura umbral por debajo de la línea

de regresión indica un caso de mayor gravedad.

Un software para facilitar los cálculos involucrados con el modelo de regresión y nuevos umbrales, lo cual puede simplificar significativamente la aplicación de dicho modelo para quienes no estén familiarizados con el análisis multivariado.

Con el nivel de conocimiento alcanzado con este trabajo, no es posible determinar de manera contundente si una anomalía encontrada con el modelo de regresión y los algoritmos integrados en el software, corresponde específicamente a cáncer de mama. Sin embargo, el presente modelo permite encontrar una temperatura umbral a partir de una máxima, y comparar el comportamiento de un paciente nuevo con el modelo de regresión multivariado. El oncólogo puede entonces considerar los valores obtenidos por el modelo al momento de realizar el diagnóstico.

Referencias bibliográficas

1. Amri, A., Saidane, A. & Pulko, S. (2011). Thermal analysis of a three-dimensional breast model with embedded tumour using the transmission line matrix (TLM) method. *Computers in Biology and Medicine*, 41, 76-86
2. Bolan, P. J., Nelson, M. T., Yee, D. & Garwood, M. (2005). Imaging in breast cancer: Magnetic resonance spectroscopy. *Breast Cancer Research*, 7, 149-152.
3. Damnjanović, Z., Petrović, D., Pantović, R. & Smiljanić, Z. (2010). Infra Red Digital Imaging in Medicine. *International Journal of Collaborative Research on Internal Medicine & Public Health*, 2, 425-434
4. Elmore, J. G., Barton, M. B. & Moceris, M. P. (1998). Mammography: False-positive results. *Southern Medical Journal*, 91, 886—887.
5. Fok, S. C., Ng, E. Y.-K. & Tai, K. (2002). Early detection and visualization of breast tumor with Thermogram and Neural Network. *Journal of Mechanics in Medicine and Biology*, 02, 185-195
6. Golnabi, A. H., Meaney, P. M., Geimer, S. D. & Paulsen, K. D. (2011). Comparison of no-prior and soft-prior regularization in biomedical microwave imaging. *Journal of Medical Physics*, 36, 159-70
7. Gonzalez, R. C., Woods, R. E. & Eddins, S. L. (2004). Digital Image processing using MATLAB. Upper Saddle River, N.J.: Pearson Prentice Hall
8. Hair Jr., J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2009). *Multivariate Data Analysis*. Upper Saddle River, N.J.: Prentice Hall
9. Harris, J. (2009). Meta-analysis: Mammography increased breast cancer risk

for high-risk women younger than 30. *Hem/Onc Today*, 40-41

10. Ibarra-Castaneda, C., Bendada, A. & Maldague, X. (2007). Thermographic Image Processing for NDT. En: *ESTRUTURALES, A. A. D. E. N. D. Y.* (ed.) IV Conferencia Panamericana de Ensayos No Destructivos. Buenos Aires, Argentina

11. Jiang, L. J., Ng, E. Y.-K., Yeo, A. C. B., Wu, S., F., P., Yau, W. Y., Chen, J. H. & Yang, Y. (2005). A perspective on medical infrared imaging. *Journal of Medical Engineering & Technology*, 29, 257-267

12. Jiang, L., Zhan, W. & Loew, M. H. (2010). Modeling static and dynamic thermography of the human breast under elastic deformation. *IOPscience*, 187-202

13. Kapoor, P., Prasad, S. V. A. V. & Patni, S. (2012). Image Segmentation and Asymmetry Analysis of Breast Thermograms for Tumor Detection. *International Journal of Computer Applications*, 50

14. Knaul, F. M., Nigenda, G., Lozano, R., Arreola-Ornelas, H., Langer, A. & Frenk, J. (2009). Cáncer de mama en México: una prioridad apremiante. *Scielo*. [Online], 51. Available:

<http://www.scielo.org.mx/pdf/spm/v51s2/v51s2a26.pdf>

15. Köşüş, N., Köşüş, A., Duran, M., Simavlı, S. & Turhan, N. (2010). Comparison of standard mammography with digital mammography and digital infrared thermal imaging for breast cancer screening. *Journal of the Turkish-German Gynecological Associatio*, 11, 152-157

16. Laufer, S., Solomon, S. B. & Rubinsky, B. (2012). Tissue characterization using electrical impedance spectroscopy data: a linear algebra approach. *Physiological Measurement*, 33, 997-1013

17. Laza Dueñas, C. M., Rego Días, A. & Carlos Pías, N. (2010). Aplicación de algoritmos para el procesamiento de imágenes a la tecnología de microarreglos para el diagnóstico. *Revista Investigación Operacional*, 31, 268-274

