

**ANÁLISIS LOCAL DE TERMOGRAFÍA INFRARROJA PARA LA DETECCIÓN
DE CARCINOMA EN GLANDULA MAMARIA**

DIEGO ANDRÉS BLANCO MORA

MELANY MARCELA MORA CARVAJAL

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE INGENIERÍAS FISICO-MECÁNICAS

ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA

Y TELECOMUNICACIONES

BUCARAMANGA

2011

**ANÁLISIS LOCAL DE TERMOGRAFÍA INFRARROJA PARA LA DETECCIÓN
DE CARCINOMA EN GLANDULA MAMARIA**

DIEGO ANDRÉS BLANCO MORA

MELANY MARCELA MORA CARVAJAL

Trabajo de Grado para optar al título de INGENIERO ELECTRÓNICO

DR. ÓSCAR GUALDRÓN GONZÁLEZ

DIRECTOR TRABAJO DE GRADO

M. VÍCTOR EDUARDO MARTÍNEZ

CODIRECTOR DEL TRABAJO DE GRADO

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE INGENIERÍAS FISICO-MECÁNICAS

ESCUELAS DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA

Y TELECOMUNICACIONES

BUCARAMANGA

2011

AGRADECIMIENTOS

A Dios por haberme dado la oportunidad de vivir y cumplir mis sueños al lado de grandes personas.

A mi madre, Nelly Carvajal por su apoyo siempre incondicional hacia mí durante mi vida ya que sin ella no sería quien soy ahora y a mi padre, Henry Mora por sus consejos que permitieron guiar mi camino.

A mis hermanos, Michael y Marlon porque a pesar de sus travesuras son el motor de mi vivir.

A mi novio, Luis Ardila quien con su amor me lleno de fuerzas para nunca decaer y lleno de felicidad mi vida.

A todos mis amigos que estuvieron conmigo desde el inicio de mi carrera y a pesar de todas las situaciones y las distancias que la vida nos colocó en el camino continúan cerca a mí.

A los profesores que me brindaron sus enseñanzas a lo largo de mi carrera.

A Víctor Martínez por su compañía y consejo en el transcurso del desarrollo de este proyecto.

A Oscar Gualdrón por todas sus enseñanzas y por ser mi guía en este proyecto.

Melany Marcela Mora Carvajal

AGRADECIMIENTOS

Le agradezco principalmente al gigante que vive en los cielos, a Él que nos regala flores y un nuevo amanecer cada día, Él quien permite que cada uno de mis pasos tenga mayor firmeza uno tras otro cuando camino de la mano con Él, a Él a mi Dios que nunca me desampara.

A mis amigos y familiares, los cuales cada día me ayudaron a seguir adelante, brindándome todo su apoyo y mejores energías, cada uno de ellos aportó un grano de arena en la consecución de los objetivos que tengo propuestos hasta el día de hoy.

Un nuevo camino comienza, si va a hacer difícil o no, nadie lo sabe, de lo que si estoy seguro es que en cada día se debe dar lo mejor, no sólo por alcanzar las metas, sino porque mi Dios, mis amigos y familiares, se merecen que ponga todo mi empeño para ser una mejor persona cada día.

Diego Andrés Blanco Mora

TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN.....	14
1. Procedimiento	15
1.1 Transformada Top Hat	15
1.2 Segmentación	16
1.3 Binarización	16
1.4 Descripción	17
1.4.1 Descriptores estadísticos de primer orden	17
1.4.2 Descriptores de región simples	19
1.5 Clasificación	20
2. Resultados y Análisis	21
2.1 Selección de la región de mayor tamaño	22
2.2 Selección de descriptores	23
2.2.1 Descriptores estadísticos	23
2.2.2 Descriptores de región simples	27
2.3 Red Neuronal	27
3. Conclusiones y trabajos futuros.....	31
BIBLIOGRAFÍA.....	32

LISTA DE FIGURAS

	Pág.
Figura 1. Etapas del procesamiento de imagen realizado.....	15
Figura 2. Gráfica de la diferencia de la media entre senos vs. la diferencia de la moda....	18
Figura 3. Gráfica de la diferencia de la media vs. la diferencia de la mediana	18
Figura 4. Gráfica de la diferencia de los valores moda, media y mediana.....	19
Figura 5. Sujeto con cáncer	21
Figura 6. Sujeto sano	22
Figura 7. Resultado de una curva ROC ideal.....	23
Figura 8. Resultados de la curva ROC para los descriptores estadísticos rango, desviación, sesgo, curtosis, energía y entropía	25
Figura 9. Resultados de la curva ROC para los descriptores estadísticos media, mediana, moda y máximo.....	26
Figura 10. Resultados de la curva ROC para los descriptores de región	27
Figura 11. Sensibilidad y especificidad para cada entrenamiento de la red neuronal.....	28
Figura 12. Desempeño del entrenamiento de la red neuronal elegida	28
Figura 13. Predicción de la red neuronal.....	29
Figura 14. Predicción de la red neuronal umbralizada	30

LISTA DE TABLAS

	Pág.
Tabla 1. Parámetros seleccionados para el entrenamiento de la red neuronal	20
Tabla 2. Descriptores estadísticos para un sujeto sano	24
Tabla 3. Descriptores estadísticos para un sujeto con cáncer	24
Tabla 4. Matriz de desempeño de la red neuronal seleccionada.....	30

RESUMEN

TÍTULO

ANÁLISIS LOCAL DE TERMOGRAFÍA INFRARROJA PARA LA DETECCIÓN DE CARCINOMA EN GLÁNDULA MAMARIA

AUTORES

Diego Andrés Blanco Mora**

Melany Marcela Mora Carvajal**

PALABRAS CLAVES

Carcinoma Mamario, Descriptores Estadísticos, Descriptores de región simples, Morfología, Segmentación, Transformada Top-Hat.

