

# UNIVERSIDAD DE SONORA

## DIVISIÓN DE INGENIERÍA



### POSGRADO EN INGENIERÍA INDUSTRIAL

MODELADO MATEMÁTICO Y TÉCNICAS INFORMÁTICAS  
PARA LA DETECCIÓN DE AFECTACIONES FISIOLÓGICAS  
POTENCIALES EN EL SENNO.

# T E S I S

PRESENTADA POR

**JOSÉ LUIS RUIZ DUARTE**

Desarrollada para cumplir con uno de los  
requerimientos parciales para obtener  
el grado de Maestro en Ingeniería

DIRECTOR DE TESIS: DR. RICARDO ALBERTO RODRÍGUEZ CARVAJAL

HERMOSILLO, SONORA, MÉXICO

AGOSTO 2013

# Universidad de Sonora

Repositorio Institucional UNISON



**“El saber de mis hijos  
hará mi grandeza”**



Excepto si se señala otra cosa, la licencia del ítem se describe como openAccess

## **RESUMEN.**

El presente documento muestra el desarrollo y aplicación de una metodología diseñada para apoyar las técnicas de detección de cáncer de mama alternativas a la mastografía, específicamente la termografía infrarroja. La problemática consiste en que el Departamento de Investigación en Física de la Universidad de Sonora actualmente realiza investigaciones enfocadas a desarrollar el conocimiento necesario para la mejor detección del cáncer de mama. Cuentan con el equipo necesario para la obtención de imágenes térmicas de las personas, pero carece de herramientas que permitan hacer un análisis de los mismos. Como parte de la presente investigación, se realizó una profunda revisión bibliográfica de los métodos de detección actuales y el conocimiento más actual en cuanto a la detección del cáncer de mama por medio de termografía. Se planteó una metodología del tipo correlacional, la cual consistió 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 del médico. Además, se desarrolló un software que apoya la aplicación de técnicas encontradas durante la investigación bibliográfica, así como la aplicación del modelo desarrollado en el presente trabajo.

## **ABSTRACT.**

The present document shows the development and application of a methodology in order to support breast cancer detection techniques different of mammography, specifically, infrared thermography. The actual problematic consists in that the Physics Research Department of the University of Sonora is conducting researches focused in the development of necessary knowledge in order to perform better breast cancer detection. They have the necessary equipment to obtain thermal images of people, but they lack proper tools to analyze such images, called thermograms. As a part of this research, a deep bibliographic review of actual methods and top knowledge of 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 was developed, as well as the application of the model developed in this work.

## **DEDICATORIAS.**

A mi querida Universidad, que debe formar no solo buenos ingenieros y profesionistas que sepan desempeñarse en sus distintas áreas, sino además debe formar excelentes seres humanos, rectos, justos y humildes, y que en su mente y corazón esté impresa en letras de oro la noble tarea de engrandecer a nuestra hermosa y gran nación, que en estos momentos pasa por un período crítico y requiere la ayuda revolucionaria de la humanidad que la Universidad está obligada a forjar. Y esa es tarea de todos los niveles.

A los futuros profesionistas y egresados de posgrado de esta Alma Máter, quienes tienen la enorme tarea de superarnos a nosotros y a todas las generaciones pasadas, pues es la única manera en que el mundo camine hacia un lugar mejor.

A las generaciones futuras. Estamos trabajando por ustedes.

## **AGRADECIMIENTOS.**

En primer lugar a mi madre y a mis hermanos, quienes, a pesar de que no fue una decisión sencilla para la familia, me apoyaron durante estos años y no dejaron de ayudarme cuando fue necesario. Además, les agradezco infinitamente su paciencia, puesto que, ante la falta de tiempo, o de atención, o de constante cansancio y dolores de cabeza o el mal humor que a veces tenía, no ha existido un solo reclamo.

A mis amigos, quienes me han apoyado durante estos dos años, y han seguido constantes a pesar que a veces resulta difícil reunirnos como lo hacíamos antes.

A mis muy queridos alumnos, a quienes durante estos dos años tuve el honor y el gusto de servir, y sobre todo, agradecerles lo que yo he aprendido de todos ustedes. En verdad uno como persona nunca deja de aprender, y ser maestro es una oportunidad de oro.

A mis compañeros maestros, quienes me han asesorado en esta hermosa labor. Especialmente al maestro Ricardo Alberto Rodríguez pues, sin su apoyo y consejos, esta travesía nunca hubiera empezado. También al maestro Carlos Anaya, quien, sin ninguna objeción, ha sido un gran impulso a esta tesis. Y con especial cariño al maestro Jorge Luis Taddei, quien ha sido, además un gran mentor académico, un excelente guía personal y un tremendo amigo.

Al DIFUS, por confiarme esta labor, espero haber cumplido con las expectativas que depositaron en mí.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) y al programa Integral de Fortalecimiento institucional (PIFI, 2012) por su apoyo económico.

# ÍNDICE.

RESUMEN. ....	i
ABSTRACT. ....	ii
DEDICATORIAS. ....	iii
AGRADECIMIENTOS. ....	iv
ÍNDICE DE FIGURAS. ....	viii
ÍNDICE DE TABLAS. ....	x
ÍNDICE DE ECUACIONES. ....	xi
ÍNDICE DE SIGLAS. ....	xii
PENSAMIENTOS. ....	xiii
1. INTRODUCCIÓN. ....	1
1.1 Presentación. ....	1
1.2 Planteamiento del problema. ....	2
1.3 Objetivo general. ....	2
1.4 Objetivos específicos. ....	3
1.5 Hipótesis. ....	3
1.6 Alcances y delimitaciones. ....	3
1.7 Justificación. ....	4
2. MARCO DE REFERENCIA. ....	5
2.1 El cáncer de mama y sus métodos de detección. ....	5
2.2 Procesamiento de imágenes digitales. ....	11
2.2.1 Conceptos básicos. ....	11
2.2.2 Transformación de la intensidad y el filtrado espacial. ....	13

2.2.3	Técnicas de segmentación.....	17
2.3	Análisis de los datos. ....	22
2.3.1	Estadística inferencial. ....	22
2.3.2	Análisis de datos multivariantes. ....	24
2.4	Software.....	27
2.4.1	Software de asistencia para la detección. ....	27
2.5	Análisis de estudios previos. ....	28
3.	METODOLOGÍA. ....	32
4.	IMPLEMENTACIÓN DE LA METODOLOGÍA .....	37
4.1	Obtener la muestra de imágenes térmicas. ....	37
4.2	Definir y extraer las propiedades de interés de las imágenes térmicas. ....	38
4.3	Desarrollar algoritmo para analizar las imágenes. ....	39
4.4	Analizar los datos. ....	50
4.4.1	Realizar propuesta de modelo multivariado. ....	51
4.4.2	Probar modelo multivariado. ....	51
4.4.3	Validar modelo multivariado. ....	52
4.4.4	Realizar propuesta de modelo univariado. ....	55
4.4.5	Probar modelo univariado. ....	56
4.4.6	Analizar correspondencia entre modelos. ....	56
4.5	Incluir modelos en la GUI.....	56
5.	RESULTADOS. ....	58
5.1	Obtención de resultados. ....	58
5.2	Resultados obtenidos. ....	59
6.	CONCLUSIONES, RECOMENDACIONES Y TRABAJOS FUTUROS. ....	60

6.1	Conclusiones. ....	60
6.2	Recomendaciones. ....	60
6.3	Trabajos futuros.....	61
7.	REFERENCIAS.....	62
8.	ANEXOS.....	67

# ÍNDICE DE FIGURAS.

Figura 1.1. Planteamiento del problema. ....	2
Figura 2.1 Negativo de un logo generado con función imadjust.....	14
Figura 2.2. Reducción de valores de intensidad de la imagen de Google. ....	14
Figura 2.3. Una imagen y su histograma antes (i) y después de aplicar histeq (d). ...	15
Figura 2.4. Aplicación de dos filtros diferentes al logo de The Offspring. ....	16
Figura 2.5. a) Imagen original. b) Filtrado espacial no lineal "@max". c) Filtrado espacial no lineal "@min". ....	17
Figura 2.6. Máscaras para detectar líneas con inclinaciones (Martín, 2002). ....	19
Figura 2.7. Operadores de derivación de Roberts (a), de Prewitt (b), de Sobel (c) y de Frei-Chen (d) (Martín, 2002). ....	20
Figura 2.8. Máscara para segunda derivación (Martín, 2002). ....	20
Figura 2.9. Matriz de datos. ....	24
Figura 2.10. Metodología propuesta por Kapoor et al. (2012) para un análisis de los termogramas.....	31
Figura 3.1. Relaciones entre técnicas de la Propuesta de solución.....	32
Figura 3.2. Metodología propuesta. ....	34
Figura 4.1. Binarización de una imagen.....	39
Figura 4.2. Diagrama de flujo de operaciones del algoritmo de análisis. ....	40
Figura 4.3. Diseño de la GUI.....	41
Figura 4.4. Jerarquía entre los objetos.....	43
Figura 4.5. Forma final de GUI .....	44
Figura 4.6. Módulo de pre procesamiento .....	46
Figura 4.7. Módulo de detección de bordes. ....	47
Figura 4.8. Módulo de ROI. ....	48
Figura 4.9. Dispersión de los datos. Temperatura máxima vs umbral.....	50
Figura 4.10. Prueba de modelo multivariado. ....	52
Figura 4.11. Línea de regresión multivariada pronosticada.....	53

Figura 4.12. Casos de validación en la línea de regresión. ....	55
Figura 4.13. Histograma de las diferencias entre temperatura máxima (X) y temperatura umbral (Y). ....	55
Figura 4.14. Módulo de cálculos.....	57

## ÍNDICE DE TABLAS.

Tabla 2.1. Clases de datos soportados por MATLAB® (Gonzalez et al., 2004). .....	13
Tabla 2.2. Control de flujo de operaciones (Gonzalez et al., 2004). .....	13
Tabla 2.3. Opciones permitidas por MATLAB® para la función imfilter (Gonzalez et al., 2004).....	16
Tabla 2.4. Situaciones que pueden ocurrir al decidir sobre $H_0$ (Montgomery y Runger, 2003).....	23
Tabla 3.1. Correspondencia entre errores estadísticos y FN, FP, VN y VP. ....	32
Tabla 4.1. Etiquetas de los objetos. ....	42
Tabla 4.2. Detectores de bordes de MATLAB®. ....	47
Tabla 4.3. Resultados de la validación.....	54
Tabla 4.4. Resultados de la prueba de bondad de ajuste.....	56

## ÍNDICE DE ECUACIONES.

(2.1) .....	13
(2.2) .....	18
(2.3) .....	19
(2.4) .....	19
(2.5) .....	20
(2.6) .....	20
(2.7) .....	22
(2.8) .....	22
(2.9) .....	24
(2.10) .....	25
(2.11) .....	25
(2.12) .....	25
(2.13) .....	25
(2.14) .....	26
(4.1) .....	48
(4.2) .....	51

## ÍNDICE DE SIGLAS.

BG	Button Group
CAD	Diagnóstico Asistido por Computadora (Computer Aided Diagnosis)
CSV	Valores separados por comas (Comma Separated Values)
DIFUS	Departamento de Investigación en Física de la Universidad de Sonora
ET	Edit text
FN	Falso Negativo
FP	Falso Positivo
GUI	Interfaz gráfica de usuario (Graphical User Interface)
GUIDE	Ambiente de desarrollo de GUI (GUI Development Environment)
$H_0$	Hipótesis Nula.
$H_1$	Hipótesis Alternativa.
MATLAB®	Matrix Laboratory
PB	Push Button
PUM	Pop-up Menu
RB	Radio Button
RI	Región de Interés
RNA	Red Neuronal Artificial
RNARP	Red Neuronal Artificial de Retro Propagación
ST	Static Text
SVC	Sistema de Visión Computarizada
VN	Verdadero Negativo
VP	Verdadero Positivo

## **PENSAMIENTOS.**

...Y poco a poco me convertí en esas historias que tanto disfrutaba leer. Con el paso del tiempo, vi cómo mis amigos se torcían, sus sonrisas se convertían en rostros deformes por el alcohol y las drogas. Cambiando sus ideas por el enajenante y vacío fulgor del momento, se olvidaron poco a poco de mí. Pero decidí que ese no era el fin, sino el principio de algo increíble. Me recosté en una almohada confeccionada con mi pasado y mis sueños, y comencé a sentir que mi cuerpo poco a poco se volvía pesado. Pero no de ese tipo de peso que te impide levantarte, sino como aquel que sientes al cargar una cubeta muy pesada de la cuál te puedes desprender. Y finalmente así pasó. Sentí que mi cuerpo quedaba inerte en el lecho del ayer, y mi alma surgía, como un halcón que emprende un vuelo solitario hacia una tierra desconocida.

Comienza la travesía. La soledad y mis confusos recuerdos de un mundo terrenal que parecía inconcebiblemente lejano, destrozaron por completo mi sentido de la orientación. Crucé desde las puertas del infierno hasta los campos elíseos. Vi miles de estrellas nacer y morir en un instante, como si comparado con ellas la vida de millones no significara nada. Me vi forzado a partir mi alma en dos para encontrar el rumbo: la luz y la oscuridad. La luz permitió navegar en los mares más violentos, mientras que la oscuridad me permitió entender el amor habitante del silencio. Fue como si todo adquiriera un sentido divino, y al siguiente instante lo perdiera. Muchas otras almas, algunas libres, otras todavía cautivas, me han permitido vislumbrar medianamente el camino, como faroles en una calle peligrosa de una ciudad violenta por la noche. De no ser por ellos, yo no sería capaz de escribir esto, puesto que justo ahora me encuentro bajo la luz de uno, mientras escribo estas palabras...

Pero debo admitir que al final fue divertido. Sólo así una persona es capaz de crear y de soñar. ¿Qué hay en la ciencia si no un poco de locura?

Escrito el 11 de mayo de 2013, a las 00:36, en uno de mis momentos de locura.

# 1. INTRODUCCIÓN.

En este apartado se enlistan los puntos que dan pie a la investigación. En una primera instancia se presentan los antecedentes de la problemática que será estudiada, es decir, sobre el cáncer de mama y sus implicaciones en la salud de las personas. En la segunda parte se manifiesta el problema que se pretende resolver. Seguido se exponen los objetivos, general y específicos. Luego se mencionan las hipótesis que se someterán a prueba durante el desarrollo del proyecto. Después se especifican los alcances y delimitaciones del proyecto. Al final del capítulo se obtiene la justificación de la investigación.

## 1.1 Presentación.

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

El Departamento de Investigación en Física de la Universidad de Sonora (DIFUS) ha mostrado su disposición e interés en desarrollar un método para detección y diagnóstico temprano de este tipo de cáncer. El DIFUS actualmente lleva a cabo investigaciones con sistemas nano estructurados a base de plata, oro y diamante, con la finalidad de desarrollar el conocimiento de fenómenos fisiológicos relacionados con procesos celulares y especialmente aquellos útiles para la detección y diagnóstico del cáncer de mama mediante un método no invasivo o mínimamente invasivo.

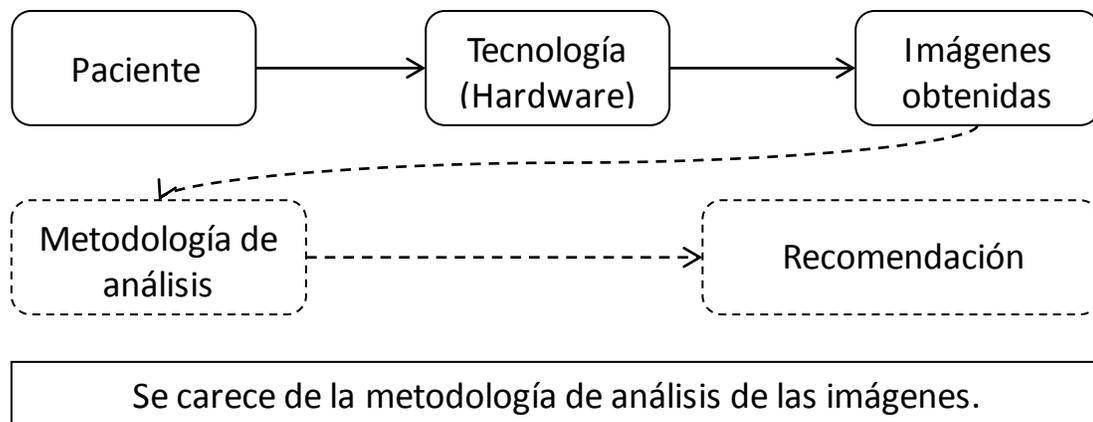
Se plantea realizar estudios termográficos de una población femenina ( $N > 100$ ) mayor de 20 años, sujeta a un protocolo de control clínico adecuado para obtener modelos del tejido mamario sano y anómalo que faciliten la detección y diagnóstico oportuno de personas afectadas por el cáncer de mama.

Por otro lado, y aunado a la tecnología física que se utilizará para la obtención de las imágenes térmicas –también conocidas como termogramas-, es importante analizar dichas imágenes y detectar en estas los diferentes gradientes de temperatura que indiquen una asimetría en el cuerpo de la persona, revelando así una anomalía y verificando si se trata de algún padecimiento.

Se considera que se pueden mejorar los indicadores de detección de los diagnósticos, utilizando un método de detección no invasivo, inocuo, sin contacto físico con el paciente y que no utilice radiación ionizante de ningún tipo.

## 1.2 Planteamiento del problema.

Tal como se muestra en la Figura 1.1, se puede observar que la situación actual consiste en que el DIFUS cuenta con la tecnología necesaria para la obtención de las imágenes, pero carecen de una metodología propia para analizarlas y emitir las recomendaciones pertinentes.



*Figura 1.1. Planteamiento del problema.*

## 1.3 Objetivo general.

Desarrollar y validar un sistema basado en técnicas de modelado matemático y aplicaciones informáticas para la detección de anomalías térmicas en los termogramas del seno de mujeres, que puedan ser causadas por algún proceso fisiológico relacionado con el cáncer de mama.

## 1.4 Objetivos específicos.

- Generar un modelo térmico, basado en datos estadísticos, de personas con anomalías térmicas y fisiológicas que sirva de comparación con el perfil de aquellas personas que se sometan a la prueba.
- Desarrollar un sistema informático basado en modelos térmicos del seno que permita determinar anomalías y asimetrías térmicas, si estas se encuentran fuera de ciertos límites estadísticos y la región donde se encuentran.

## 1.5 Hipótesis.

El desarrollo y la implementación de un sistema de tratamiento de imágenes facilitarán la detección de anomalías fisiológicas potenciales a partir de anomalías térmicas.