18. Marino, A. A., Morris, D. M., Schwalke, M. A., Iliev, I. G. & Rogers, S. (1994). Electrical potential measurements in human

breast cancer and benign lesions. *Tumour biology: the journal of the International Society of Oncodevelopmental Biology and Medicine*, 15, 147-152

19. Marques De Miranda, D., Mamede, M., Rezende De Souza, B., Guimarães De Almeida Barros, A., Magno, L. A., Alvim-Soares Jr., A., Valadão Rosa, D., José De Castro Jr., C., Malloy-Diniz, L., Gomez, M. V., De Marco, L. A., Correa, H. & Romano-Silva, M. A. (2012). Molecular medicine: a path towards a personalized medicine. *Revista Brasileira de Psiquiatria*, 34, 82-91

20. Mcknight, A. L., Kugel, J. L., Manduca, A., Hartmann, L. C. & Ehman, R. L. (2002). MR elastography of breast cancer: preliminary results. *American Journal of Roentgenology*, 178, 1411-1417

21. Megahed, F. M., Woodall, W. H. & Camelio, J. A. (2011). A Review and Perspective on Control Charting with Image Data. *Journal of Quality Technology*, 43, 83-98.

22. Moghbel, M. & Mashohor, S. (2011). A review of computer assisted detection/diagnosis (CAD) in breast thermography for breast cancer detection. *Artificial Intelligence Review*

23. Montgomery, D. C. & Runger, G. C. (2003). *Applied Statistics and Probability for Engineers*. New York: John Wiley & Sons, Inc

24. Montgomery, D. C. (2005). *Introduction to Statistical Quality Control*. New York: Wiley.

25. Ng, E. Y.-K. & Kee, E. C. (2008). Advanced integrated technique in breast cancer thermography. *Journal of Medical Engineering & Technology*, 32, 103-114

26. Ng, E. Y.-K. & Sudharsan, N. (2004). Computer simulation in conjunction with medical thermography as an adjunct tool for early detection of breast cancer. *BMC Cancer* 2004 [Online], 4. Available: <http://www.biomedcentral.com/1471-2407/4/17> [Accesado el 20 de sep de 2012]

27. Ng, E. Y.-K., Chen, Y. & Ung, L. N. (2001a). Computerized breast thermography: study of image segmentation and temperature cyclic variations. *Journal of Medical Engineering & Technology*, 25, 12-16

28. Ng, E. Y.-K., Fok, S. C., Peh, Y. C., Ng, F. C. & Sim, L. S. J. (2002). Computerized detection of breast cancer with artificial intelligence and thermograms. *Journal of Medical Engineering & Technology*, 26, 152-157
29. Ng, E. Y.-K., Ung, L. N., Ng, F. C. & Sim, L. S. J. (2001b). Statistical analysis of healthy and malignant breast thermography. *Journal of Medical Engineering & Technology*, 25, 253-263
30. Ohuchi, N., Yoshida, K., Kimura, M., Ouchi, A., Shiiba, K., Ohnuki, K., Fukao, A., Abe, R., Matsuno, S. & Mori, S. (1995). Comparison of false-negative rates among breast-cancer screening modalities with or without mammography - Miyagi trial. *Japanese Journal of Cancer Research*, 86, 501-506
31. Oran, O. F. & Ider, Y. Z. (2012). Magnetic resonance electrical impedance tomography (MREIT) based on the solution of the convection equation using FEM with stabilization. *Physics in Medicine and Biology*, 57
32. Peña, D. (2002). *Análisis de datos multivariantes*. Madrid: McGraw-Hill/Interamericana
33. Pérez López, C. (2005). *Muestreo Estadístico. Conceptos y problemas resueltos*. Madrid: Pearson.
34. Scopinaro, F., Pani, R., De Vincentis, G., Soluri, A., Pellegrini, R. & Porfiri, L. M. (1999). High-resolution scintimammography improves the accuracy of technetium-99m methoxyisobutylisonitrile scintimammography: use of a new dedicated gamma camera. *European Journal of Nuclear Medicine*, 26, 1279-1288
35. Sree, S. V., Ng, E. Y.-K., Acharya, R. U. & Faust, O. (2011). Breast imaging: A survey. *World Journal of Clinical Oncology*, 2, 171-178
36. Tan, J. M. Y., Ng, E. Y.-K., Archarya U., R., Keith, L. G. & Holmes, J. (2009). Comparative Study on the Use of Analytical Software to Identify the Different Stages of Breast Cancer Using Discrete Temperature Data. *Journal of Medical Systems*, 33, 141-153.