DESCRIPCIÓN

En este artículo se presenta un análisis de termografía infrarroja basado en descripción local para la detección de anomalías en la glándula mamaria. Este estudio se basa en la descripción local de la región más grande y de altas temperaturas en cada seno. En principio para el desarrollo de este proceso, fueron extraídas las regiones con mayores temperaturas de cada seno utilizando morfología matemática, específicamente la transformada white top-hat y desarrollando una posterior binarización de la imagen. Con el fin de localizar distribuciones asimétricas de temperatura entre los senos se seleccionó para su descripción la zona de mayor tamaño en cada seno. Esta zona se describió mediante sus características tanto internas como externas. Esta zona fue representada utilizando los descriptores estadísticos de primer orden y también los descriptores de región simples; área, perímetro y densidad. Los descriptores más relevantes para su ingreso al sistema de inferencia fueron seleccionados por discriminación mediante el área bajo la curva ROC. Posteriormente por medio de una RNA se desarrolló un sistema de inferencia, cuyas entradas corresponden al valor absoluto de la diferencia entre los descriptores del seno derecho y el seno izquierdo respectivamente. Finalmente el desempeño del sistema es evaluado, determinando la especificidad y sensibilidad de éste.

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería Eléctrica, Electrónica y Telecomunicaciones. Director: Dr. Óscar Gualdrón González. Codirector: M. Víctor Eduardo Martínez.

SUMMARY

TITLE

Local Analysis of infrared thermograph for detection of carcinoma in mammary gland *

AUTHORS

Melany Marcela Mora Carvajal**

Diego Andrés Blanco Mora**

KEYWORDS

Mammary carcinoma, morphology, segmentation, statistical descriptors, simple region descriptors, top hat transform.

DESCRIPTION

In this article is exposed an infrared thermographical analysis based on local description for the detection of abnormalities in mammary gland. This study is based on the local description of the biggest region with highest temperatures in each breast using mathematical morphology, specifically it is used the white top-hat transform and then it was done a binarization of the image. With the aim to find asymmetric distribution of temperature between breasts, it was chosen for their description the biggest zone in each breast. This zone was described through its internal and external characteristics. This zone was represented through the first order statistical descriptors and the region descriptors: area, perimeter and density. The most relevant descriptors to ingress to the inference system were selected by discrimination through the area under the curve ROC. The inference system was made, using an ANN (Artificial Neuronal Network). The inputs of the Neuronal Network correspond to the absolute value of the descriptors difference between right and left breast. Finally the performance of the system is evaluated, finding the sensibility and sensitivity of it. This study is part of the project "Evaluation of Infrared Thermography in the Detection of Carcinoma of the Mammary Gland" code 1102-343-19202 supported by Colciencias.

* Degree Work.

** Faculty of Physic mechanical Engineering. School of electric, electronic and telecommunications. Project director: Óscar Gualdrón González. Project co-director: Víctor Eduardo Martínez.

INTRODUCCIÓN

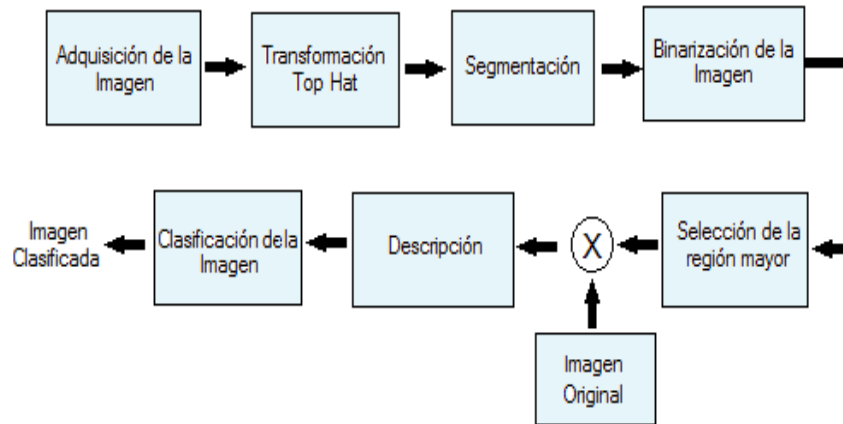
El análisis de imágenes de termografía infrarroja se ha presentado desde hace algunos años como un método que puede complementar a la mamografía en la detección del carcinoma en la glándula mamaria en fases iniciales [1]. La termografía infrarroja pasiva cuantifica la temperatura superficial y ofrece ventajas con respecto a otros métodos de detección de carcinoma. Es un método no invasivo, indoloro y se puede detectar por medio de ésta cambios en las temperaturas de los tejidos (debido a la actividad de las células cancerosas) en etapas tempranas del desarrollo del tumor [2]. Para la detección de carcinoma existen métodos como la mamografía, que sólo detectan el cáncer cuando el tumor está desarrollado y además resultan ser métodos incómodos para el sujeto.

Este artículo presenta un método para la detección de carcinoma en glándula mamaria basado en un análisis local de la termografía infrarroja de los senos y el uso de técnicas de inteligencia artificial para el diagnóstico. Las variables utilizadas para el análisis realizado corresponden al valor absoluto de la diferencia entre los descriptores de cada seno. El método desarrollado fue aplicado a las imágenes termográficas de 35 sujetos, de las cuales 5 presentaban un diagnóstico positivo de cáncer confirmado por biopsia.

Este artículo está organizado de la siguiente manera: Inicialmente se realiza una breve introducción del proyecto. La primera sección detalla cada una de las etapas del método desarrollado para el diagnóstico de cáncer. La sección dos presenta los resultados encontrados en el desarrollo del estudio. Finalmente la sección tres presenta las conclusiones más importantes de este trabajo.

1. Procedimiento

Figura 1. Etapas del procesamiento de imagen realizado.



Fuente: Autores

1.1 Transformada *white top hat*

La asimetría de la distribución de temperatura entre los senos, es indicio de posibles anomalías en la glándula mamaria. Un tumor que se esté desarrollando en la mama presenta un aumento de actividad metabólica asociado a un incremento de la vascularización local. El análisis que se desarrolló busca extraer de la imagen estas zonas con mayor vascularización utilizando morfología matemática. En particular el filtro morfológico utilizado fue la transformada White Top-Hat en niveles de gris.