## 1.6 Alcances y delimitaciones.

El proyecto contempla desarrollar el prototipo de una herramienta capaz de detectar una anomalía térmica y determinar la región donde se encuentra, mediante comparación con patrones conocidos, así como si dicha anomalía térmica tiene una probabilidad relativamente elevada de corresponder a una anomalía fisiológica. No es parte del alcance del proyecto conocer si la anomalía corresponde a una anomalía fisiológica plenamente identificada por un especialista.

Algunas limitaciones del proyecto son:

- Acceso a información disponible en las bases de datos que tiene acceso la Universidad de Sonora y aquellas libres.
- La resolución de las imágenes térmicas está en función de la capacidad del equipo que utilice el DIFUS para obtenerlas.
- No se tiene conocimiento médico sobre las diversas anomalías térmicas que se relacionen potencialmente con anomalías fisiológicas.
- El tiempo para realizar la investigación debe estar apegado a lo requerido por el DIFUS.

- El tiempo para desarrollar el proyecto debe estar apegado a la duración de la Maestría, por lo que no es parte del proyecto su validación con expertos en medicina.

## **1.7 Justificación.**

Esta investigación forma parte de una más amplia que busca beneficiar a la población total de mujeres con edad a partir de los 20 años, debido a que el cáncer de mama se ha detectado desde estas edades (Knaul et al., 2009).

Los métodos que actualmente sugieren las normatividades mexicanas para la detección del cáncer, particularmente la mastografía de rayos X, exhibe baja sensibilidad en la presencia de tejido glandular denso, un rango elevado de falsos positivos y requiere compresión del seno y exposición a radiación; mientras que un termograma -obtenido mediante métodos no invasivos, sin exposición a radiación-anormal ha demostrado ser un indicador confiable de un riesgo alto de cáncer de mama en etapas tempranas (Jiang et al., 2010). La termografía tiene una sensibilidad y especificidad promedio de 90%. En los últimos años, y gracias al Diagnóstico Asistido por Computadora, este nivel ha superado los de la mastografía (Moghbel y Mashohor, 2011). Una detección temprana permite más opciones clínicas en el tratamiento.

Por las razones ya mencionadas, muchos países están considerando a la termografía como el sistema de detección primario de cáncer; por ejemplo en Japón, existen más de 1500 hospitales y clínicas, que actualmente utilizan dicho método (Moghbel y Mashohor, 2011). En México aún no se aprovecha dicha tecnología.

El principal producto de esta investigación busca brindar al equipo de investigadores del DIFUS las herramientas necesarias para analizar las imágenes térmicas obtenidas a partir de equipos sofisticados para tal propósito. Estas herramientas servirán para encontrar variaciones de temperatura en los termogramas de las personas que posiblemente se deberán a causas fisiológicas.

## **2. MARCO DE REFERENCIA.**

En este apartado se recopila y expone la base teórica sobre la problemática, misma que sustentará la metodología que será aplicada. En un principio se presenta información sobre el cáncer de mama y los métodos de detección actuales, haciendo énfasis en el método de la termografía infrarroja y se comienza a tocar el tema de las imágenes, lo que guía al lector de manera natural hacia la segunda parte, referente al procesamiento de imágenes. En ella se hablan de algunas técnicas existentes para procesar imágenes y obtener datos estadísticos. Esto orienta la discusión hacia el siguiente punto, el análisis de los datos estadísticos obtenidos. En este apartado se abordan técnicas de estadística univariada y multivariada. Después, una sección recupera de manera general algunos trabajos sobre los tipos de software que se han utilizado para ayudar a la detección del cáncer de mama a partir de imágenes térmicas. Al final del capítulo se muestra el análisis de algunos estudios previos.

### **2.1 El cáncer de mama y sus métodos de detección.**

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 cáncer puede ser categorizado como invasivo o no invasivo basándose en su patrón de crecimiento característico: si este está restringido se considera tumor benigno. Si es rápido y tiene habilidad para progresar a otras partes del cuerpo se considera tumor maligno o carcinoma (Ng y Kee, 2008). Una detección temprana permite más opciones clínicas en el tratamiento (Jiang et al., 2010). El cáncer de mama tiene una tasa de supervivencia a diez años de 95% en la etapa 0, 88% en la etapa I, 66% en la etapa II, 36% en la etapa III y 7% en la etapa IV (Köşüş 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. Si un caso es correctamente diagnosticado como saludable se le conoce como verdadero negativo (VN), si el diagnóstico es incorrecto se le conoce como falso negativo (FN). Si un caso es correctamente diagnosticado como enfermo se le conoce como verdadero positivo (VP), si el diagnóstico es incorrecto se le conoce como falso positivo (FP) (Ng y Kee, 2008).

Estudios realizados por Elmore et al. (1998) han revelado que el 23.8% de los casos han obtenido un falso positivo en mastografía, 13.4% por medio de exploración de pecho y 31.7% al menos un falso positivo por alguno de los dos métodos. Además, si se realizan pruebas anuales durante un periodo de 10 años, el riesgo de obtener un falso positivo por medio de mastografía es del 49.1% y de 22.3% si se hace por exploración de pecho. Padecer un falso positivo en la detección de cáncer de mama tiene un efecto psicológico, aunque no produce morbilidad psicológica (Valera, 2010). Estudios realizados por Ohuchi et al. (1995) han comparado las proporciones de falsos negativos en estudios de exploración física con o sin mastografía. Los porcentajes de detección son de 0.28% y 0.09% para exploración física con y sin mastografía, mientras 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). En la siguiente parte se abordan de manera general los casos de mastografía, ultrasonido e imágenes térmicas.

Los indicadores para la evaluación de un método de detección son, entre otros, según la obra de Ng y Kee (2008):

- Sensitividad (S) es la probabilidad que una prueba sea positiva en una población no saludable. Se calcula como el cociente de VP entre la suma de VP y FN.
- Especificidad (E) es la probabilidad que la prueba sea negativa en una población saludable. Se calcula como el cociente de VN entre la suma de VN y FP.

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

#### **a) Mastografía.**

La mamografía –o 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).

Estudios realizados por Harris (2009) han detectado que la exposición a bajas dosis de radiación durante la mastografía incrementa el riesgo de desarrollar cáncer de mama hasta 1.5 veces en mujeres que cuentan con predisposición congénita. Además, aquellas mujeres menores de 20 años o con más de 5 exposiciones a radiaciones son 2.5 veces más propensas de desarrollar el cáncer, que aquellas que no han sido expuestas.

#### **b) Ultrasonido**

Ultrasonido se usa para detectar lesiones de seno y es una herramienta adjunta para detectar lesiones sospechosas. El transductor de ultrasonido dirige ondas de sonido de alta frecuencia en los tejidos del seno y detecta las ondas de sonido que se reflejan, mostrando imágenes en 2D, así como imágenes continuas en tiempo real. El ultrasonido es muy útil para detectar anomalías palpables y no palpables en el seno. La mastografía por sí sola puede dejar pasar muchos casos de cáncer, para los cuales el ultrasonido ha mejorado la detección (Sree et al., 2011).

### **c) Imágenes térmicas y su sustento.**

La imagenología infrarroja ha sido utilizada en aplicaciones como oncología, dolor, desórdenes vasculares, artritis, reumatismo, entre otras (Jiang et al., 2005). Fue aprobada por la Food and Drugs Administration (FDA) de EEUU en 1982 como complemento a la mastografía para detectar cáncer de mama (Köşüş et al., 2010).

La termografía es un método de imagen no invasivo que es económico, rápido y no causa ningún daño sobre el paciente (Ng y Kee, 2008). Una de sus ventajas es que puede ser usada repetidas veces sin causar daño al paciente y puede detectar cambios sutiles que indiquen una variedad de enfermedades y anomalías del seno (Ng et al., 2001b). En su estudio, Damnjanović et al. (2010) indican que la visión térmica tiene, entre otras, las siguientes ventajas:

- Es un método inofensivo, pues no hay radiación o influencia dañina porque sólo registra la radiación del cuerpo.
- Puede diagnosticar enfermedades en etapas tempranas, antes que haya cambios en la funcionalidad de los órganos.
- No requiere preparaciones especiales para tomar una imagen.

La termografía es complementaria de las otras técnicas. Por sí sola no es capaz de generar diagnóstico, por ahora, pero es útil para brindar más información al diagnóstico obtenido por otros métodos (Ng y Kee, 2008). Cuando la termografía infrarroja se utiliza como un enfoque multimodal (examen clínico, mamografía y termografía) se pueden detectar hasta el 95% de los casos de cáncer temprano (Jiang et al., 2005).

La termografía en la detección de cáncer de mama se basa en el principio de que si existe un tumor 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 investigación de Amri et al. (2011) indica que la temperatura de la piel sobre un tumor maligno es frecuentemente entre 1 y 3° más caliente que la piel alrededor.

Asimismo, un tumor puede causar que la distribución de temperatura se altere, causando asimetría térmica entre los senos (Ng et al., 2001b). Además, se asume que el cuerpo humano tiene simetría térmica, por lo que las variaciones en ambos senos deben ser similares (Ng et al., 2002).

Los desarrollos recientes en cámaras térmicas digitales y técnicas de análisis computarizado de termogramas han mejorado la confiabilidad de la detección de cáncer basado en termogramas (Jiang et al., 2010). La termografía ha ayudado a los físicos a diferenciar tejido benigno y maligno por medio de la caracterización de diferentes patrones en la señal infrarroja emitida por el tejido (Köşüş et al., 2010).

La utilización de termografía permite detectar cáncer hasta 10 años antes de que otras modalidades de imagen lo permitan, 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 por la edad ni por la densidad de los senos (Moghbel y Mashohor, 2011).

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

Una vez que la temperatura de cada pixel es compilada, se pueden aplicar técnicas bioestadísticas para analizarlos; por ejemplo, determinando la media, la mediana y la temperatura modal de una región del seno (Ng y Kee, 2008).

Las imágenes del seno pueden seccionarse en múltiples áreas, para evitar errores, entonces el especialista puede comparar la distribución de la temperatura de manera sistemática en las áreas correspondientes de ambos senos (FOK et al., 2002).

**d) Otros.**

Otros métodos de detección, mencionados en el trabajo de Ng y Kee (2008), son:

➤ **A partir de la teoría de ondas:**

**Scintimamografía de rayos gamma.** Ha sido propuesta como un complemento ideal para la mamografía de rayos X y el ultrasonido debido a su habilidad para evaluar la naturaleza de las lesiones de seno (Scopinaro et al., 1999).

**Tomografía de emisión de positrones.** Es una técnica de imagenología funcional no invasiva con buena resolución, buena sensibilidad y cuantificación precisa sobre procesos fisiológicos, bioquímicos y farmacológicos (Marques de Miranda et al., 2012). Detecta pares de rayos gama, que son emitidos de un radio nucleído que es introducido al cuerpo humano. Los tumores malignos, que están caracterizados por un metabolismo de glucosa elevado, se observan mediante contraste. Los pacientes están sujetos a exposición de radiación (Sree et al., 2011).

**Imagenología de microondas.** Se basa en la recuperación de propiedades eléctricas del tejido y en que existe un contraste dieléctrico entre tejido maligno y normal (Golnabi et al., 2011).

**Imagen óptica de luz infrarroja.** Es una técnica de imagenología no invasiva que es capaz de analizar imágenes objetivos a bajo costo (Nakayama et al., 2011).

➤ **A partir de propiedades eléctricas.**

**Imagenología de impedancia eléctrica.** Consiste en medir la habilidad de los tejidos para conducir y almacenar electricidad (Oran y Ider, 2012). Se ha demostrado que los tumores difieren del tejido sano circundante e incluso se han desarrollado métodos para mejorar los procedimientos de posicionamiento de agujas, lo cual es necesario cuando se somete al paciente a biopsias o cirugías de mínima invasión (Laufer et al., 2012).

**Mediciones de potencial eléctrico.** Los cambios eléctricos pueden brindar información sobre crecimiento normal o canceroso, ya que se ha encontrado que los

carcinomas de células basales son significativamente más electropositivos comparados con células de tejido normal (Marino et al., 1994).

➤ **A partir del campo del sonido y magnetismo.**

**Imagen de resonancia magnética.** Utiliza núcleos de hidrógenos y sus propiedades magnéticas para propósitos de imágenes, ya que es abundante en agua y grasa. El paciente es posicionado en un campo magnético y una onda de frecuencia de radio se aplica para crear imágenes de alto contraste del seno (Sree et al., 2011).

**Espectroscopia de resonancia magnética.** Se ha demostrado que los estudios de espectroscopia de resonancia magnética utilizando hidrógeno han identificado ciertos componentes que están presentes en lesiones malignas, pero no en lesiones benignas ni en tejido normal (Bolan et al., 2005).

**Elastografía.** Se sugiere que las propiedades elásticas del tejido pueden ayudar a determinar la existencia de cáncer. Un tejido canceroso puede ser de 5 a 20 veces más difícil de doblar que el tejido adiposo glandular (McKnight et al., 2002).

## 2.2 Procesamiento de imágenes digitales.

Debido a que se ha mencionado con anterioridad la utilización de imágenes térmicas para la detección del cáncer de mama, es importante mencionar algunas herramientas para el análisis y procesamiento de imágenes.

### 2.2.1 Conceptos básicos.

Algunos conceptos claves definidos por Gonzalez et al. (2004) son:

- **Imagen.** Función de dos dimensiones  $f(x, y)$  donde  $x$  e  $y$  son coordenadas espaciales y la amplitud de  $f$  en cada pareja de coordenadas  $(x, y)$  representa la intensidad del nivel de gris en ese punto. Si  $x$ ,  $y$  y  $f$  son finitos, podemos entonces decir que la imagen es una imagen digital. Las imágenes a color se forman mediante la combinación de varios arreglos de dos dimensiones. Por ejemplo, una imagen RGB consta de 3 arreglos individuales, uno para el color Rojo, otro para el Verde y uno para el Azul.

- **Pixel.** Elementos finitos de una imagen que cuentan con localización y un valor específico.
- **Resolución.** Cantidad de pixeles que forman una imagen, escritos en un arreglo [Reglones, Columnas].

Los valores de la intensidad de una imagen dependen si esta es en blanco y negro, en escala de gris o a color. Para imágenes a blanco y negro,  $f(x,y)$  es binaria, donde 0 es negro y 1 es blanco. 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 con valores de 0 a 255 para cada componente (Megahed et al., 2011).

De acuerdo con Jiang et al. (2005), el procesamiento de imágenes infrarrojas consiste en aplicación de técnicas de procesamiento de imágenes, por ejemplo:

- Realzado de imagen. Mejora de la resolución, remoción el ruido, etc.
- Medición de temperatura.

Algunas herramientas de procesamiento de imágenes incluyen isoterms, perfiles de línea, histogramas de área, etc. (Jiang et al., 2005).

De acuerdo con Gonzalez et al. (2004), el software Matrix Laboratory® (MATLAB®) tiene la capacidad para analizar imágenes y convertirlas en arreglos numéricos, los cuales pueden someterse a operaciones matemáticas diversas. MATLAB® cuenta con una serie de funciones para el procesamiento de imágenes, las cuales pueden contar con más de una sintaxis (se recomienda revisar los archivos de ayuda para cada caso). Algunas de estas funciones son:

- **Imread.** Sirve para leer una imagen dentro del ambiente de MATLAB®. Convierte la imagen en una matriz. Los formatos reconocidos por este comando son TIFF, JPEG, GIF, BMP, PNG y XWD.
- **Size.** Arroja el número de renglones y columnas que tiene una imagen, en forma de un vector renglón de 1x2.
- **Imshow.** Muestra la imagen a partir de una variable.
- **Imwrite.** Escribe un archivo de imagen a partir de una variable.

Las clases de datos son las diversas maneras de representar la intensidad de los valores de un pixel. MATLAB® soporta las clases de datos indicadas en la Tabla 2.1. Las herramientas para controlar el flujo de las operaciones se muestra en la Tabla 2.2.

Nombre	Descripción
double	Doble precisión. Valores que van de $-10^{308}$ a $10^{308}$
uint8	Valores que se encuentran en el rango [0,255]
uint16	Valores que se encuentran en el rango [0,65535]
uint32	Valores que se encuentran en el rango [0,4294967295]
int8	Valores que se encuentran en el rango [-128,127]
int16	Valores que se encuentran en el rango [-32768,32767]
int32	Valores que se encuentran en [-2147483648,2147483647]
single	Precisión sencilla. Valores que van de $-10^{38}$ a $10^{38}$
char	Caracteres
logical	Valores 0 o 1

**Tabla 2.1.** Clases de datos soportados por MATLAB® (Gonzalez et al., 2004).

Comando	Descripción
if	En conjunto con "else" y "elseif" ejecutan un grupo de comandos basados en una condición lógica.
for	Ejecuta un grupo de comandos un número determinado de veces.
while	Ejecuta un grupo de comandos un número no definido de veces, basados en una condición lógica específica.
break	Termina una ejecución de un ciclo "for" o "while".
continue	Pasa a la siguiente iteración de un ciclo "for" o "while", sin tomar en cuenta el resto de los comandos.
switch	Junto con "case" y "otherwise", ejecutan diferentes grupos de comandos dependiendo de un valor determinado.
return	Causa que se regrese a la función de invocación.
try...catch	Cambia el control del flujo si se detectan errores durante la ejecución.

**Tabla 2.2.** Control de flujo de operaciones (Gonzalez et al., 2004).

Los procesos sobre una imagen se pueden denotar tal como en la Ecuación (2.1) (Gonzalez et al., 2004):

$$g(x,y) = T[f(x,y)] \quad (2.1)$$

Donde  $f$  es la imagen de entrada,  $g$  es la imagen de salida y  $T$  es un operador en  $f$ .

### 2.2.2 Transformación de la intensidad y el filtrado espacial.

De acuerdo con Gonzalez et al. (2004), se pueden dividir en varios tipos de procesos sobre la imagen.

### a) Funciones de transformación de la intensidad.

Se dan cuando se obtiene una imagen nueva a partir de una imagen existente que es alterada pixel por pixel en función de los pixeles que rodean cada pixel procesado. El caso más sencillo es cuando sólo se trabaja sin vecindad.

La función *imadjust*. Es la herramienta más sencilla de MATLAB®. Calcula la intensidad de los valores en la imagen  $f$  y asigna nuevos valores en la imagen  $g$ . Se puede utilizar para generar un negativo, como en la Figura 2.1, o para amplificar o reducir los valores de intensidad de la imagen, como en la Figura 2.2.



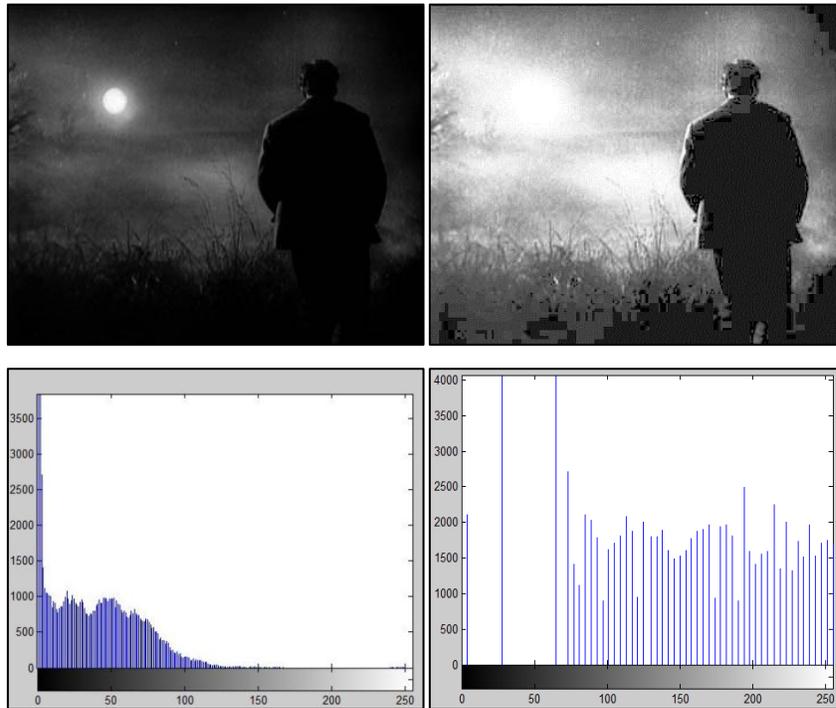
Figura 2.1 Negativo de un logo generado con función *imadjust*.



Figura 2.2. Reducción de valores de intensidad de la imagen de Google.

### b) Procesamiento de histograma.

El histograma de una imagen nos indica el número de pixeles que tienen un determinado nivel de gris de la imagen (Gonzalez et al., 2004). Las funciones más importantes sobre el histograma de una imagen son *imhist*, que muestra dicho histograma, y *histeq*, que ayuda a ecualizar (distribuir la cantidad de pixeles durante todo el rango de intensidades) el histograma, asimilándose a una distribución de probabilidad especificada por el usuario, como se muestra en la Figura 2.3.



**Figura 2.3.** Una imagen y su histograma antes (i) y después de aplicar histeq (d).

### c) Filtrado espacial.

El procesamiento de filtrado espacial consiste en el siguiente procedimiento:

- Definir un punto central  $(x,y)$ .
- Llevar a cabo una operación que involucre solamente los píxeles en un vecindario predefinido alrededor de dicho punto central.
- El resultado de la operación sea el resultado del proceso en el punto central.
- Repetir el proceso para cada punto de la imagen.

Al mover el punto central se generan nuevas vecindades. El procedimiento de un filtrado espacial lineal consiste en multiplicar cada píxel en la vecindad por un coeficiente correspondiente y sumar los resultados para obtener la respuesta en cada punto  $(x,y)$ . Los coeficientes están arreglados como una matriz, conocida como filtro o máscara. En cada punto  $(x,y)$ , la respuesta del filtro en dicho punto es la suma de productos de los coeficientes del filtro y la vecindad correspondiente a los píxeles en el área que abarcó el filtro. Para una máscara de tamaño  $m \times n$ , se asume

normalmente que  $m = 2a + 1$  y  $n = 2b + 1$ , donde  $a$  y  $b$  son enteros no negativos. Es mejor cuando las matrices son de tamaño impar, pues así se tiene un centro único. Correlación es el proceso de pasar la máscara  $\omega$  a través del arreglo de imagen  $f$ . Convolución es el proceso de correlación con la excepción que  $\omega$  sea rotado  $180^\circ$  antes de pasarlo a través de  $f$ . Para que la máscara pueda trabajar desde el primer pixel, MATLAB® añade los pixeles faltantes a la imagen con un valor de 0. MATLAB® utiliza la función *imfilter* para el filtrado espacial lineal, que permite las opciones que se observan en la Tabla 2.3. La aplicación de diversas máscaras en una imagen se puede observar en la Figura 2.4.

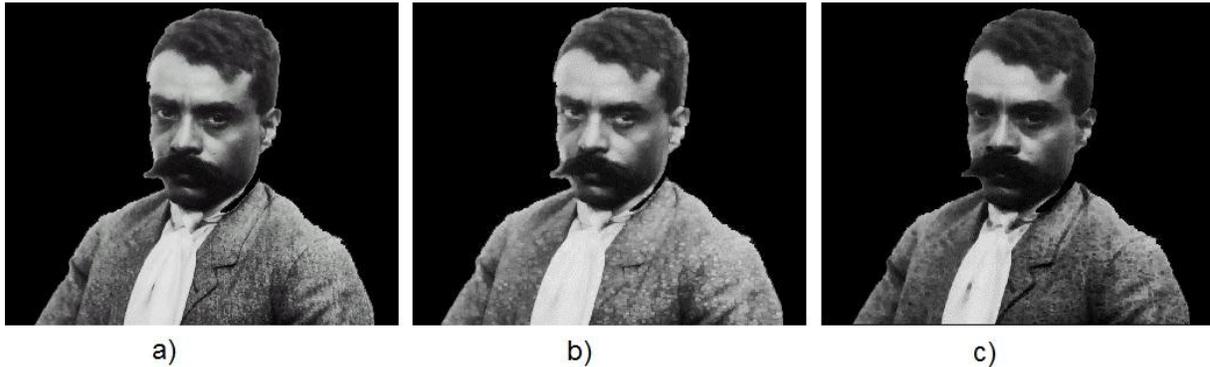
Opción	Descripción
Modo de filtro	
'corr'	Permite usar correlación. Es predeterminado.
'conv'	Permite usar convolución.
Opciones de límite.	
P	Los límites de la imagen son extendidos en P valores. El predeterminado es 0.
'replicate'	El tamaño de la imagen es extendido replicando valores en su límite exterior.
'symmetric'	El tamaño de la imagen es extendido mediante reflexión de espejo en sus bordes.
'circular'	El tamaño de la imagen es extendido tratando a la imagen como una función de periodo 2D.
Opciones de tamaño	
'full'	Las salidas tienen el mismo tamaño que la imagen extendida.
'same'	Las salidas tienen el mismo tamaño que la imagen e entrada. Es el predeterminado.

**Tabla 2.3.** Opciones permitidas por MATLAB® para la función *imfilter* (Gonzalez et al., 2004).



**Figura 2.4.** Aplicación de dos filtros diferentes al logo de The Offspring.

La diferencia entre el Filtrado espacial lineal y el Filtrado espacial no lineal son las operaciones que sufren los píxeles en la vecindad. MATLAB® provee dos funciones para filtrado no lineal: *nfilter* y *colfilt*. Utilizando las funciones *mean* y *sum* para filtrado no lineal, se obtienen los resultados de la Figura 2.5.



**Figura 2.5.** a) Imagen original. b) Filtrado espacial no lineal "@max". c) Filtrado espacial no lineal "@min".

La herramienta de MATLAB® Image Processing Toolbox cuenta con filtros estándares (Gonzalez et al., 2004). El filtrado Gaussiano remueve el ruido de patrón fijo y el ruido de columna fijo y mejora el índice señal-ruido (Joro et al., 2008).

### 2.2.3 Técnicas de segmentación.

Segmentación de la imagen se refiere a clasificación de píxeles en Píxeles de Fondo y Píxeles Señal (Laza Dueñas et al., 2010). Esta se agrupa en cuatro tipos:

1. Técnicas basadas en cotas. Todos los píxeles cuyo valor de intensidad caiga en cierto rango pertenecen a una clase. Hay un umbral que permite definir cuando un píxel pertenece a la señal o al fondo.
2. Métodos basados en bordes. Los valores de los píxeles cambian bruscamente alrededor de los bordes de los objetos. Estos se pueden detectar.
3. Métodos basados en regiones. Píxeles vecinos dentro de una región dada tienen valores de intensidad similar. El procedimiento general compara cada píxel con los de cierta región y si se satisface la condición de homogeneidad de interés se agrega a la misma.

4. Técnicas que combinan bordes y regiones. Incluyen segmentación morfológica y ajuste de superficie.

Una vez que se ha llevado a cabo la segmentación, sigue la etapa de extracción de características, en la que se calculan las medidas de interés para las regiones segmentadas en la imagen: áreas, volúmenes, promedios. Se pueden utilizar las intensidades de los píxeles de la región a medir, pero pudieran usarse también la cantidad de píxeles de los puntos pertenecientes a una región dada (Laza Dueñas et al., 2010).

El siguiente apartado es un extracto del trabajo reportado por Martín (2002). El primer paso para cualquier proceso de análisis es la segmentación, es decir, dividir la imagen en los objetos que la forman. La segmentación termina cuando se han detectado todos los objetos de interés.

Los algoritmos de segmentación se basan en dos propiedades básicas de los niveles de gris: la discontinuidad (basado en los cambios bruscos en los niveles de gris) y similitud (técnicas de umbrales, crecimientos de regiones, y de división y de fusión).

**a) Detección de discontinuidades.**

El método más común es la correlación de la imagen con una máscara. Se realiza el producto correspondiente a los píxeles de la imagen encerrados por la máscara. La respuesta a la máscara (en este caso de una de tamaño 3x3) de cualquier píxel de la imagen viene dado por la Ecuación (2.2) (Martín, 2002).

$$R = \sum_{i=1}^9 w_i x \quad (2.2)$$

Donde  $x_i$  es el nivel de gris de la imagen y  $w_i$  el coeficiente de la máscara.

**b) Detección de puntos.**

Es inmediata. Se detecta un punto si  $|x_i| > T$ , donde  $T$  es un umbral. Se mide la diferencia entre el píxel central y sus vecinos, un píxel será un punto aislado siempre que sea suficientemente distinto de sus vecinos.

**c) Detección de líneas.**

Utilizando las máscaras de la Figura 2.6 se pueden detectar líneas con diversas inclinaciones.

<table border="1"> <tbody> <tr><td>-1</td><td>-1</td><td>-1</td></tr> <tr><td>2</td><td>2</td><td>2</td></tr> <tr><td>-1</td><td>-1</td><td>-1</td></tr> </tbody> </table> <p>Horizontal</p>	-1	-1	-1	2	2	2	-1	-1	-1	<table border="1"> <tbody> <tr><td>-1</td><td>-1</td><td>2</td></tr> <tr><td>-1</td><td>2</td><td>-1</td></tr> <tr><td>2</td><td>-1</td><td>-1</td></tr> </tbody> </table> <p>45°</p>	-1	-1	2	-1	2	-1	2	-1	-1	<table border="1"> <tbody> <tr><td>-1</td><td>2</td><td>-1</td></tr> <tr><td>-1</td><td>2</td><td>-1</td></tr> <tr><td>-1</td><td>2</td><td>-1</td></tr> </tbody> </table> <p>Vertical</p>	-1	2	-1	-1	2	-1	-1	2	-1	<table border="1"> <tbody> <tr><td>2</td><td>-1</td><td>-1</td></tr> <tr><td>-1</td><td>2</td><td>-1</td></tr> <tr><td>-1</td><td>-1</td><td>2</td></tr> </tbody> </table> <p>-45°</p>	2	-1	-1	-1	2	-1	-1	-1	2
-1	-1	-1																																					
2	2	2																																					
-1	-1	-1																																					
-1	-1	2																																					
-1	2	-1																																					
2	-1	-1																																					
-1	2	-1																																					
-1	2	-1																																					
-1	2	-1																																					
2	-1	-1																																					
-1	2	-1																																					
-1	-1	2																																					

**Figura 2.6.** Máscaras para detectar líneas con inclinaciones (Martín, 2002).

#### d) Detección de bordes

**Bordes preliminares.** Es el procedimiento más empleado para detectar discontinuidades. Un borde es la frontera entre 2 regiones con nivel de gris relativamente diferente. Un detector de bordes se basa en el cálculo de un operador local de derivación. En las imágenes reales los bordes están ligeramente desenfocados. El valor de una primer derivada nos sirve para detectar la presencia de bordes, mientras que el signo de la segunda derivada nos indica si el pixel pertenece a una zona clara (negativo), oscura (positivo) o es el borde (cero). Para hacerlo en 2 dimensiones se debe definir el perfil en la dirección perpendicular a la del borde. La primera derivada viene de la magnitud del gradiente y la segunda viene por el operador Laplaciano.

Mediante un gradiente. El gradiente de una imagen  $\nabla f$  en la posición  $(x,y)$  viene dado por la Ecuación (2.3) (Martín, 2002), cuya magnitud se puede obtener mediante la Ecuación (2.4) (Martín, 2002). El gradiente es un vector que siempre apunta en la dirección de la máxima variación de la imagen. La dirección del gradiente se obtiene mediante la Ecuación (2.5) (Martín, 2002).

$$\nabla f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} \quad (2.3)$$

$$\|\nabla f(x,y)\| = \sqrt{\left(\frac{\partial f(x,y)}{\partial x}\right)^2 + \left(\frac{\partial f(x,y)}{\partial y}\right)^2} \quad (2.4)$$

$$\alpha(x,y) = \tan^{-1} \frac{\frac{\partial f(x,y)}{\partial x}}{\frac{\partial f(x,y)}{\partial y}} \tag{2.5}$$

Algunos operadores de derivación se observan en la Figura 2.7. Nota: la suma de los coeficientes de la máscara debe ser nula.

Mediante un Laplaciano. Este método juega un papel secundario. Es la derivada de orden dos definida por la Ecuación (2.6) (Martín, 2002).

$$\nabla^2 f(x,y) = \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2} \tag{2.6}$$

El coeficiente asociado al pixel central debe ser positivo y los demás valores negativos o cero, como se ve en la Figura 2.8.

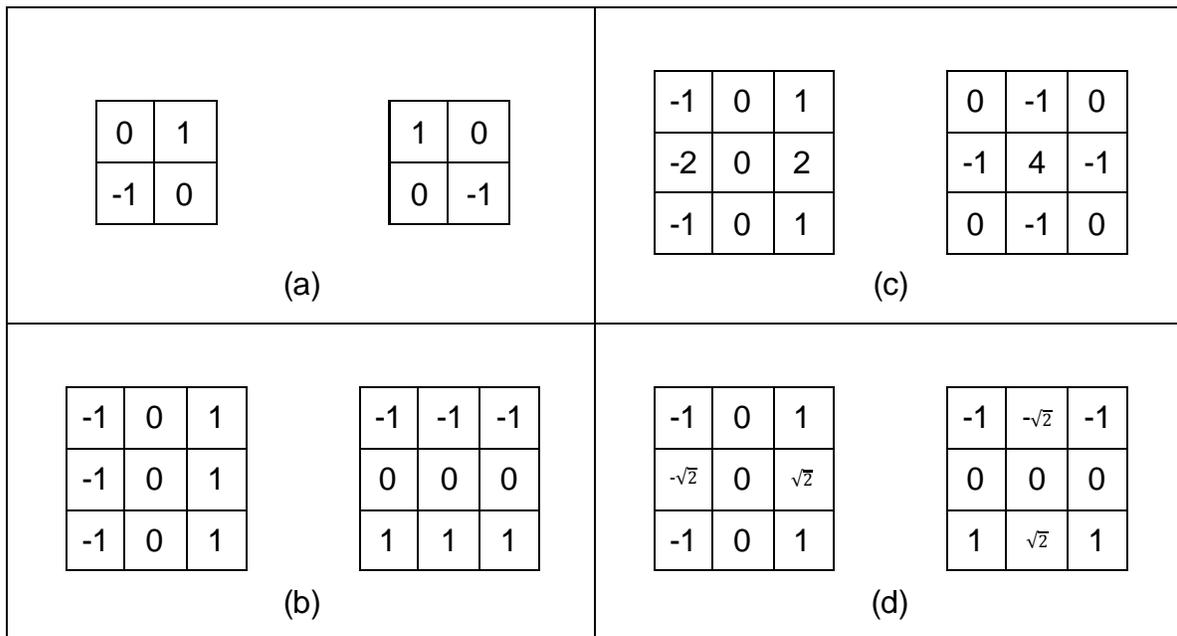


Figura 2.7. Operadores de derivación de Roberts (a), de Prewitt (b), de Sobel (c) y de Frei-Chen (d) (Martín, 2002).

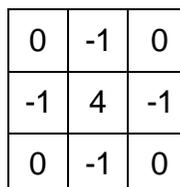


Figura 2.8. Máscara para segunda derivación (Martín, 2002).

**e) Enlazado de bordes.**

A veces los bordes tienen discontinuidad debido al ruido u otros factores de la imagen. Procesado local.

1. Analizar las características de los píxeles en un vecindario pequeño (3x3) alrededor de cada punto  $(x,y)$  donde se ha detectado un borde. Los puntos similares se enlazan, dando lugar a un contorno de píxeles que comparten ciertas propiedades en común.
2. Se utilizan la magnitud del operador de gradiente que ha dado lugar al píxel del borde y la dirección del gradiente.
3. Se dice que un píxel de borde  $(x',y')$  es similar a un píxel  $(x,y)$  si  $|\nabla f(x,y) - \nabla f(x',y')| \leq T$ .  $T$  es un umbral de amplitud.
4. El ángulo de un píxel de borde  $(x',y')$  va a ser similar al ángulo de un píxel  $(x,y)$  si  $|\alpha(x,y) - \alpha(x',y')| \leq A$ .  $A$  es un umbral angular.

**f) Técnicas de umbrales.**

Si una imagen  $f(x,y)$  está compuesta de objetos claros sobre un fondo oscuro de tal forma que los niveles de gris están agrupados en dos modos predominantes, se pueden separar los objetos del fondo mediante un umbral  $T$ , por lo tanto, cualquier píxel que cumpla con  $f(x,y) > T$  es etiquetado como objeto. De igual manera se puede aplicar en el caso que existan objetos agrupados en más de un nivel de gris. Un umbral puede depender del nivel de gris, de alguna propiedad local del píxel, de las coordenadas del píxel, o de una combinación.

**g) Segmentación orientada a regiones.**

Cuando se segmenta una imagen en regiones todos los píxeles deben pertenecer a una y solamente una región. El crecimiento de regiones es un procedimiento mediante el cual se agrupan píxeles o subregiones mayores. A partir de píxeles semilla se agregan píxeles vecinos que tienen propiedades similares. Una propiedad puede ser el nivel de gris. Se debe seleccionar la semilla y las propiedades para ir añadiendo píxeles durante el proceso de crecimiento. En una imagen infrarroja los píxeles claros son una elección natural para las semillas.