La transformada u operación morfológica White Top hat se basa en la combinación de las operaciones erosión y dilatación. Su definición es la siguiente:

$$f_{Top\ hat} = f - f \circ b \quad (1)$$

Donde b es el elemento estructurante. La transformada White Top hat consiste en la sustracción de la apertura de la imagen con la imagen original. La operación apertura de una imagen consiste en la dilatación de la erosión de la imagen. En dicha operación las características brillantes de la imagen se atenúan mientras las características oscuras permanecen intactas. Así la operación White Top hat permite resaltar las características brillantes de la imagen sobre las oscuras [3].

En las imágenes procesadas las temperaturas de mayor valor corresponden a un nivel de gris brillante, es decir, que al aplicar la transformada Top-Hat a la imagen será posible segmentar las zonas de mayor temperatura. Las formas y tamaños de las regiones de mayor temperatura tienen diferentes características; por ello se seleccionó como elemento estructurante un disco, debido a que éste contiene todos los ángulos de dirección. En cuanto al tamaño del disco, luego de desarrollar algunas pruebas con diferentes radios de disco el valor seleccionado para éste fue de seis.

1.2 Segmentación

El algoritmo de segmentación utilizado para esta etapa del proceso, fue el desarrollado en [4]. Éste procedimiento consta de tres procesos: reconocimiento y acotamiento de la región, detección de bordes y aplicación de la transformada de Hough para localización de parábolas.

1.3 Binarización

Luego de realizar la transformación Top-Hat y la segmentación en la imagen se realizó una binarización de la imagen resultante. El valor umbral se determinó utilizando el método de Otsu [5]. Este método permite encontrar el umbral

deseado para realizar una correcta binarización, maximizando la varianza entre clases.

Las regiones encontradas luego de la binarización son etiquetadas y dependiendo del área de cada una de las regiones se selecciona la región de mayor área en cada seno. Finalmente, la imagen que contiene la mayor región encontrada en cada uno de los senos es utilizada como una máscara con la imagen termográfica.

1.4 Descripción

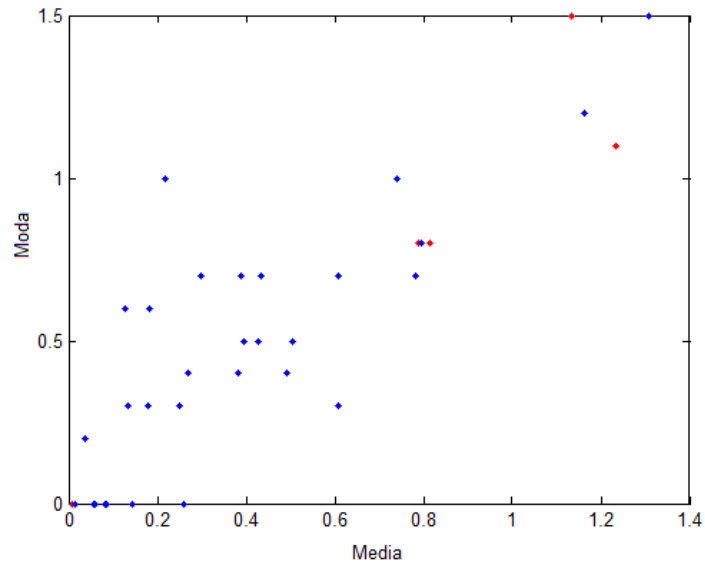
En el proceso de descripción realizado a la región extraída de cada uno de los senos se incluyeron los descriptores estadísticos de primer orden y los descriptores de región simples.

1.4.1 Descriptores Estadísticos de primer orden

Los descriptores de primer orden calculados fueron los siguientes: Media, Mediana, Moda, Desviación, Máximo, Rango, Sesgo, Curtosis, Energía y Entropía. Se seleccionaron éstos como posibles descriptores por ser utilizados en investigaciones previas sobre el tema [4] [7]. Las variables estudiadas corresponden al valor absoluto de la diferencia entre los descriptores del seno derecho y del seno izquierdo. La selección de los descriptores más importantes se basó en el análisis mediante la curva ROC. Los descriptores seleccionados fueron media, mediana, moda y máximo.

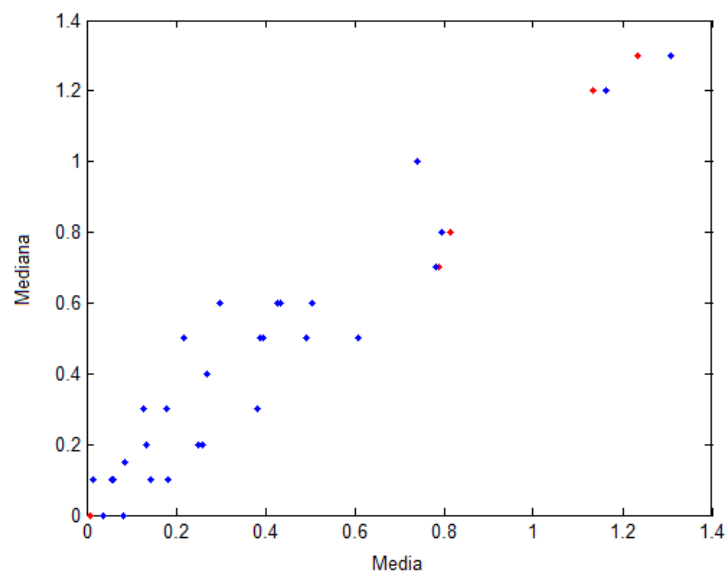
Las Fig. (2-4) representan los datos de los descriptores seleccionados calculados para cada sujeto. Los puntos rojos hacen referencia a los sujetos con cáncer y los azules a los sujetos sanos.

Figura. 2. Gráfica de la diferencia de la media entre senos vs. la diferencia de la moda.