Para la extracción de objetos, Ng et al. (2001a) proponen aplicar el operador de Gradiente Cruz de Roberts, dado por la Ecuación (2.7) (Martín, 2002).

$$\nabla f = |f(x,y) - (f(x+1,y+1))| + |f(x+1,y) - f(x,y+1)| \quad (2.7)$$

Luego mejorar la imagen mediante alargamiento de contraste, seleccionar los umbrales necesarios, utilizar un filtro de mediana para eliminar el ruido, y finalmente utilizar la Ecuación (2.8) (Martín, 2002) para seleccionar los pixeles que nos interesen.

$$\left\{ \begin{array}{l} f(x,y) = f(x,y), \quad \text{si está dentro de los límites} \\ f(x,y) = 255, \quad \text{si no está dentro de los límites} \\ f(x,y) = \begin{cases} f(x,y), f \geq T \\ 255, f < T, \end{cases} \quad \text{si } f(x,y) \text{ está dentro de los límites.} \end{array} \right. \quad (2.8)$$

## 2.3 Análisis de los datos.

En esta sección se abordan los temas principales para el análisis de datos de una sola variable (estadística inferencial) y multivariable.

### 2.3.1 Estadística inferencial.

En su obra, Montgomery (2005) nos indica que el objetivo de la inferencia estadística es obtener conclusiones acerca de una población basándonos en una muestra de la misma. La estadística inferencial ha sido empleada como una herramienta muy poderosa y confiable para encontrar posibles causas de variación de los sistemas.

En sus estudios sobre control estadístico del proceso, Montgomery (2005) nos indica que todo sistema cuenta con una variabilidad inherente a los procesos naturales e inevitables del mismo, pero que esta suele ser acotada y no afectan en gran medida. También afirma que ocasionalmente se pueden presentar otras causas que afectarán de una manera más drástica esa variabilidad, la cual puede deberse a la presencia de una causa especial. El método para detectar que existen estas causas involucra un estudio probabilístico de las variables de interés. Para el caso de la variación de la temperatura puede resultar importante encontrar cuánta variación se debe a los procesos y metabolismo natural del seno.

### a) Pruebas de hipótesis.

En muchos problemas de la ingeniería, según dicen Montgomery y Runger (2003), se requiere que tomemos una decisión sobre aceptar o no alguna hipótesis de algún parámetro en una población de datos. Este procedimiento es conocido como prueba de hipótesis y es muy útil en el caso de experimentos comparativos. La definición formal de hipótesis, según Montgomery y Runger (2003), es “una afirmación acerca de los parámetros de una o más poblaciones”. Ya que se utilizan distribuciones de probabilidad para representar las poblaciones, la hipótesis puede ser acerca de la distribución de probabilidad de la variable aleatoria.

Una hipótesis nula ( $H_0$ ) es aquella a la que someteremos a la prueba. Rechazar esta hipótesis siempre nos lleva a aceptar una hipótesis alternativa ( $H_1$ ). Un error del tipo I consiste en rechazar una  $H_0$  cuando es verdadera. Un error del tipo II consiste en no rechazar una  $H_0$  cuando es falsa. La potencia de una prueba se define como rechazar una hipótesis nula cuando en realidad esta es falsa. En la Tabla 2.4 se muestran las cuatro situaciones existentes al determinar el estado de la decisión final.

Decisión	$H_0$ es Verdadera	$H_0$ es Falsa
No rechazar $H_0$	No error	Error tipo II
Rechazar $H_0$	Error tipo I	No error

**Tabla 2.4.** Situaciones que pueden ocurrir al decidir sobre  $H_0$  (Montgomery y Runger, 2003).

A la probabilidad de cometer el error tipo I se le conoce como nivel de significancia y se denota por  $\alpha$ .

### b) Distribuciones de probabilidad y prueba de bondad de ajuste.

Siguiendo con la obra de Montgomery y Runger (2003), podemos ver que un caso especial de las pruebas de hipótesis se aplica 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 ( $\chi^2$ ). Este estadístico compara las observaciones obtenidas en una serie de  $k$  intervalos en una muestra dada contra las observaciones esperadas en la misma serie para una distribución de probabilidad

dada. El cálculo del estadístico se hace mediante la Ecuación (2.9), donde  $O_i$  representa las frecuencias observadas y  $E_i$  las frecuencias esperadas.

$$\chi_0^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (2.9)$$

Luego este estadístico se compara contra el nivel crítico de la Chi-Cuadrada  $\chi_{\alpha, k-p-1}^2$ , donde  $\alpha$  representa el nivel de significancia,  $k$  el número de intervalos y  $p$  el número de parámetros estimados. Si el valor de  $\chi_0^2 < \chi_{\alpha, k-p-1}^2$ , entonces no se rechaza  $H_0$ .

### 2.3.2 Análisis de datos multivariantes.

De acuerdo con Peña (2002), describir cualquier situación real requiere tener en cuenta simultáneamente varias variables. El análisis de datos multivariantes tiene por objeto el estudio estadístico de varias variables medidas en los elementos de una población. Entre los objetivos del análisis de datos multivariantes se encuentran: resumir el conjunto de variables; encontrar grupos en los datos, si existen; clasificar nuevas observaciones; y relacionar dos conjuntos de variables.

#### a) Datos multivariantes.

La información para estos estudios consiste en un conjunto de  $p$  variables obtenido de  $n$  elementos muestrales. Esta información se presenta generalmente en una matriz de datos  $\mathbf{X}$  de dimensiones  $n \times p$ , donde cada columna representa una variable escalar (univariante) tomada de cada elemento de la muestra, y cada renglón representa el vector de variables tomado de un solo elemento de la muestra, tal como se presenta en la Figura 2.9 (Peña, 2002).

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$

**Figura 2.9.** Matriz de datos.

#### b) El vector de medias.

De acuerdo con Peña (2002), se utiliza el vector de medias como medida de centralización para describir los datos multivariantes. Este es un vector de dimensión

$p$  que contiene las medias de cada variable, tal como se observa en la Ecuación (2.10).

$$\bar{x} = \frac{1}{n} \sum x_i = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix} \quad (2.10)$$

El vector de medias se encuentra en el centro de los datos, tal que  $\sum(x_i - \bar{x}) = 0$ .

**c) La matriz de varianzas y covarianzas.**

La variabilidad de una variable con respecto a su media se mide con su varianza. Así mismo, la relación lineal entre dos variables se mide por su covarianza. Cuando se trabaja con datos multivariados, se pueden presentar por medio de la matriz de varianzas y covarianzas  $S$ , definida por la Ecuación (2.11).

$$S = \frac{1}{n} \sum (x_i - \bar{x})(x_i - \bar{x})' \quad (2.11)$$

La matriz  $S$  contiene en su diagonal principal las varianzas y fuera de la diagonal las covarianzas entre las variables (Peña, 2002).

**d) La matriz de precisión.**

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. Dentro de cada fila, sin contar la diagonal principal, se encuentran los términos proporcionales a los coeficientes de regresión múltiple de esa variable explicada por todos los demás. Si se llaman  $s^{ij}$  a los elementos de la matriz de precisión, tenemos la Ecuación (2.12).

$$s^{ij} = -\frac{\hat{\beta}_{ij}}{s_r^2(i)} \rightarrow \hat{\beta}_{ij} = -s^{ij}s_r^2(i) \quad (2.12)$$

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).

**e) La normal k-dimensional.**

De acuerdo a lo establecido en la obra de Peña (2002), un vector  $x$  sigue una distribución normal  $p$ -dimensional si su función de densidad es aquella que aparece en la Ecuación (2.13).

$$f(x) = |\mathbf{V}|^{-1/2} (2\pi)^{-p/2} \exp\left\{-\frac{1}{2}(x - \mu)' \mathbf{V}^{-1}(x - \mu)\right\} \quad (2.13)$$

Sus propiedades principales son:

- ✓ La distribución es simétrica alrededor de  $\mu$ .
- ✓ La distribución tiene un único máximo en  $\mu$ .
- ✓ La media del vector aleatorio normal es  $\mu$  y su matriz de varianzas y covarianzas es  $\mathbf{V}$ .
- ✓ Si  $p$  variables aleatorias tienen distribución conjunta normal y están incorreladas, son independientes.
- ✓ Cualquier vector  $\mathbf{x}$  normal  $p$ -dimensional con matriz  $\mathbf{V}$  no singular se puede convertir en un vector  $\mathbf{z}$   $p$ -dimensional con un vector de medias 0 y matriz de varianzas y covarianzas igual a la matriz identidad.
- ✓ Sus distribuciones marginales son normales.
- ✓ Si se cortara con hiperplanos paralelos al definido por las variables, se obtienen curvas de nivel definidas por la Ecuación (2.14).

$$D^2 = (\mathbf{x} - \mu)' \mathbf{V}^{-1} (\mathbf{x} - \mu) \quad (2.14)$$

- ✓ La variable  $D^2$ , conocida como distancia de Mahalanobis, se distribuye como una  $\chi^2$  con  $p$  grados de libertad.

#### f) Muestreo.

El muestreo, según Pérez López (2005), es un conjunto de técnicas estadísticas que estudian la forma de seleccionar una muestra lo suficientemente representativa de una población cuya información permita inferir sobre los parámetros de la misma.

El muestreo aleatorio simple es un procedimiento de selección de muestras, donde cada elemento tiene la misma probabilidad de ser seleccionado. Consiste en obtener la muestra unidad a unidad de forma aleatoria, sin reposición (sin que el elemento se repita), a la población de las unidades previamente seleccionadas.

En el caso de las distribuciones multivariadas, Hair Jr. et al. (2009) recomiendan utilizar un tamaño de muestra tal que, en combinación con el *tamaño del efecto* (considerado en la obra como la correlación entre las variables existentes) y el nivel de significancia  $\alpha$ , logre una potencia relativamente alta (por lo menos 80%). De acuerdo con su obra, con un tamaño de muestra 100 o superior se puede lograr una potencia aceptable para un tamaño de efecto moderado (0.5 o superior).

## **2.4 Software.**

Un análisis manual de una serie de estudios médicos puede ser ineficiente y propenso a error humano (FOK et al., 2002). Por lo tanto, en situaciones reales suele ser recomendable utilizar software que permita automatizar una serie de procesos de análisis y realizarlos de una manera más rápida, reduciendo de esta manera el error humano.

### **2.4.1 Software de asistencia para la detección.**

A continuación se mencionan algunos de los métodos en los cuales es utilizado el software para asistir a la detección de enfermedades.

#### **a) Diagnóstico asistido por computadora.**

El diagnóstico asistido por computadora (CAD, por sus siglas en inglés) ayuda a los médicos a la parte de la interpretación de las imágenes adquiridas. Las técnicas más usadas son, según Jiang et al. (2005):

- Redes neuronales artificiales (RNA).
- Redes bayesianas.
- Análisis de discriminación lineal.
- Lógica difusa.
- Fusión de Lógica difusa y RNA.

Utilizando CAD, se ha logrado obtener una precisión general cercana al 90% (Moghbel y Mashohor, 2011). Con el desarrollo de RNA y la introducción de sistemas basados en Lógica Difusa para clasificación de imágenes y la disponibilidad de utilizar cámaras de segunda generación, los sistemas termográficos CAD recibieron una nueva ola de interés por parte de investigadores. Los mejores diseños de CAD desde 1977 hasta 2008 se basan en RNA y Lógica difusa (Moghbel y Mashohor, 2011).

#### **b) Sistemas de visión computarizada.**

Un sistema de visión computarizada (SVC) es un sistema de cómputo que utiliza uno o más dispositivos de captura de imagen para proveer datos de imágenes para

análisis e interpretación. Una de las ventajas de los SVC es que tienen la habilidad de cubrir todos los rangos del espectro electromagnético (Megahed et al., 2011).

**c) Imagenología térmica computarizada.**

Imagenología térmica computarizada (ITC) se usa para estudiar la termografía con reconstrucción digital de las imágenes. Es un sistema de detección de calor y de procesamiento que usa las cámaras termo-sensibles para capturar imágenes digitales basadas en la radiación del cuerpo. Una vez capturadas, las imágenes son analizadas por un algoritmo de computadora y mostrada para interpretación (Köşüş et al., 2010). ITC analiza los datos usando algoritmos para correlacionar la información infrarroja del seno que es examinado en patrones infrarrojos que son asociados con tejido mamario benigno o maligno. El ITC puede llegar a ser una herramienta complementaria de diagnóstico a la mamografía tradicional (Köşüş et al., 2010).

**d) Redes Neuronales Artificiales.**

El cerebro humano está compuesto de neuronas. Un conjunto de estas puede llevar a cabo tareas simultáneas; esto se conoce como red neuronal (Ng y Kee, 2008).

Una RNA es un modelo bio-fisiológico del cerebro humano que intenta emular sus procesos operacionales (Moghbel y Mashohor, 2011).

Para la detección de defectos en la industria, se han aplicado RNA en las que primero se computan las imágenes luego procede un análisis de la misma con valores usados como entrada a la red y se obtiene una nota de “defecto” o “no defecto” como salida (Ibarra-Castaneda et al., 2004).

## **2.5 Análisis de estudios previos.**

- En sus estudios, Ng et al. (2001b) realizaron 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. En sus estudios se encuentran con diversos resultados, como intervalos de confianza para las temperaturas de ambas mamas, pero sin duda lo más importante es que afirman que las diferencias entre el seno derecho y el izquierdo

entre personas sanas, con tumores benignos o malignos, es insignificante. Se estima que estos autores manejaron el promedio de todo el seno, por lo que consideramos importante una etapa previa de segmentación, para que el promedio se vea realmente afectado por la temperatura elevada.

- En su estudio, Ng et al. (2002), muestran la utilización de redes neuronales artificiales. Se obtiene una red donde sus resultados son correctos en un 61.54% de las ocasiones. Esta red maneja como entradas la media, mediana y moda de la temperatura, así como la edad, historial familiar, edad de menarquía, presencia de bulto palpable, biopsia o cirugía previa, presencia de descarga del pezón, dolor de seno, menopausia a la edad menor a los 50, primer hijo a la edad menor de los 30.

- Otro estudio es el de Tan et al. (2009), donde mediante la aplicación de redes neuronales logran la clasificación de diferentes diagnósticos (normal, benigno, cáncer o sospechoso) con más del 80% de precisión.

- En su obra, Moghbel y Mashohor (2011) hacen 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, Ng y Kee (2008) hablan de una propuesta de aplicación de redes neuronales artificiales y métodos bio-estadísticos, como utilizar regresión lineal para seleccionar las variables más importantes al momento de diseñar la red neuronal que decida si la persona está enferma o no. 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 nos 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.

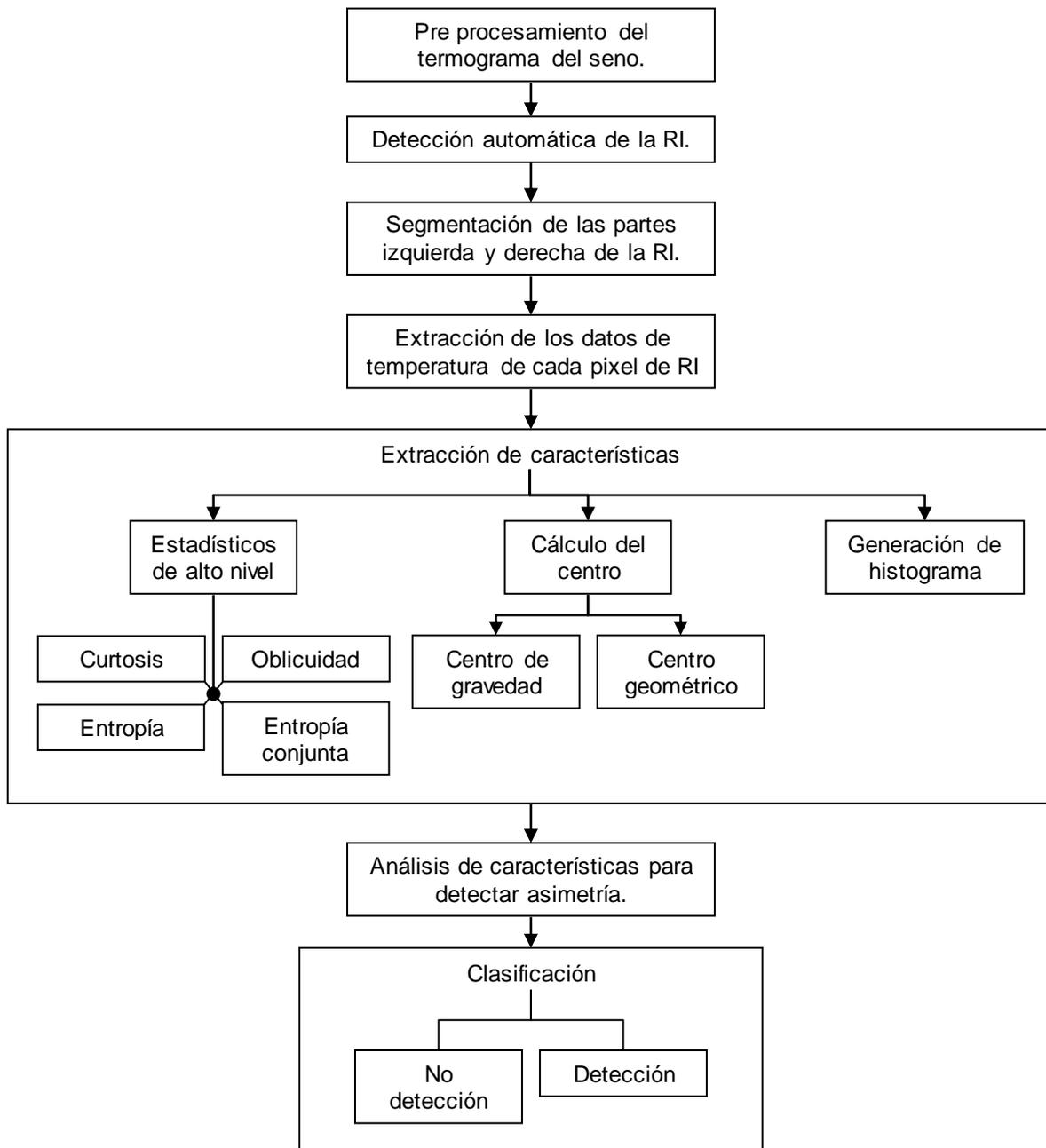
- En su investigación, FOK et al. (2002) utilizaron una combinación de termogramas y redes neuronales artificiales en las cuales se podían aplicar técnicas de procesamiento de imágenes antes de ingresarlas a la red, y una segunda neurona

en la cual se utilizaban datos fisiológicos en lugar de termogramas. Ellos encontraron que una 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) desarrollan varios algoritmos que permiten minimizar los resultados no reales en los termogramas.

- El estudio de Amri et al. (2011) se enfocó a investigar numéricamente 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. El incremento de temperatura fue significativo para tumores localizados en una profundidad menor o igual a 20mm.

- En su investigación, Kapoor et al. (2012) mencionan que el análisis de los termogramas requieren 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. En dicha metodología, mostrada en la Figura 2.10, se habla de una serie de acciones de pre procesamiento de los termogramas, además del desarrollo de un sistema automatizado para la detección de las regiones de interés (RI), segmentación de la imagen, luego obtiene la información de cada pixel y recopila los datos estadísticos necesarios. A partir de esto, lleva a cabo un análisis estadístico para obtener nueva información de los datos y detectar la asimetría térmica en el seno de las personas. Basado en eso, hace una distinción entre un paciente “normal” y uno “anormal”.

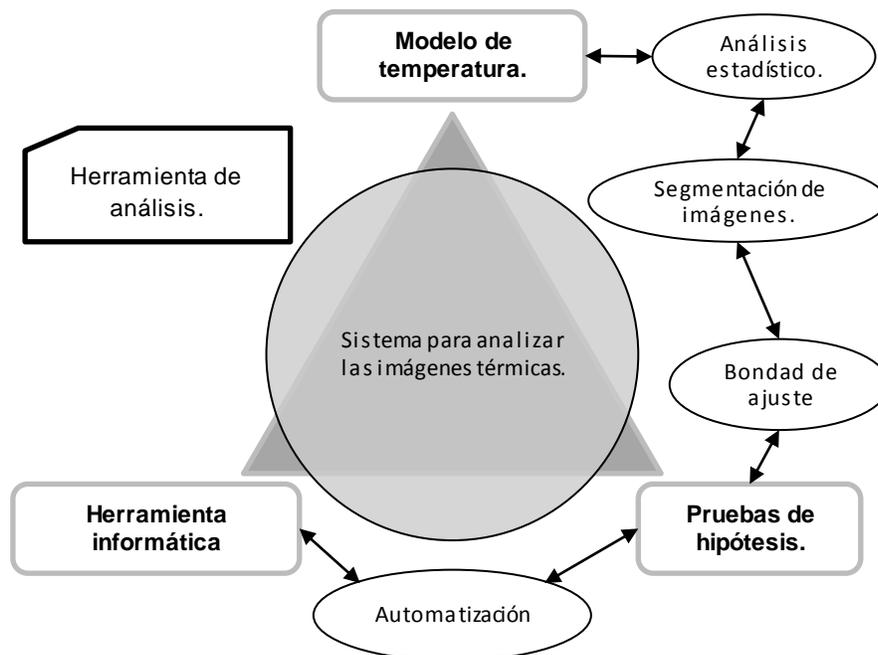


**Figura 2.10.** Metodología propuesta por Kapoor et al. (2012) para un análisis de los termogramas.

De los casos anteriores se puede observar un factor común en lo relacionado con la utilización del software para facilitar el procesamiento y análisis de los datos. Dependiendo del enfoque, algunos autores utilizan software más especializado que otro. Los elementos mencionados en el marco teórico, serán necesarios para desarrollar la metodología que se presenta en el siguiente capítulo.

### 3. METODOLOGÍA.

El enfoque que se propone, mostrado en la Figura 3.1, surge a partir del análisis hecho en el Capítulo 2. Este es multidisciplinario, ya que explota técnicas de análisis estadístico, procesamiento de imágenes y desarrollo de herramientas informáticas para el problema de distinguir, mediante la temperatura, entre un seno sano y un seno potencialmente enfermo.



**Figura 3.1.** Relaciones entre técnicas de la Propuesta de solución.

Dentro de la parte de estadística inferencial, sirvió de inspiración percatarse que existe una relación entre falso negativo (FN), falso positivo (FP), verdadero negativo (VN) y verdadero positivo (VP) y los errores tipo I y II, cuando la  $H_0$  consiste en declarar un paciente sano, como se puede apreciar en la Tabla 3.1.

Decisión	$H_0$ es Verdadera (Es persona sana)	$H_0$ es Falsa (Es persona no sana)
	No rechazar $H_0$ (Declarar persona Sana)	No error (VN)
Rechazar $H_0$ (Declarar persona No Sana)	Error tipo I (FP)	No error (VP)

**Tabla 3.1.** Correspondencia entre errores estadísticos y FN, FP, VN y VP.

Por lo tanto se podrían realizar pruebas de hipótesis sobre la temperatura del seno y determinar si una persona tiene cáncer de mama.

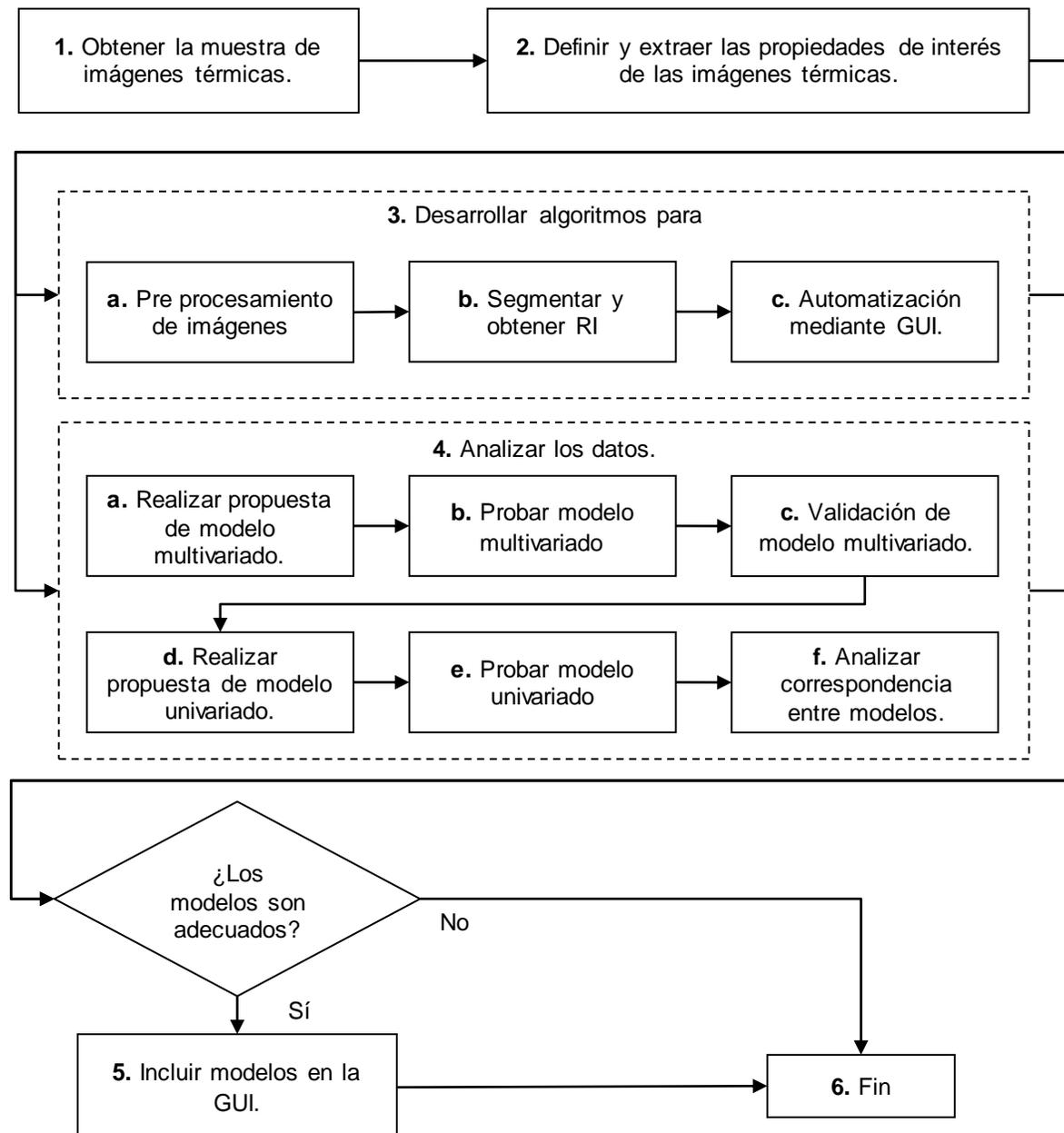
Dentro del área de la estadística multivariada, se puede considerar a la temperatura máxima de un seno y la temperatura umbral a partir de la cual se comienza a detectar la anomalía como un solo conjunto de datos multivariados. De tal manera que se intentará encontrar un modelo que se ajuste a su comportamiento y que nos permita predecir la temperatura umbral a la cual se detectaría la anomalía para casos en los que no se tiene información anterior.

Siguiendo con la estadística inferencial, se puede aplicar una prueba de bondad de ajuste a la diferencia de ambas temperaturas para obtener más información y soportar el modelo multivariado.

Por otro lado, utilizar técnicas de procesamiento de imágenes ayudará a estandarizar los procesos de detección de las regiones de interés de las imágenes térmicas, así como la extracción de sus datos estadísticos y facilitarán los procesos de prueba de hipótesis. Las técnicas de procesamiento de imágenes pueden incluir técnicas de mejora de contraste que permitan diferenciar las regiones de interés.

Una herramienta informática puede ser muy útil para personas que no estén familiarizadas con los paquetes de procesamiento de imágenes. Se pueden automatizar una serie de operaciones de procesamiento de tal manera que no se requiera mucho conocimiento para realizarlos. Además, como mencionan en su obra FOK et al. (2002), pueden ayudar a reducir la incidencia de errores.

Teniendo en cuenta lo anterior se ha llegado a la decisión de aplicar la metodología que se muestra en la Figura 3.2, la cual, se puede definir como una investigación correlacional, puesto que este tipo de estudios tienen por objetivo saber cómo se puede comportar un concepto o variable conociendo el comportamiento de otra u otras variables relacionadas (Hernández Sampieri et al., 2006).



**Figura 3.2.** Metodología propuesta.

A continuación se describen los puntos de la metodología.

1. **Obtener la muestra imágenes térmicas** de casos plenamente identificados de personas con cáncer de mama, acompañados con su respectivo diagnóstico en el cual se indica la región de interés del tumor cancerígeno. Estos datos serán proporcionados por parte del equipo de investigación del DIFUS. De acuerdo con Hair Jr. et al. (2009) se requiere un tamaño de

muestra grande para cualquier análisis de tipo multivariado. Se utilizará una muestra de tamaño 100 o superior, dependiendo de los datos que sean brindados por el DIFUS. En este caso, cada imagen representa un elemento de la muestra.

2. **Definir y extraer las propiedades de interés de las imágenes térmicas.** En este caso nos interesan dos propiedades de cada imagen: la temperatura máxima del seno donde se encuentra el tumor (es decir, la temperatura máxima del tumor) y la temperatura umbral a la cual se detecta el tumor identificado por el oncólogo.

**Nota:** Los pasos 3 y 4 se pueden hacer de manera paralela.

3. **Desarrollar algoritmos para analizar las imágenes.**

- a. Realizar un algoritmo para pre procesamiento de las imágenes, de tal manera que sea más sencillo detectar las regiones de interés con otro algoritmo.
- b. Desarrollar un algoritmo de procesamiento para separar y obtener las RI de los senos.
- c. Automatizar dichos algoritmos por medio de una Interfaz Gráfica para Usuarios (GUI) que sea fácil de utilizar por el equipo del DIFUS.

4. **Analizar los datos.**

- a. Realizar propuesta de un modelo multivariado. Debido a que se utilizan varias propiedades de cada imagen, se intentará encontrar si estas se ajustan a alguna distribución multivariada, específicamente a una distribución normal bivariada.
- b. Probar el modelo propuesto en el paso anterior mediante las curvas de nivel descritas en el apartado 2.3.2, inciso e), correspondiente a la normal k-dimensional, las cuales se distribuyen como una  $\chi^2$  con  $p$  grados de libertad.
- c. Validar modelo propuesto. Se tomarán imágenes al azar que no hayan sido sujetas a ningún análisis, y por medio de ellas se confirmará que el modelo sea correcto.

- d. Realizar propuesta de modelo univariado. Se medirá la diferencia entre las temperaturas máximas y mínimas (umbrales de detección) del tumor y se planteará un modelo de distribución de probabilidad univariada para esta diferencia.
  - e. Probar modelo univariado. Se hará una prueba de bondad de ajuste a dicho modelo para comprobar si la propuesta fue correcta.
  - f. Analizar si existe correspondencia entre el modelo univariado y el multivariado.
5. Si los modelos son adecuados, incluirlos en la GUI para uso futuro en otras imágenes.
  6. Documentar los resultados obtenidos.

## 4. IMPLEMENTACIÓN DE LA METODOLOGÍA

Durante este capítulo se irán detallando las acciones que se llevaron a cabo para la implementación de la metodología expuesta en el capítulo anterior.

Durante la implementación, se notó que el software que adquiere y lee las imágenes térmicas tiene capacidad para exportarlas en varios formatos, siendo los que más nos interesan: En escala de grises, en escala de colores RGB y en escala de colores IRON, siendo esta última una gama que va desde colores púrpuras para las temperaturas bajas, hasta bronce y blanco para temperaturas altas. Además, permite exportar en su totalidad una matriz con la temperatura obtenida en cada pixel en formato Comma Separated Values (CSV) que Excel y MATLAB® pueden interpretar sin dificultad. Esto nos facilita la etapa de recolección de los datos. Se detalla cómo se obtuvieron los resultados de la aplicación; estos serán expuestos en el Capítulo 5.

### 4.1 Obtener la muestra de imágenes térmicas.

El equipo de investigación del DIFUS se ha puesto en contacto con múltiples centros dedicados a la oncología y otros estudios sobre el cáncer, especialmente sobre el cáncer de mama. Una de las personas con las que se hizo contacto, el oncólogo Martín del Campo, ha trabajado durante años en el estudio del cáncer de mama obteniendo imágenes térmicas de personas a las cuales se les ha comprobado cáncer por medio de biopsia, por lo que se tiene una certeza del 100% en el diagnóstico.

El oncólogo ha trabajado con un software especializado para la obtención de las imágenes, de la marca FLIR, que es igual a la utilizada por el equipo del DIFUS; de tal manera que las imágenes del doctor tienen el mismo formato (\*.fff) que aquellas que se obtendrán por parte del DIFUS. Al momento, el doctor Martín del Campo nos ha permitido trabajar con alrededor de 180 imágenes obtenidas, de las cuales 157 venían acompañadas de un documento de Word en el cual aparecía explícitamente su región de interés. De estas, se dejaron por fuera 23 imágenes que

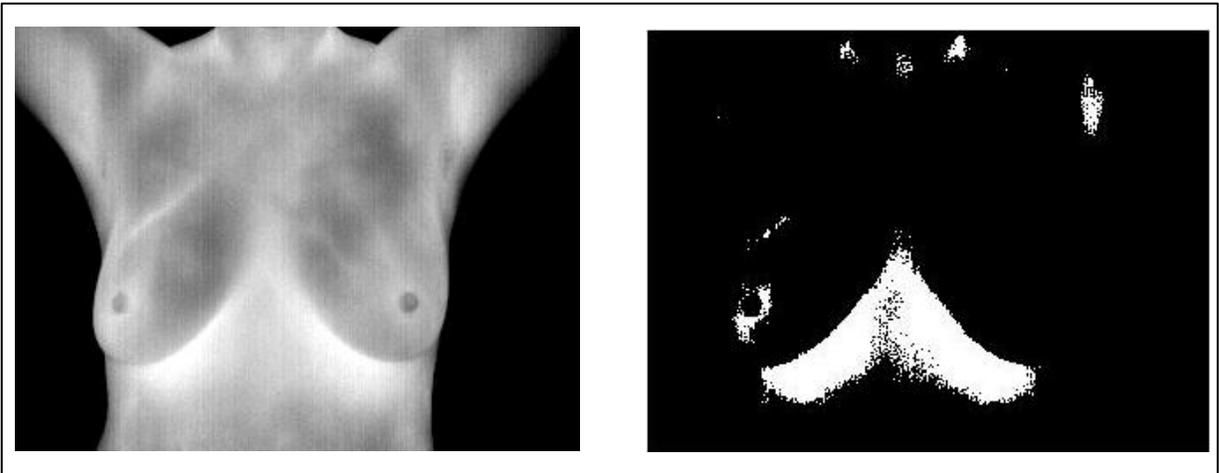
nos permitirán corroborar nuestros resultados, de las cuales se descartaron 4 que no venían acompañadas del diagnóstico completo, quedando 19 imágenes para propósitos de validación. Del resto, 4 pacientes tenían el cáncer en una etapa muy desarrollada, y ya que este trabajo busca proveer herramientas para un método de detección temprano, fueron descartadas. Las 130 imágenes restantes serán utilizadas para los análisis estadísticos correspondientes.

## 4.2 Definir y extraer las propiedades de interés de las imágenes térmicas.

A partir de la muestra del doctor Martín del Campo, se seleccionaron 130 imágenes de las cuales se extrajeron 2 características importantes: La temperatura máxima en el seno, y el umbral a partir de cual se detecta una anomalía térmica. Para obtener la temperatura máxima en el seno, se utilizó el software de FLIR y se seleccionó manualmente el área de cada seno y se pidió que mostrara su valor máximo. Para establecer la temperatura umbral, se utilizó el siguiente algoritmo:

```
CSV=csvread('Aquí va el archivo CSV deseado');
for i=1:240
for j=1:320
if CSV(i,j)>'Temperatura umbral'
Z(i,j)=1;
else
Z(i,j)=0;
end
end
end
end
imshow(Z)
```

Ello nos permitía observar en una imagen binaria Z (donde 0 corresponde al color negro y 1 al color blanco) las temperaturas que se encuentran sobre el umbral. El valor 'Temperatura umbral' se ajusta hasta encontrar el valor en el cual fuera visible el área detectada por el doctor Martín del Campo y explicitada en el documento de Word que acompaña a la imagen, pero otras regiones dentro del seno permanecieran de color negro. Un ejemplo de una imagen binaria se puede observar en la Figura 4.1.