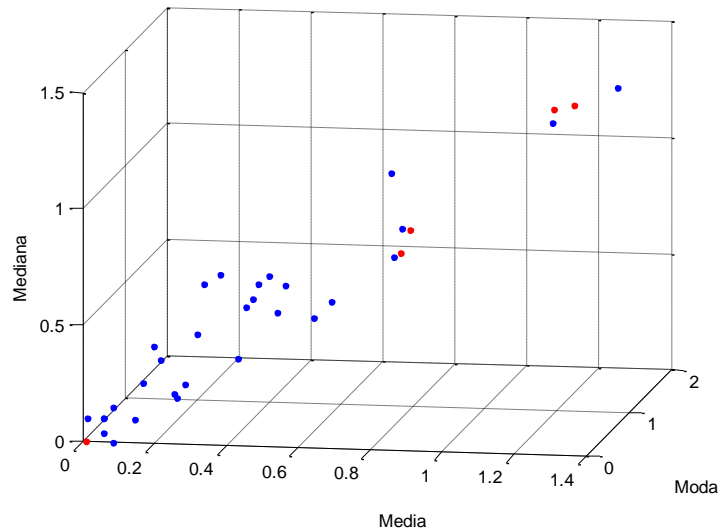
Fuente: Autores

Figura. 3. Gráfica de la diferencia de la media entre senos vs. la diferencia de la mediana.



Fuente: Autores

Figura. 4. Gráfica de la diferencia de los valores de moda, media y mediana.



Fuente: Autores

1.4.2 Descriptores de región simples.

Los descriptores de región simples (área, perímetro y densidad) fueron calculados para la región de máxima área de cada uno de los senos de las pacientes.

Del mismo modo que se realizó el análisis con los descriptores de primer orden, las variables estudiadas corresponden al valor absoluto de la diferencia entre los descriptores del seno derecho y del seno izquierdo. Sin embargo, estos descriptores al ser analizados por medio de la curva ROC no generaron buenos resultados.

1.5 Clasificación

La clasificación de sujetos se puede conseguir mediante un método de inferencia. En este método se seleccionó una red neuronal Feedforward entrenada mediante Backpropagation algoritmo Levenberg-Marquardt, con una configuración de dos capas; la primera capa consiste en cuarenta neuronas con función de activación tangente sigmoideal y la capa de salida contiene una neurona con función de activación lineal.

Debido al reducido tamaño de la población, ésta se dividió en 60% para entrenamiento, 20% para validación, y el restante 20% son independientes de la prueba de generalización.

Los parámetros seleccionados para el entrenamiento de la red neuronal se muestran en la Tabla 1.

Tabla 1. Parámetros seleccionados para el entrenamiento de la red neuronal.

Nombre	Valor	Significado
Epochs	1000	Número máximo de iteraciones
Goal	0	Límite de la función error
Min_grad	1e-10	Mínimo valor del gradiente
Mu_inc	10	Valor de incremento de los pesos
Mu_dec	0.1	Valor de decrecimiento de los pesos

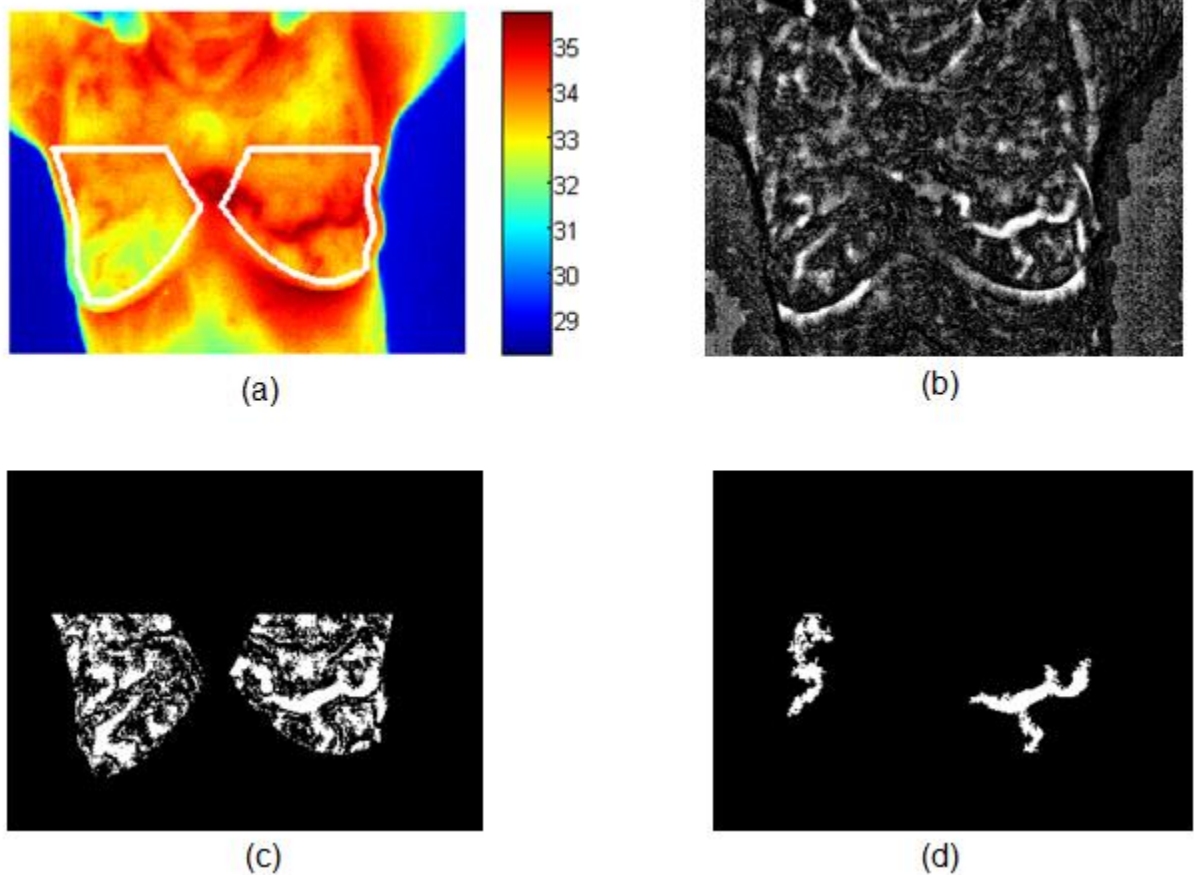
Fuente: Autores.

Para la validación del sistema de clasificación se usó la curva ROC. Esta curva permite evaluar el desempeño de las diferentes neuronales desarrolladas.

2. Resultados y análisis

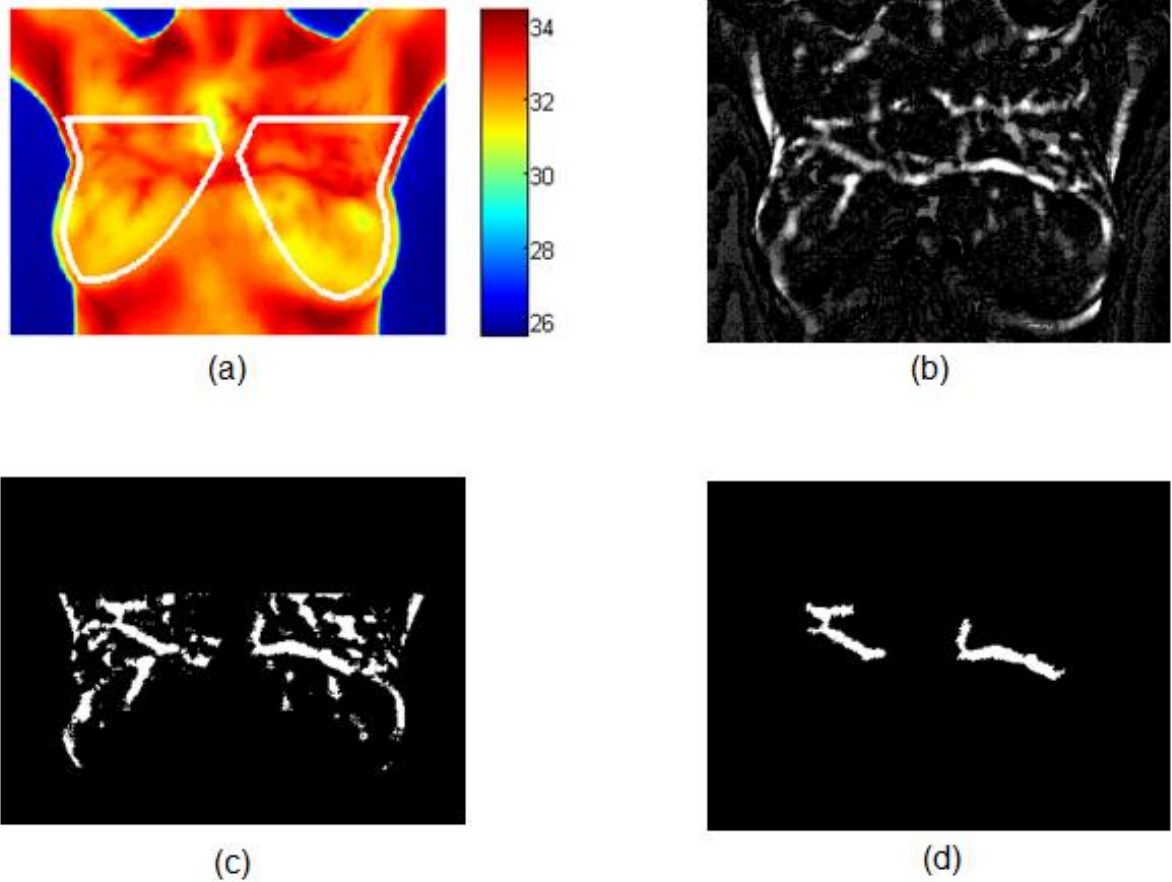
Al analizar las imágenes de una sujeto con cáncer y una sujeto saludable se puede observar la uniformidad y simetría por parte de la sujeto saludable Fig. 5, además de la diferencia de predominancia de temperaturas entre un seno y otro en las zonas de mayor área que caracterizan a la sujeto con cáncer, Fig. 6.

Figura. 5. Sujeto con cáncer a) Imagen termográfica segmentada. b) Transformada Top Hat de la imagen. c) Imagen segmentada y binarizada. d) Regiones de mayor área.



Fuente: Autores

Figura. 6. Sujeto sano. a) Imagen termográfica segmentada. b) Transformada Top Hat de la imagen. c) Imagen segmentada y binarizada. d) Regiones de mayor área.



Fuente: Autores

2.1 Selección de región de mayor tamaño

Los resultados encontrados con el método propuesto se muestran a continuación. La fig.5 muestra el proceso realizado sobre la imagen termográfica de una sujeto con cáncer y la fig.6 de una sujeto sana.

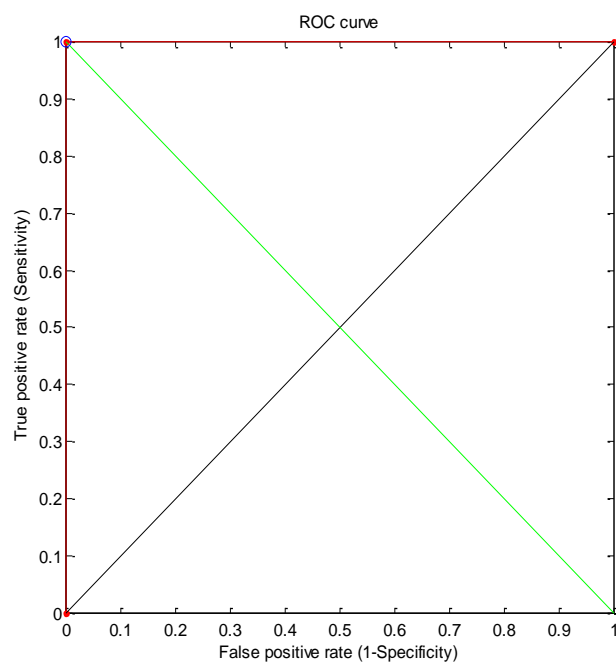
2.2 Selección de descriptores

La selección de los descriptores más importantes se realizó mediante el análisis de la curva ROC. A continuación se presentan los resultados de la curva ROC que se obtuvo de los diferentes descriptores utilizados.

2.2.1 Descriptores estadísticos

El desempeño de una técnica cuando se utiliza la curva ROC se valora mediante el parámetro AUC (Area Under the Curve) el cual en el caso ideal vale 1 y corresponde a una sensibilidad y especificidad de 100%. Idealmente la forma de la curva ROC se presenta en la Fig. 7.

Figura. 7. Resultado de una curva ROC ideal.



Fuente: Autores

En la Fig. 8 se presentan los resultados de la curva ROC de los descriptores que no fueron seleccionados. Se puede observar el valor del área bajo la curva de ellas.