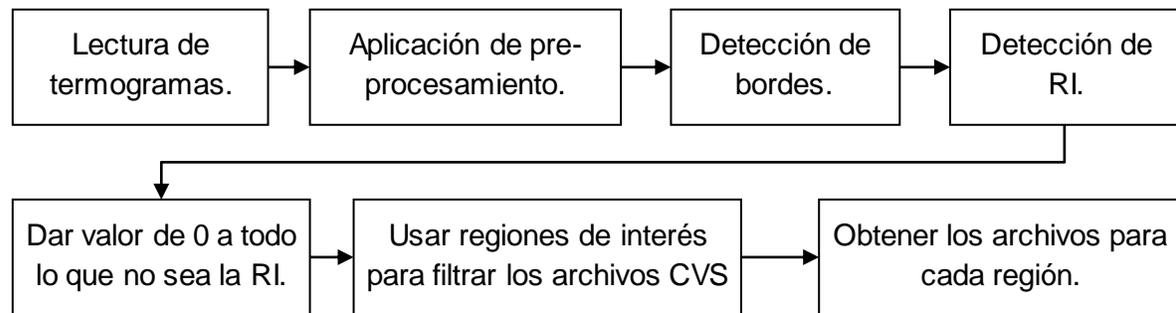
*Figura 4.1. Binarización de una imagen.*

### **4.3 Desarrollar algoritmo para analizar las imágenes.**

Este algoritmo nos debe arrojar como producto los datos de la temperatura de cada pixel para cada seno.

Una Interfaz gráfica de usuario (GUI, por sus siglas en inglés) es un despliegue gráfico, en una o más ventanas, que contiene controles, llamados componentes, los cuales permiten al usuario realizar tareas interactivas. El usuario de la GUI no necesita entender los detalles de cómo funciona su script. La mayoría de las GUI esperan a su usuario para manipular un control y responder con una acción. MATLAB® contiene un Ambiente de Desarrollo de GUI (GUIDE, por sus siglas en inglés) (Mathworks, 2013).

Un *callback* es una rutina escrita por el usuario que consiste en un código ejecutable de MATLAB®. Para expresarlo mejor, se realizó un diagrama de flujo que se muestra en la Figura 4.2.

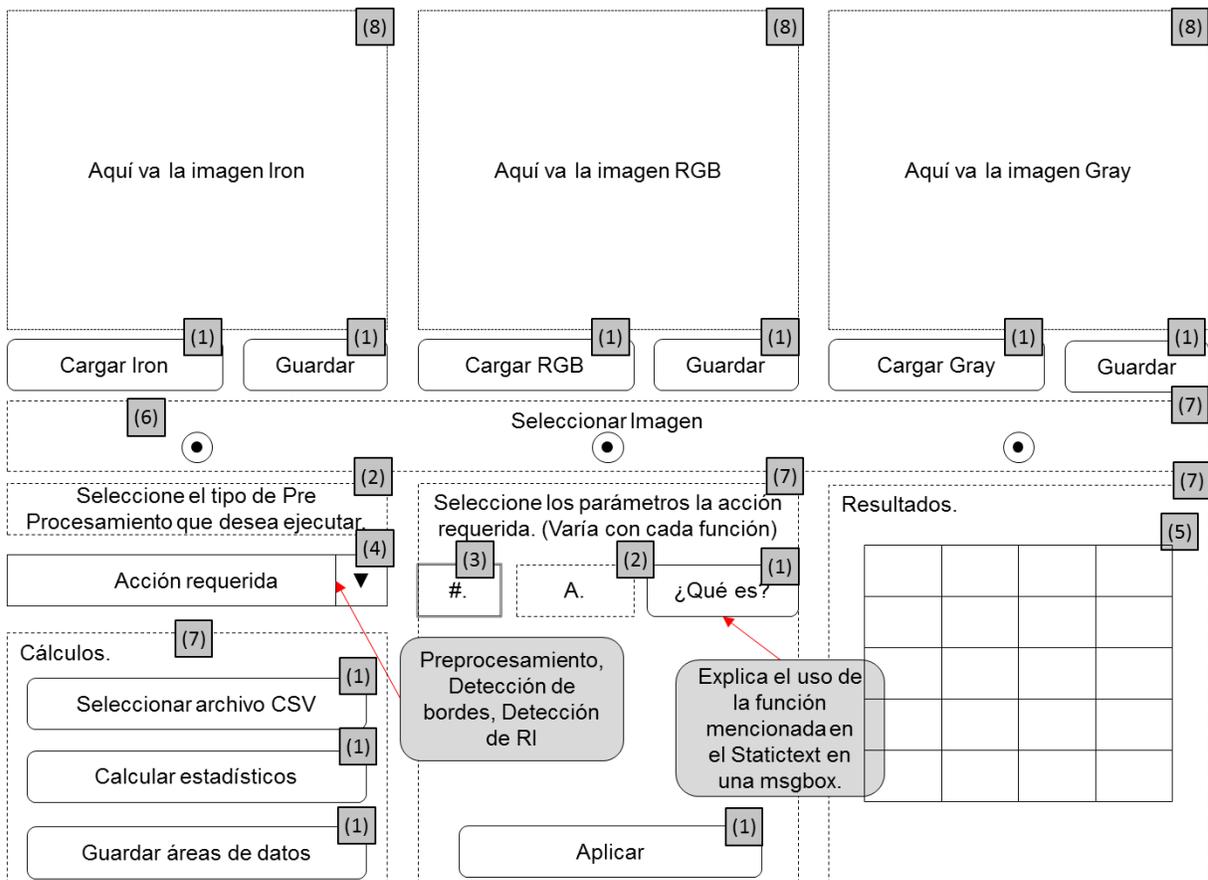


**Figura 4.2.** Diagrama de flujo de operaciones del algoritmo de análisis.

Los requisitos que debe tener esta GUI son los siguientes:

- Que permita leer la imagen del paciente.
- Que permita especificar ciertos tipos de procesamiento a las imágenes.
  - Permitir entre varios tipos de técnicas de pre procesamiento y sus respectivos parámetros.
- Que permita hacer selecciones con el mouse dentro de esas imágenes.
  - Existen varias funciones de MATLAB® para eso.
- Que permita definir el umbral que diferencie una zona de la otra.
  - Probarlo en las 3 imágenes, para ver cuál de los bordes es más representativo.
- Que permita filtrar los archivos SCV.
  - Extraer los datos en otro archivo nuevo.

El diseño de la GUI se muestra en la Figura 4.3, donde los cuadrados con (1) representan un Push Button, (2) Static Text, (3) Edit Text, (4) Pop-up Menu, (5) Table, (6) Radio Button (Button Group), (7) un Panel y (8) Axes.



**Figura 4.3.** Diseño de la GUI.

Un **Push Button** (PB) es aquél que genera una acción cuando es pulsado. Cuando el botón es pulsado, aparece presionado, pero una vez que se libera, el botón aparece alzado. Un **Radio Button** (RB) permite realizar una selección. Son mutuamente excluyentes, es decir, cuando se selecciona uno dentro de un grupo, el resto aparece como no seleccionado. Un **Button Group** (BG) permite trabajar con una serie de Radio Buttons. Un **Edit Text** (ET) es un campo que permite al usuario insertar o modificar cadenas de texto o números. Para el caso de los números, estos deben ser previamente convertidos a sus equivalentes numéricos. Un **Static Text** (ST) muestra líneas de texto no modificables. Un **Pop-up Menu** (PUM) muestra una serie de opciones cuando se pulsa su flecha. Un **Table** permite insertar un componente tipo tabla. Un **Axes** permite a los GUI mostrar elementos gráficos, tales

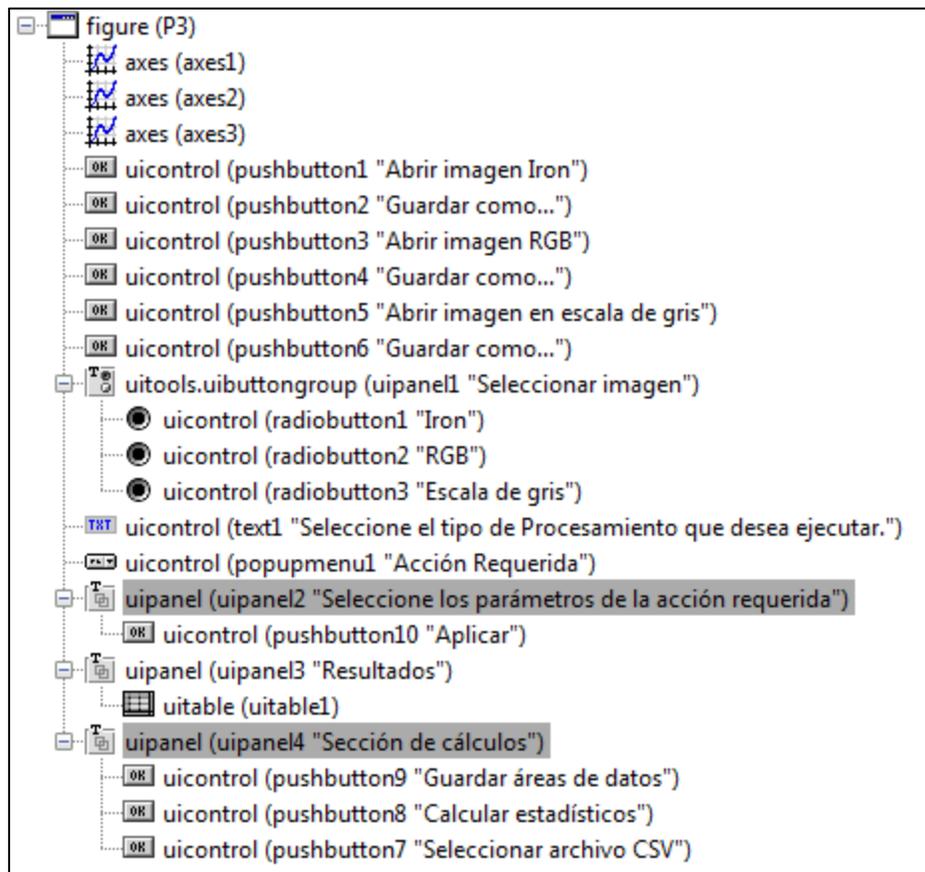
como imágenes y gráficas. Un **Panel** permite acomodar los componentes de la GUI en grupos, que sean visualmente agradables para el usuario y fáciles de entender.

Los objetos más importantes de este algoritmo se etiquetaron, para facilitar su identificación durante la programación, de acuerdo a la Tabla 4.1.

<b>Push Buttons</b>	
Etiqueta	Función
PB_Iron1	Carga la imagen con escala de colores Iron.
PB_Iron2	Guarda la imagen con escala de colores Iron modificada.
PB_RGB1	Carga la imagen con escala de colores RGB.
PB_RGB2	Guarda la imagen con escala de colores RGB modificada.
PB_Gr1	Carga la imagen con escala de gris.
PB_Gr2	Guarda la imagen con escala de gris modificada.
PB_SCV	Carga archivo SCV con datos de temperatura.
PB_Calc	Hace los cálculos de media, desviación estándar y la tabla de frecuencias. Los posiciona en "Tabla_Res".
PB_Save	Guarda un archivo con los datos separados por cada área.
PB_AplicarX	Aplica los cambios a las imágenes de acuerdo a los parámetros indicados para la acción seleccionada.
PB_QueEsX	Al dar clic, indica para qué sirve cada parámetro disponible en cada acción.
<b>Radio Buttons</b>	
RB_Iron	Selecciona la imagen con escala de colores Iron.
RB_RGB	Selecciona la imagen con escala de colores RGB.
RB_Gr	Selecciona la imagen con escala de gris.
<b>Edit Text</b>	
ET_X	Introduce el valor para el parámetro especificado.
<b>Pop-up Menu</b>	
POM_Pro	Selecciona el tipo de procesamiento a que se someterá la imagen.
<b>Table</b>	
Tabla_Res	Muestra los resultados obtenidos de "PB_Calc".
<b>Axes</b>	
A_Iron	Muestra la imagen escala Iron y sus modificaciones en tiempo real.
A_RGB	Muestra la imagen escala RGB y sus modificaciones en tiempo real.
A_Gr	Muestra la imagen escala gris y sus modificaciones en tiempo real.

**Tabla 4.1.** Etiquetas de los objetos.

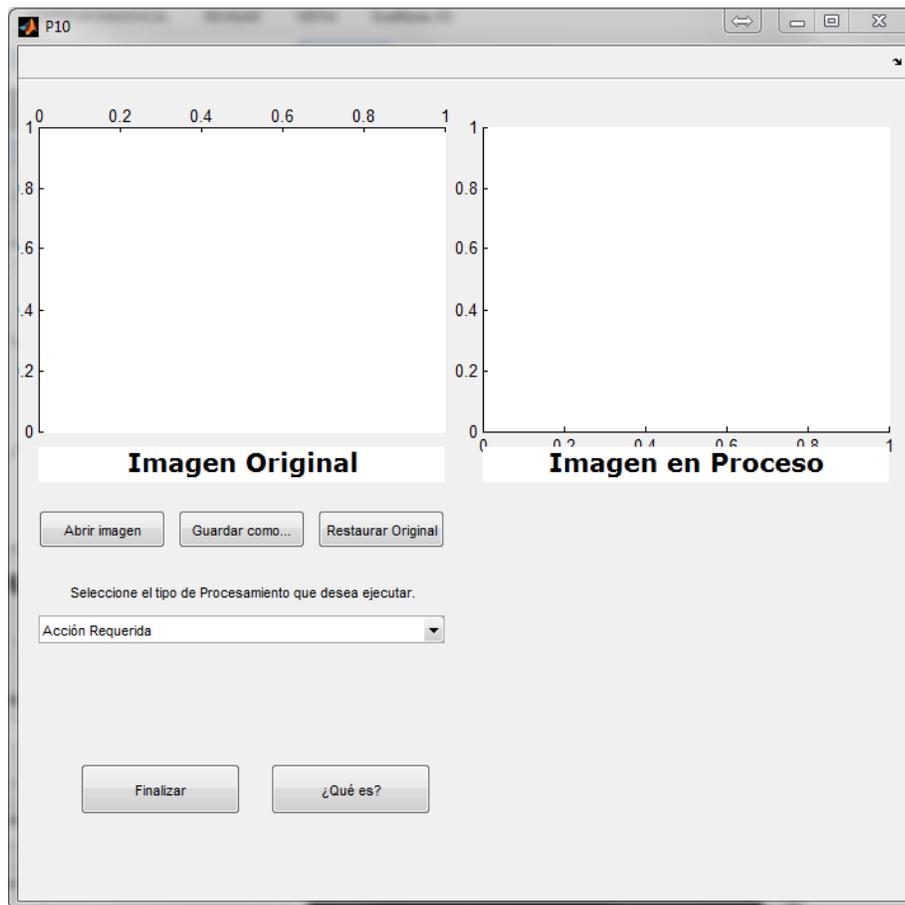
Cuando un objeto se encuentra dentro de otro, por ejemplo, un PB dentro de un Panel, el Panel se convierte en un objeto Padre del PB y el hijo en un objeto Hijo del Panel, dentro de la jerarquía entre los objetos. Dentro de este algoritmo, la jerarquía es la que se muestra en la Figura 4.4.



**Figura 4.4.** Jerarquía entre los objetos.

Dentro de las funciones de pre procesamiento, se encuentran las funciones `imadjust` e `histeq`. Los códigos se detallan en las siguientes secciones.

Después de una serie de ajustes que se dieron durante el desarrollo del presente trabajo, la forma final de la GUI se modificó de su boceto original, tal como se observa en la Figura 4.5. El código completo de la GUI se puede consultar en el Capítulo 8, Anexos.



**Figura 4.5.** Forma final de GUI

#### a) Lectura de termogramas.

En el PB con la leyenda “Abrir Imagen...” se ejecuta el siguiente algoritmo:

```
[GrName GrPath] = uigetfile('*.*jpg','Abrir imagen');
FullPathGr = strcat(GrPath,GrName);
axes(handles.A_Gr);
GrJPG = imread(FullPathGr);
GrJPG = rgb2gray(GrJPG);
imshow(GrJPG)
set(handles.RB_Gr,'UserData',GrJPG);
```

La función `uigetfile` permite abrir una ventana para buscar un archivo. Las variables `GrName` y `GrPath`, guardan el nombre y la dirección de la imagen dentro de la computadora, respectivamente. La función `strcat` une las variables `GrPath` y `GrName` en la variable `FullPathGr`. La función `axes` crea un objeto gráfico que lo asigna al Axes `A_Gr`. La función `imread` lee una imagen y la asigna a la variable

GrJPG. La función `rgb2gray` transforma la imagen leída a una imagen en blanco y negro, eso debido a que es más sencillo aplicar acciones de procesamiento a este tipo de imágenes. La función `imshow` proyecta la variable GrJPG. La función `set`, asigna la variable GrJPG, a la propiedad `UserData` del objeto `RB_Gr`. Es decir, la imagen será almacenada temporalmente en el objeto `RB_Gr` y a partir de ahí se estará llamando desde otros objetos para sus transformaciones.

Además, se presenta la opción de restaurar la imagen original después de aplicar ciertos procesamientos.

### **b) Aplicación de pre procesamiento**

La función `imadjust` ajusta el rango de los niveles de gris de una imagen. Requiere de 3 entradas: el nombre de la variable a ajustar, el rango de entrada y el rango de salida. Los rangos de entrada y de salida, son cada uno un vector de dos valores entre 0 y 1, que corresponden al valor mínimo y máximo a considerar del rango. La función transforma el mínimo y el máximo valor del rango de entrada en el mínimo y el máximo valor asignado del rango de salida, ajustando así su contraste.

La función `histeq` ajusta el histograma de las imágenes a cierta distribución de probabilidad de acuerdo a la expresión:  $J = \text{histeq}(I, \text{hgram})$ , Donde  $I$  es la imagen inicial,  $J$  es la imagen de salida y `hgram` es un vector donde se aparecen las frecuencias de una distribución de probabilidad a la cual deseamos ajustar nuestro histograma. Según Gonzalez et al. (2004), si el número de valores  $L$ , -o clases, si consideramos que es un histograma- que tiene `hgram` es mucho menor al número de niveles de gris que tiene la imagen, el histograma de la imagen de salida se ajusta mejor al propuesto. Para nuestro trabajo, se le asignará un tamaño constante de  $L = 40$ . Se propondrá que el histograma se ajuste a una distribución uniforme (por default), normal (con media y sigma ajustables) y exponencial (con media ajustable). La media y la desviación estándar serán, por default 20.5 y 7, respectivamente.