Los resultados de las curvas ROC de los descriptores estadísticos seleccionados para el análisis de los registros son ilustrados en la Fig. 9.

En las Tablas 2 y 3 se pueden observar la cuantificación de estas características mediante la obtención de los descriptores estadísticas para las sujetos mostradas en la Fig. 5 y 6.

Tabla 2. Descriptores estadísticos para un sujeto sano.

Sujeto Sano	Seno derecho	Seno izquierdo	Diferencia
Media	32,82	33,21	0,39
Mediana	32.8	33.3	0.5
Moda	32.8	33.3	0.5
Máximo	33.3	33.8	0.5

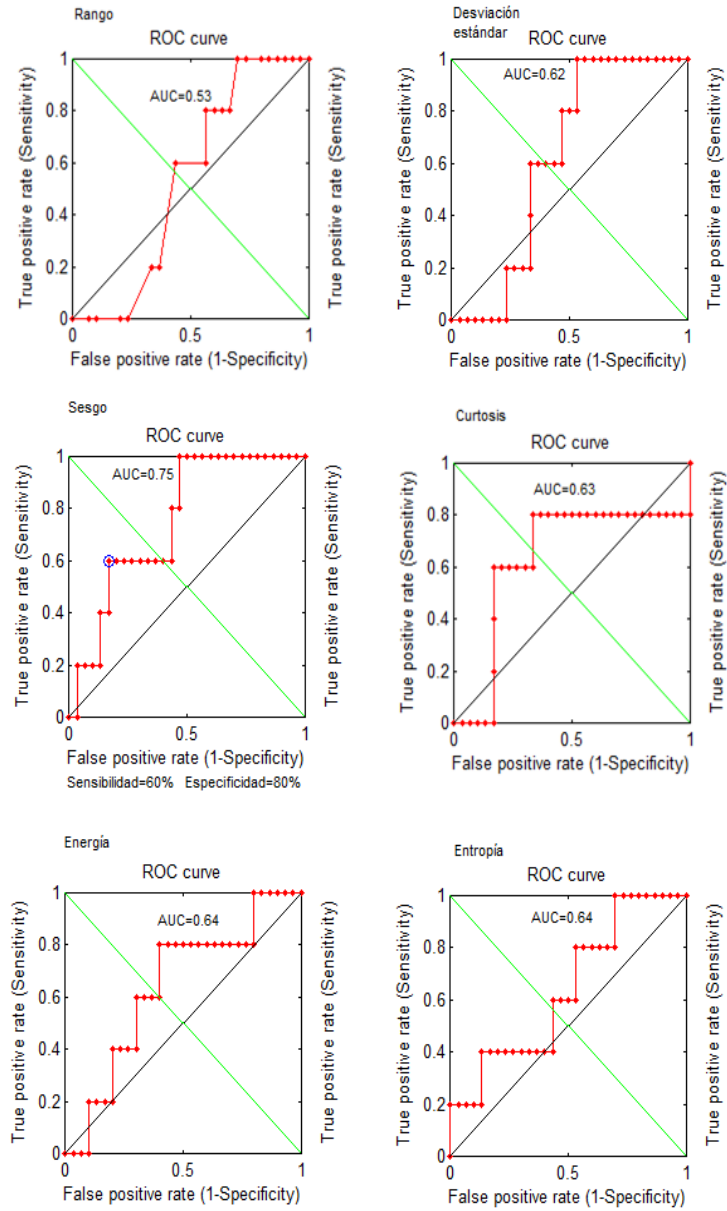
Fuente: Autores.

Tabla 3. Descriptores estadísticos para un sujeto con cáncer.

Sujeto con cáncer	Seno derecho	Seno izquierdo	Diferencia
Media	33,72	34,85	1,13
Mediana	33,7	34,9	1,2
Moda	33.6	35.1	1.5
Máximo	34.3	35.5	1.2

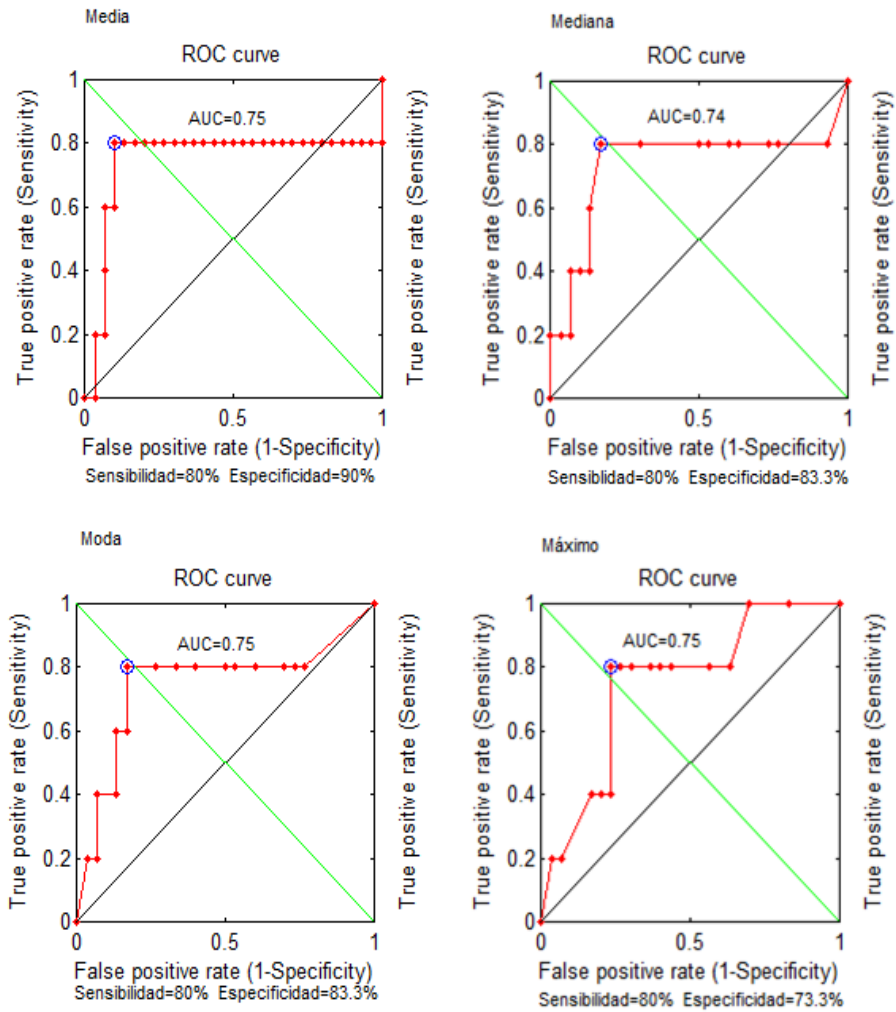
Fuente: Autores.

Figura. 8. Resultados de la curva ROC para los descriptores de rango, desviación, sesgo, curtosis, energía y entropía.



Fuente: Autores

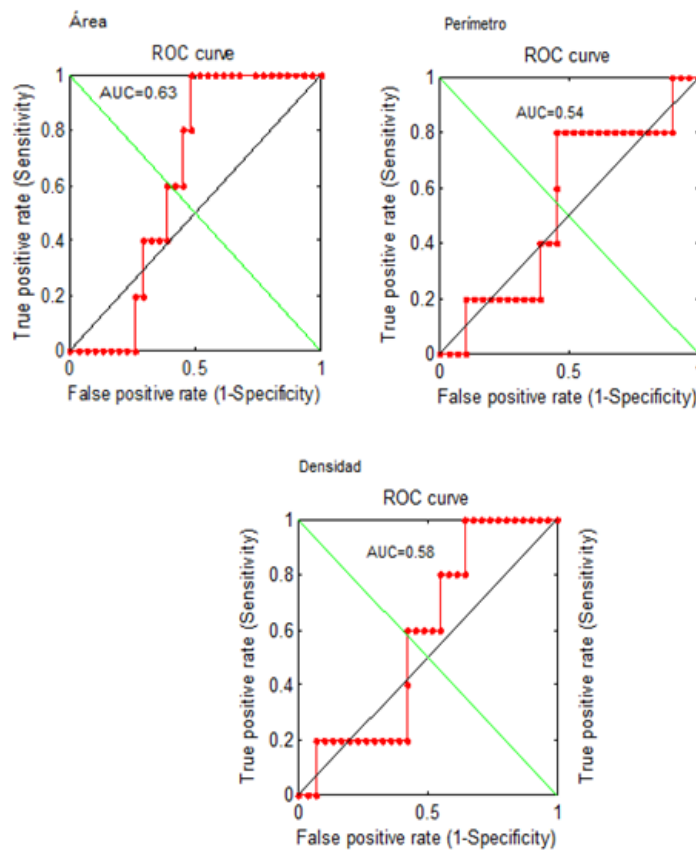
Figura. 9. Resultados de la curva ROC para los descriptores de media, mediana, moda y máximo.



Fuente: Autores.

2.2.2 Descriptores de región simples

Figura 10. Resultados de la curva ROC para los descriptores de región simples.

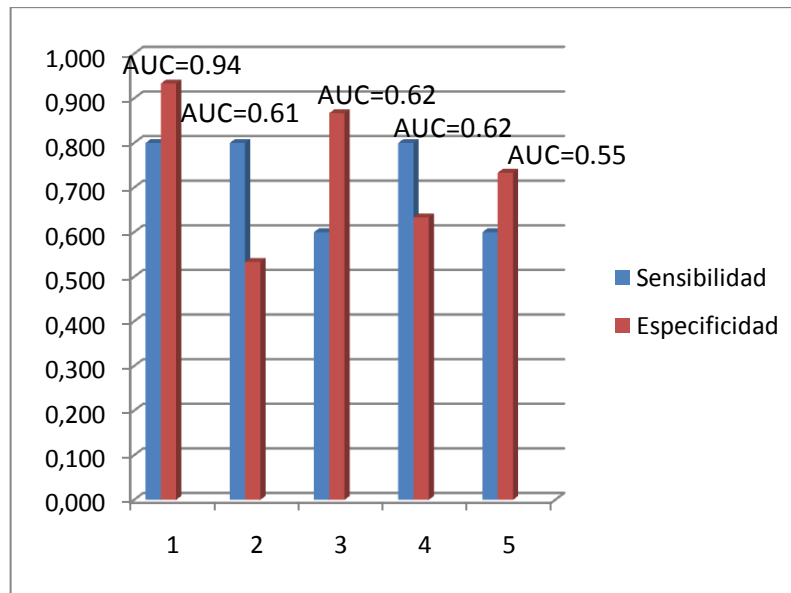


Fuente: Autores

2.3 Red neuronal

La red neuronal se entrenó y probó con las mismas condiciones, fueron realizadas diferentes iteraciones. La Fig. 11 ilustra la sensibilidad, la especificidad y el área bajo la curva ROC que se obtuvo para diferentes redes neuronales desarrolladas.

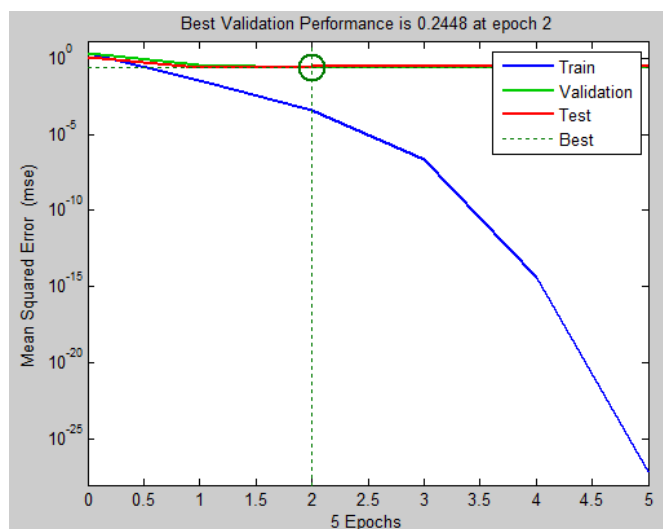
Figura. 11. Sensibilidad y especificidad para cada entrenamiento de la red neuronal.