Para un ajuste uniforme, no es necesario hacer un algoritmo.

Para el caso normal se utilizará el siguiente algoritmo.

```
media=get(handles.Media,'string');
desvest=get(handles.Sigma,'string');
media=str2num(media);
desvest=str2num(desvest);
L=(1:1:40);
P=normcdf(L,media,desvest)-normcdf(L-1,media,desvest);
X=round(numel(Gr_JPG)*P);
```

Para el caso exponencial se utilizará el siguiente algoritmo.

```
media=get(handles.Media,'string');
media=str2num(media);
L=(1:1:40);
P=expcdf(L,media)-expcdf(L-1,media);
X=round(numel(Gr_JPG)*P);
```

Este algoritmo lee el valor de la media y desviación estándar que el usuario ingrese, luego, con base a esos valores, calcula las probabilidades que le corresponderían a cada una de las clases de  $L$ , después multiplica esa probabilidad por el número de píxeles de la imagen determinado por `numel`. El módulo se observa en la Figura 4.6

The figure displays two screenshots of a software interface for image preprocessing. The top screenshot shows the 'Ajuste de Imagen (imadjust)' window. It has a title bar 'Seleccione el tipo de Procesamiento que desea ejecutar.' and two dropdown menus: 'Pre procesamiento' and 'Ajuste de Imagen (imadjust)'. Below these are two buttons: 'Finalizar' and '¿Qué es?'. To the right, under the title 'Seleccione los parámetros de ajuste', there are two sliders labeled 'Inicial, Bajo' and 'Inicial, Alto'. The bottom screenshot shows the 'Mejora de histograma (histeq)' window. It has a title bar 'Seleccione el tipo de Procesamiento que desea ejecutar.' and two dropdown menus: 'Pre procesamiento' and 'Mejora de histograma (histeq)'. Below these are two buttons: 'Finalizar' and '¿Qué es?'. To the right, under the title 'Seleccione el tipo de Histograma', there is a dropdown menu set to 'Normal', and two input fields: 'Media' with the value '20.5' and 'Sigma' with the value '7'. Below these is a 'GO!' button.

**Figura 4.6.** Módulo de pre procesamiento

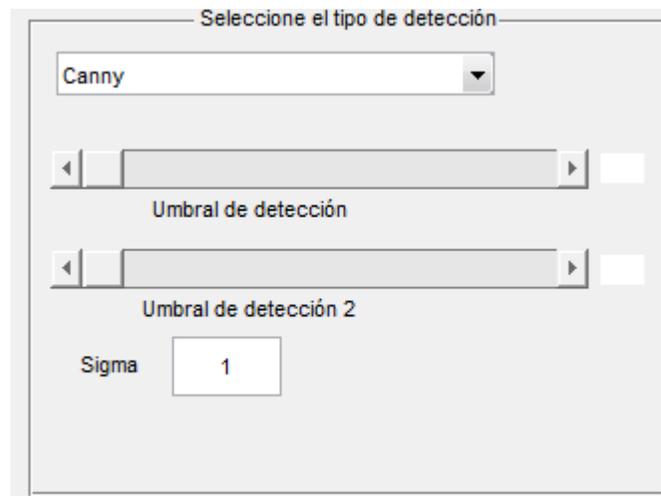
### c) Detección de bordes.

Para este caso se utilizó la función `edge` de MATLAB®, la cual cuenta con diversos tipos de detectores de bordes, tal como se muestran en la Tabla 4.2 extraída de la obra de Gonzalez et al. (2004). Las aproximaciones de derivación se muestran en la Figura 2.7.

Detector	Propiedades
Sobel	Utiliza la aproximación de derivación Sobel.
Prewitt	Utiliza la función de derivación Prewitt.
Roberts	Utiliza la función de derivación Roberts.
Laplaciano de Gaussiano	Busca cruces por cero después de filtrar la imagen con un filtro Gaussiano.
Canny	Busca máximos locales del gradiente de la imagen. El gradiente se calcula usando la derivada del filtro Gaussiano. El método utiliza 2 umbrales para detectar bordes fuertes y débiles e incluye los bordes débiles en la salida sólo si están conectados a bordes fuertes.

**Tabla 4.2.** Detectores de bordes de MATLAB®.

El módulo se observa en la Figura 4.7.



**Figura 4.7.** Módulo de detección de bordes.

### d) Detección de región de interés.

Se realizó un módulo que permitiera encerrar en un círculo nuestra región de interés. Para eso es necesario especificar el centro y el radio de dicho círculo. La función `getline` de MATLAB® nos permite hacer clic con el mouse sobre la imagen

especificada para dibujar una poli línea y obtener como resultado los puntos que corresponden a los vértices de dicha poli línea. En este caso solo nos interesan el punto del centro y la longitud del radio, por lo que nuestra poli línea será una línea con sólo dos puntos.

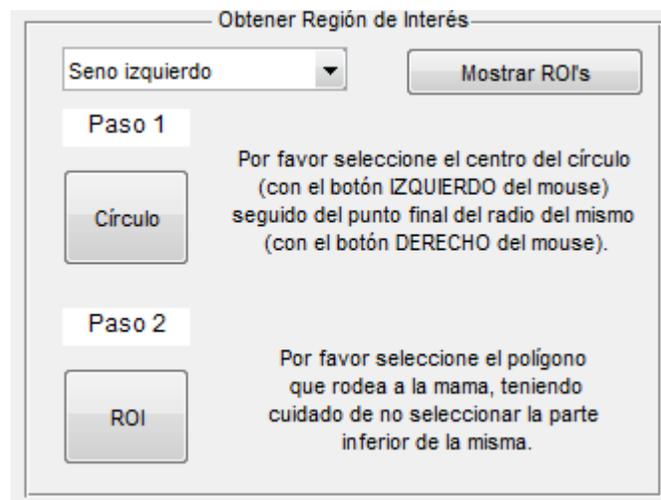
Con los dos puntos se utiliza la ecuación del círculo que se muestra en la Ecuación (4.1), en la cual también se presenta su correspondiente despeje de  $Y$ .

$$(X - X_0)^2 + (Y - Y_0)^2 = r^2$$

Ó

$$Y = Y_0 \pm \sqrt{r^2 - (X - X_0)^2} \quad (4.1)$$

Donde  $r = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$  es el radio -redondeado al siguiente entero con la función `ceil`- (que corresponde a la distancia entre dos puntos) y  $X_0$  e  $Y_0$  corresponden a las coordenadas del centro. Como complemento se utilizó la función `roipoly`, la cual obtiene la región de interés a partir de una poli línea que el usuario traza con el mouse en la imagen. El módulo se observa en la Figura 4.8.



**Figura 4.8.** Módulo de ROI.

**e) Dar valor de 0 a todo lo que no sea la región de interés.**

Para cumplir con este objetivo es necesario tener los pixeles correspondientes a las figuras. Al momento de utilizar, por ejemplo, la ecuación del círculo, MATLAB® puede guardar automáticamente los pixeles correspondientes a los puntos del borde del mismo. Después se genera una nueva imagen con el mismo tamaño que la

imagen original, la cual servirá como auxiliar para operaciones futuras. A partir de las coordenadas de los pixeles correspondientes a los bordes del círculo (o cualquier forma geométrica utilizada) se debe hacer una discriminación para convertir en 1 todo lo que cumpla con el siguiente par de condicionales y 0 el caso contrario:

$$Si: X_{min} \leq x \leq X_{max} \quad \vee \quad Si: Y(x)_{neg} \leq y \leq Y(x)_{pos}$$

Donde:  $x$  e  $y$  son las coordenadas de cada pixel;  $X_{min}$  y  $X_{max}$  corresponden a la coordenada del centro habiendo restado y sumado, respectivamente, el radio en el eje  $X$ ;  $Y(x)_{neg}$  y  $Y(x)_{pos}$  corresponden a la coordenada  $Y$  en función de  $X$  para cuando el valor de  $Y$  es negativo y positivo, respectivamente. Esta operación se puede adecuar para otras formas geométricas hasta obtener la región de interés.

#### f) Usar regiones de interés para filtrar los archivos CSV.

Con el archivo obtenido en el punto anterior, podemos hacer una serie de operaciones para encontrar solamente las temperaturas que nos interesen. El archivo CSV de un termograma tiene la temperatura de cada pixel en una matriz del tamaño de la imagen obtenida.

Para abrir el archivo CSV, se utilizó el siguiente algoritmo:

```
[CSVName CSVPath] = uigetfile('*.csv','Abrir archivo de datos CSV');
FullPathCSV = strcat(CSVPath,CSVName);
CSV_data = csvread(FullPathCSV);
```

En este caso nos interesa solamente la temperatura para cada región y el resto de la imagen convertirlos a 0 y calcular su temperatura máxima de una manera más sencilla. Se utilizó el siguiente algoritmo:

```
i = 1;
j = 1;
for i = 1:240
for j = 1:320
    X(i,j)=Y(i,j)*CSV(i,j);
end
end
```

La variable `CSV` representa el archivo CSV original,  $Y$  representa la matriz de 1 y 0 que denota la región de interés obtenida con anterioridad y  $X$  representa el archivo

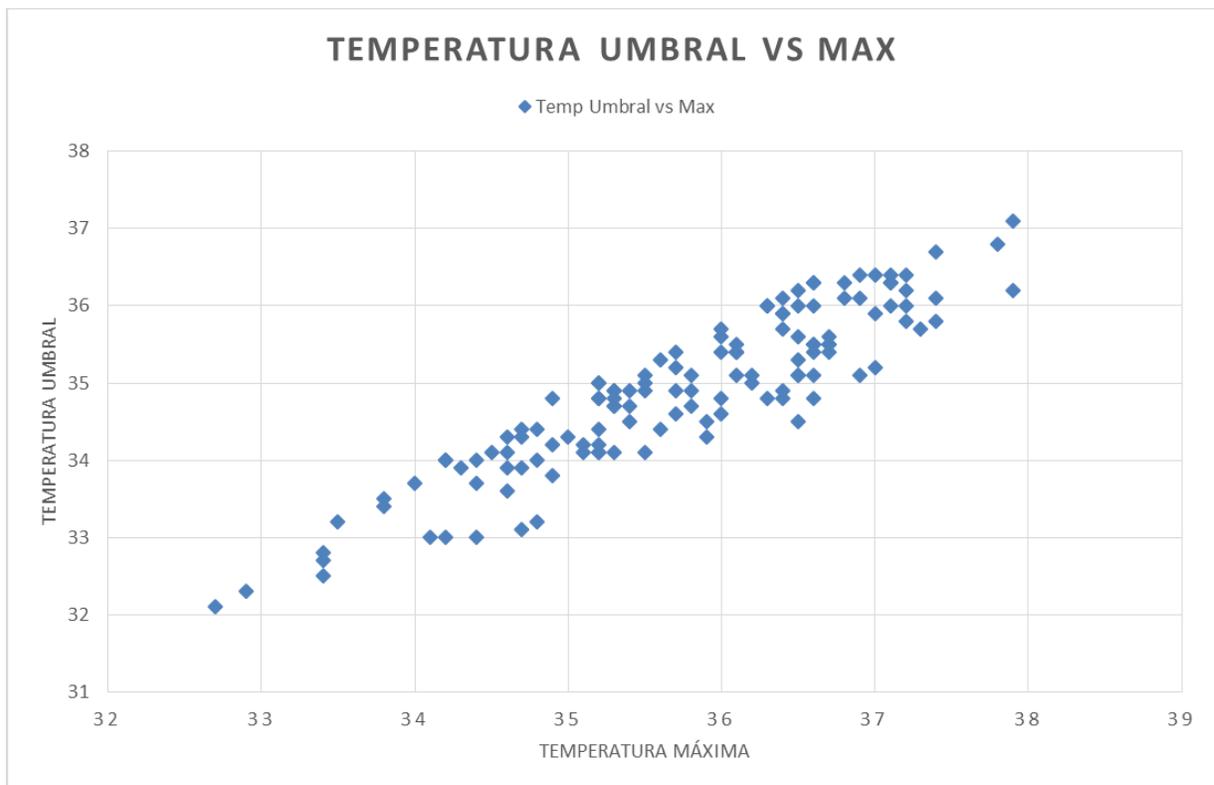
filtrado. Si el valor de  $\Upsilon$  es 0, convierte el valor de la temperatura en 0; en otro caso, conserva el valor original.

**g) Obtener los archivos para cada región.**

La variable generada puede ser guardada como un nuevo archivo que será útil para nuevos análisis posteriores. Esto se logra mediante la función `csvwrite`.

## 4.4 Analizar los datos.

Al momento de graficar los datos obtenidos, poniendo como variable independiente ( $x$ ) la temperatura máxima del tumor y como variable dependiente ( $y$ ) la temperatura umbral, se obtiene el gráfico que se observa en la Figura 4.9. Por lo tanto, se pensó en la idea de proponer que los datos siguen una distribución normal bivariada.



**Figura 4.9.** *Dispersión de los datos. Temperatura máxima vs umbral.*

La temperatura máxima tiene un promedio de 35.7446 y la temperatura umbral un promedio de 34.9046. Es decir, su vector de medias es:

$$\bar{x} = \begin{bmatrix} 35.7446 \\ 34.9046 \end{bmatrix}$$

La matriz de varianzas y covarianzas de los datos es:

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

La correlación entre la temperatura máxima y la temperatura umbral es de 0.9170.

Su matriz de precisión es:

$$S^{-1} = \begin{bmatrix} 5.1370 & -5.0745 \\ -5.0745 & 5.9615 \end{bmatrix}$$

#### 4.4.1 Realizar propuesta de modelo multivariado.

Se propone probar que los datos sigan una distribución normal bivariada con lo cual, aplicando el principio que indica que la matriz de precisión permite encontrar los coeficientes de regresión de una variable explicada por todas las demás variables, tenemos el modelo de la Ecuación (4.2).

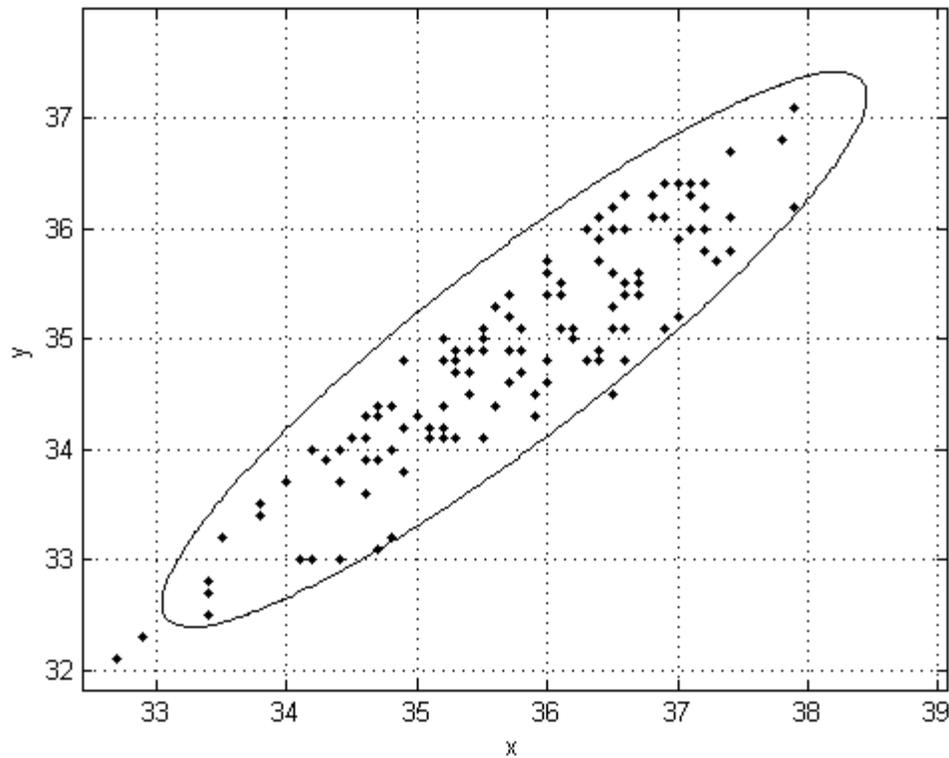
$$\begin{aligned} (y - \mu_y) &= \beta_{yx}(x - \mu_x) \\ (y - 34.9046) &= 0.8512(x - 35.7446) \quad \text{ó} \quad (4.2) \\ y &= 34.9046 + 0.8512(x - 35.7446) \end{aligned}$$

Donde  $x$  corresponde a la temperatura máxima del nuevo sujeto a analizar e  $y$  la temperatura umbral estimada. Los cálculos de  $\beta_{yx}$  se muestran a continuación:

$$\hat{\beta}_{yx} = -s^{ij} s_r^2(y) = -(-5.0745) \times \left( \frac{1}{5.9615} \right) = 0.8512$$

#### 4.4.2 Probar modelo multivariado.

El modelo multivariado se probó con la propiedad que indica que si se cortara con hiperplanos paralelos al hiperplano definido por las variables, se obtienen curvas de nivel definidas por la distancia de Mahalanobis, distribuida como una  $\chi^2$  con  $p$  grados de libertad. Al graficar los datos dentro de dicha distancia, tal como se observa en la Figura 4.10, podemos afirmar que los datos se comportan como una normal bivariada, con una confianza del 95% y 2 grados de libertad.