Fuente: Autores

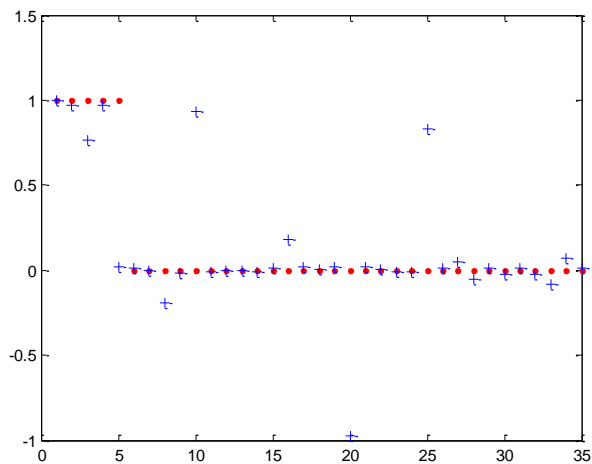
De acuerdo a la evaluación realizada a cada una de las redes, fue seleccionada la primera red neuronal mostrada en la fig.11. En la fig.12 se observa el desempeño que tuvo esta red. La fig. 13 muestra la salida de la red neuronal.

Figura.12. Desempeño del entrenamiento de la red neuronal elegida.



Fuente: Autores

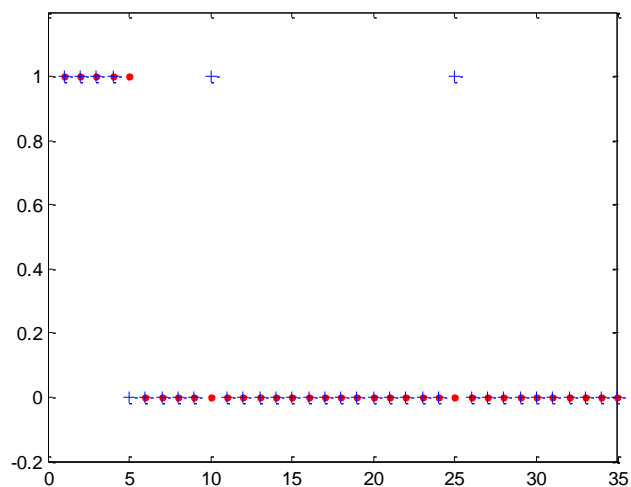
Figura 13. Predicción de la red neuronal.



Fuente: Autores.

La salida de la red neuronal fue binarizada de acuerdo al punto de corte que resultó de la curva ROC que fue de 0.1798 y se encuentra ilustrada en la fig. 14. En esta figura los indicadores rojos representan los resultados previamente conocidos mediante exámenes y biopsias y los indicadores azules representan la predicción por parte de la red neuronal para cada sujeto.

Figura.14. Predicción de la red neuronal umbralizada, luego de aplicar el punto de corte sugerido por la curva ROC (Punto de corte =0.1798)



Fuente: Autores

La matriz de desempeño de la red neuronal se observa en la tabla 4. La cual representa un 80% de sensibilidad y un 93.33% de especificidad para la clasificación de los registros manipulados.

Tabla 4. Matriz de desempeño de la red neuronal seleccionada.

		Reales	
		Casos Positivos	Casos Negativos
Resultados	Casos Positivos	4	2
	Casos Negativos	1	28

Fuente: Autores.

3. Conclusiones y trabajos futuros.

La simetría en el patrón térmico de los senos es de carácter natural. Debido a ello un sujeto saludable debe presentar una distribución de temperatura similar en ellos. Con este estudio se comprueba que esta característica de simetría de temperatura en los senos, es útil para el tamizaje de sujeto con carcinoma en la glándula mamaria.

La descripción de las zonas de mayor vascularización en la glándula mamaria, es de ayuda para el diagnóstico de cáncer ya que caracterizan cada uno de los senos y permite realizar una comparación de la vascularización entre ellos.

De acuerdo con las pruebas realizadas, los descriptores que presentan mejor desempeño son los descriptores estadísticos: media, mediana, moda y máximo en comparación con los descriptores de región simples.

Se podría mejorar el desempeño de este proceso en investigaciones futuras, con un aumento de la población de sujetos a analizar. Este estudio se enmarcó en el proyecto “Evaluación de la Termografía Infrarroja en la Detección del Carcinoma de Glándula Mamaria” código 1102-343-19202 financiado por Colciencias.

BIBLIOGRAFIA

- [1] S.C. Fok, E.Y.K. Ng y K. Tai, *Early Detection and Visualization of Breast Tumor with thermogram and Neuronal Network*, Journal of Mechanics in Medicine and Biology, Vol 2, N°2, 2002, pags 185-195.
- [2] Gautherie Michel y Gros Charles, *Breast thermography and cancer risk prediction*, Cancer Vol. 45, 1980, pag 51-56.
- [3] Tcheslavski Gleb, *Morphological Image Processing: Gray-scale morphology*, 2010.
- [4] Alferéz Santiago, *Detección de carcinoma de glándula mamaria fusionando variables clínicas y termográficas*, Bucaramanga: Universidad Industrial de Santander, 2010.
- [5] Otsu Nobuyuki, *A Threshold Selection Method from Gray – Level Histograms*, IEEE. Trans. SMC-9:1, Enero 1979, pags. 62-66.
- [6] Tang Xianwu, Ding Haishu, *Asymmetry Analysis of Breast Thermograms with Morphological Image Segmentation*, IEEE, Proceedings of the 2005, Septiembre 2005, pags 1680-1683.
- [7] E.Y.K. Ng, L. N. Ung., F.C. Ng. y L.S.J. Sim, *Statistical analysis of healthy and malignant breast thermography*, Journal of Medical Engineering & Technology, Vol 25, N°2, Noviembre/Diciembre 2001, pags 253-263.
- [8] González, Rafael. y Woods, Richard, *Procesamiento Digital de Imágenes*, 3a. Ed. New Jersey: Pearson Prentice Hall, 2008.

[9] Haykin Sadiku. *Neuronal Networks: A comprehensive foundation*. 2a. Ed. Hamilton, Ontario, Canadá : Prentice Hall, 1999.