*Figura 4.10. Prueba de modelo multivariado.*

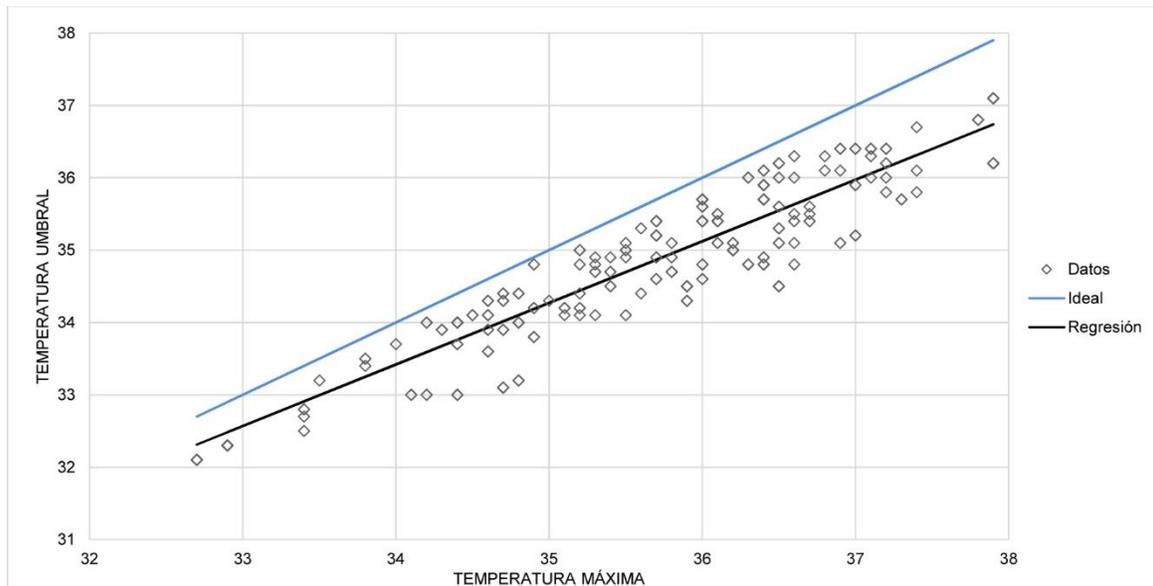
#### **4.4.3 Validar modelo multivariado.**

Con esas características el modelo es capaz de arrojar una aproximación a la temperatura umbral, dada una temperatura máxima. Teniendo en cuenta lo anterior se estimó la temperatura umbral para las 19 imágenes que se habían dejado para validación. Además, la temperatura umbral se ajustó hasta encontrar la anomalía térmica, misma que se define, para nuestro caso, 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. Dado que las nuevas imágenes no contarán con información de diagnóstico, no se conocerá el punto donde se encuentra la anomalía. Por lo anterior y para el estudio de los casos de validación, se define como temperatura umbral el mínimo de las dos temperaturas máximas de ambos senos.

El modelo a su vez traza una línea de regresión con las temperaturas pronosticadas, tal como se mostró en la Ecuación (4.1) de la sección 4.4.1, y como se muestra en la línea negra en la Figura 4.11.

El caso ideal sería cuando un paciente que se practica una termografía tiene una temperatura umbral exactamente igual a la temperatura máxima. Esta línea se muestra de color azul en la Figura 4.11.

Se propone que mientras más cercano se encuentre un punto a la línea azul (y sobre la línea negra de regresión), existe menos probabilidad que en el paciente exista un proceso fisiológico sospechoso; mientras que, si el punto cae por debajo de la línea negra, se puede pensar que hay altas probabilidades que exista un proceso fisiológico que altere la temperatura del paciente.



**Figura 4.11.** Línea de regresión multivariada pronosticada.

Al analizar los datos que quedaron al final para validación, se obtuvieron los resultados mostrados en la Tabla 4.3. En dicha tabla aparece, en la primera columna, el número de muestra analizado; seguido por su temperatura máxima; luego la temperatura umbral pronosticada por el modelo de regresión; después la temperatura umbral determinada de manera manual; posteriormente se observa el cálculo de la diferencia entre la temperatura umbral obtenida manual y la del modelo de regresión; por último se observa el diagnóstico que el oncólogo asignó a dicho caso.

Se puede observar que a medida que la diferencia tiende a ser un número negativo más alejado de cero, el caso se observa con mayor gravedad, mientras que un número positivo más alejado de cero se refiere a casos no detectados contundentemente. Esta información hace pensar que nuestro modelo es capaz de realizar afirmaciones sobre la probabilidad de padecer alguna anomalía dependiendo si un determinado caso se encuentra por encima o por debajo de la línea de regresión multivariada.

No. Muestra	Temp. máxima	Umbral reg. ( $\hat{Y}_i$ )	Temp. umbral ( $Y_i$ )	$Y_i - \hat{Y}_i$	Diagnóstico
2	35.8	34.95	35.3	0.35	Nódulo de mama benigno.
3	36.6	35.63	36.5	0.87	No detectado en estas imágenes
6	37.1	36.06	35.5	-0.56	Alto riesgo de cáncer con componente inflamatorio
9	36.6	35.63	34.6	-1.03	Alto riesgo malignidad
17	37.4	36.31	35.3	-1.01	Alto riesgo de cáncer con componente inflamatorio
24	36.5	35.55	35.7	0.15	Hallazgo sospechoso
46	37.4	36.31	36.1	-0.21	Hallazgo inflamatorio
51	35.8	34.95	34.7	-0.25	Alto riesgo malignidad
66	33.7	33.16	33.1	-0.06	Alto riesgo malignidad
75	36	35.12	35.7	0.58	Sospechoso
83	37	35.97	35.7	-0.27	Alto riesgo malignidad
92	36.4	35.46	34.8	-0.66	Confirmatorio de malignidad
95	35.7	34.87	34.4	-0.47	Alto riesgo malignidad
99	34.6	33.93	33.6	-0.33	Alto riesgo malignidad
117	37.1	36.06	36.7	0.64	Hallazgo benigno
121	34.8	34.10	33.9	-0.20	Alta probabilidad de riesgo
123	34.5	33.85	34	0.15	Riesgo de malignidad
127	36.1	35.21	35.7	0.49	Hallazgo benigno
135	35.8	34.95	35.4	0.45	Riesgo de malignidad

**Tabla 4.3.** Resultados de la validación.

En la Figura 4.12 se pueden observar estos casos comparados con la línea de regresión. El número de muestra aparece enseguida de cada punto.

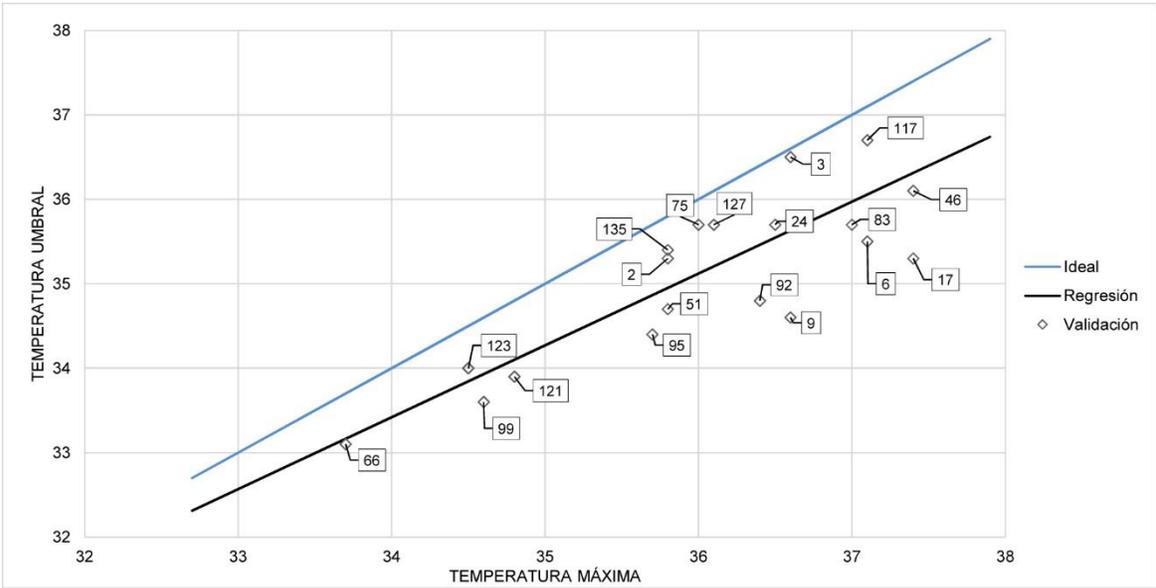


Figura 4.12. Casos de validación en la línea de regresión.

#### 4.4.4 Realizar propuesta de modelo univariado.

Después, con la idea de tener un modelo univariado del cuál podríamos obtener información adicional, se procuró calcular la diferencia entre la temperatura máxima y la temperatura umbral detectada manualmente para cada caso.

Con los datos de las diferencias entre la temperatura máxima y la temperatura umbral de cada seno se obtiene el histograma que se muestra en la Figura 4.13.

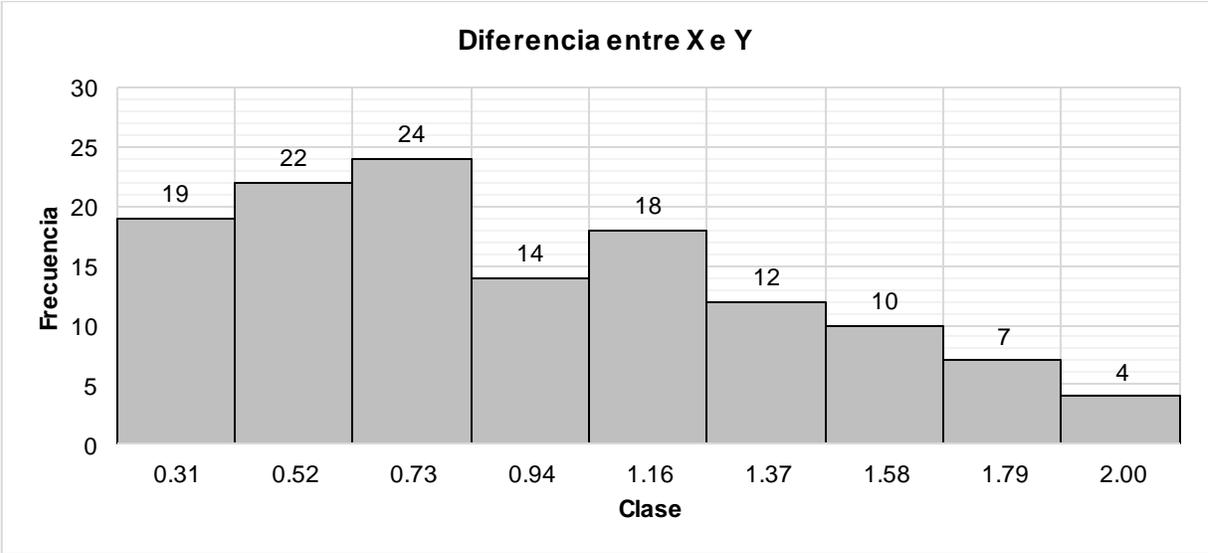


Figura 4.13. Histograma de las diferencias entre temperatura máxima (X) y temperatura umbral (Y).

Con dicha información, se harán las pruebas de bondad de ajuste con diversas distribuciones de probabilidad (exponencial, normal, lognormal y gamma) y se determinará si la distribución de los datos se comportan como alguna de ellas.

#### 4.4.5 Probar modelo univariado.

Se realizarán las siguientes pruebas de hipótesis.

$H_0$ : La diferencia entre la temperatura máxima del seno y la temperatura umbral se comporta de acuerdo a una distribución (depende de cada prueba) con parámetros (depende de cada prueba).

$H_1$ : La diferencia entre la temperatura máxima del seno y la temperatura umbral no se comporta de acuerdo a una distribución (depende de cada prueba) con parámetros (depende de cada prueba).

Las pruebas de bondad de ajuste, mediante la prueba de la Chi-Cuadrada (utilizando la función `chi2gof` de MATLAB®) dieron los resultados que se muestran en la Tabla 4.4.

Distribución	Valor P	Decisión
Normal	0.0161	Rechazar $H_0$ .
Exponencial	0.00005	Rechazar $H_0$ .
Lognormal	0.0092	Rechazar $H_0$ .
Gamma	0.0084	Rechazar $H_0$ .

*Tabla 4.4. Resultados de la prueba de bondad de ajuste.*

#### 4.4.6 Analizar correspondencia entre modelos.

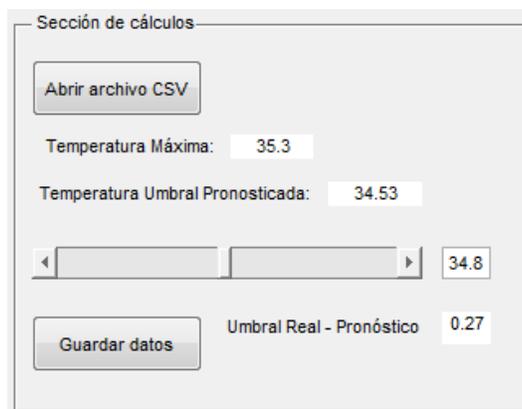
Dado que el modelo univariado no se logró concretar, solamente existe el modelo multivariado.

### 4.5 Incluir modelos en la GUI.

Debido a que solamente el modelo multivariado pasó la prueba de hipótesis, este se incluirá en la parte de pruebas de hipótesis. Para ello, se agregará un nuevo elemento al PUM que despliega las funciones que tiene la GUI. Este módulo lee el

archivo CSV y calcula la temperatura máxima de la región de interés obtenida con anterioridad.

Después encuentra, a partir de la temperatura máxima, la temperatura umbral pronosticada, misma que se puede ajustar esta temperatura umbral con un Slider hasta encontrar un umbral que resulte adecuado para detectar anomalías térmicas. Asimismo, el software calculará la diferencia entre la temperatura umbral del Slider y la pronosticada lo que permitirá al usuario tener una mejor visión del caso. El módulo se observa en la Figura 4.14.



Sección de cálculos

Abrir archivo CSV

Temperatura Máxima: 35.3

Temperatura Umbral Pronosticada: 34.53

34.8

Guardar datos Umbral Real - Pronóstico 0.27

**Figura 4.14.** Módulo de cálculos

## 5. RESULTADOS.

En este capítulo se detallan los resultados obtenidos por la implementación de la metodología descrita en el capítulo anterior, mismos que nos permitirán alcanzar las conclusiones del trabajo. En una primera instancia se recapitula cómo se obtuvieron los datos y su análisis; después se muestra un resumen de los resultados más relevantes obtenidos.

### 5.1 Obtención de resultados.

Los datos, que consisten en la temperatura máxima del seno enfermo y la temperatura umbral a la cual se detecta el tumor se obtuvieron, como se menciona en las secciones 4.1 y 4.2, a partir de tres fuentes de manera simultánea:

- De los diagnósticos de los pacientes en los cuales el doctor Martín del Campo había señalado cuál era la región enferma para cada persona.
- De las imágenes `*.fff` que nos brindó el doctor Martín del Campo y mediante el software de FLIR se obtuvieron las temperaturas máximas del seno enfermo.
- De los archivos `*.csv` que, procesados mediante MATLAB®, se obtuvo la temperatura umbral de cada seno enfermo.

Con estos datos se realizaron todos los estudios estadísticos del presente trabajo. Dicho análisis consistió en probar si los datos se podrían ajustar a una distribución normal bivariada para predecir el umbral para nuevas imágenes, y ver si es capaz esta predicción dar información sobre el estado del paciente. Además se probó si la diferencia entre ambas variables podrían servir para realizar un modelo univariado; es decir, si los datos serían capaces de ajustarse a una distribución de probabilidad teórica que nos permitiera obtener información extra con ellos.

Por otro lado, se desarrolló un software que permitiera analizar las imágenes y sus archivos de temperatura (CSV). Esta GUI consiste en una serie de módulos capaces de leer una imagen, aplicarle pre procesamiento (ajuste de niveles de gris y de

histograma), detectar los bordes de la imagen y seleccionar una región de interés que se utiliza para filtrar los archivos CSV; es decir, calcular la temperatura máxima solo de las áreas que nos importan, para evitar que temperaturas mayores ubicadas en otro lugar de la imagen que no sea el seno den resultados falsos. Además, se calcula una temperatura umbral para dicha temperatura máxima y una temperatura umbral que se puede ajustar mediante un slider. El software también calcula la diferencia entre la temperatura umbral real y la pronosticada. Si esta diferencia es negativa (lo que indicaría un punto debajo de la línea de regresión) el texto se pone de color rojo; en caso contrario de color negro.

## **5.2 Resultados obtenidos.**

Un resumen de los resultados más importantes obtenidos de esta investigación se muestra a continuación:

1. Un modelo basado en una distribución normal bivariada para predecir la temperatura máxima y la temperatura umbral a la cual se detectan anomalías en el seno. Este consiste en una ecuación que nos permite predecir el estado del paciente. El modelo se muestra en la sección 4.4.1, su prueba en la sección 4.4.2 y su validación en la sección 4.4.3.
2. La evidencia para afirmar que la diferencia entre ambas temperaturas no se puede ajustar a una distribución de probabilidad teórica. Esto se observa en la sección 4.4.5.
3. Un software que permite seccionar las imágenes de las personas y extraer solamente el área del seno. El desarrollo de ese software se detalla en la sección 4.3.
4. Un software que permite aprovechar la información del modelo basado en la distribución normal bivariada para predecir nuevos umbrales. Esto se detalla en la sección 4.5.

## **6. CONCLUSIONES, RECOMENDACIONES Y TRABAJOS FUTUROS.**

En esta sección se engloban las conclusiones obtenidas a partir del trabajo de investigación y su correspondiente aplicación.

### **6.1 Conclusiones.**

- Se define, para nuestro caso, 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.

### **6.2 Recomendaciones.**

El modelo presentado fue obtenido a partir de los datos dados por el doctor Martín del Campo. Debido a que estos datos corresponden a una población de una localidad específica, es posible que el modelo sea ligeramente diferente para

poblaciones de otro lugar. Se recomienda entonces realizar un muestreo preliminar que permita corroborarlo.

Es importante mencionar que el modelo presentado así como el software desarrollado no deben ser utilizados como sustitución del trabajo del oncólogo; sin embargo puede ser utilizado como una herramienta de apoyo a la toma de decisiones.

### **6.3 Trabajos futuros.**

Como parte de las investigaciones que se sugieren a partir de esta investigación se encuentran:

- Verificar qué tan sensible es este método para la detección específica del cáncer.
- Realizar los cálculos de la probabilidad de riesgo de tener alguna anomalía debida a un proceso metabólico a partir de la distancia entre el punto encontrado y la línea de regresión, de tal manera que se pueda llegar a una conclusión cuantificable.

## 7. REFERENCIAS.

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.

Bolan, P. J., Nelson, M. T., Yee, D. & Garwood, M. 2005. Imaging in breast cancer: Magnetic resonance spectroscopy. *Breast Cancer Research*, 7, 149-152.

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.

Elmore, J. G., Barton, M. B. & Mocerri, M. P. 1998. Mammography: False-positive results. *Southern Medical Journal*, 91, 886-887.

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.

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.

Gonzalez, R. C., Woods, R. E. & Eddins, S. L. 2004. *Digital Image processing using MATLAB*, Upper Saddle River, N.J., Pearson Prentice Hall.

Hair Jr., J. F., Black, W. C., Babin, B. J. & Anderson, R. E. 2009. *Multivariate Data Analysis*, Prentice Hall.

Harris, J. 2009. Meta-analysis: Mammography increased breast cancer risk for high-risk women younger than 30. *Hem/Onc Today*, 40-41.

Hernández Sampieri, R., Fernández-Collado, C. & Baptista Lucio, P. 2006. *Metodología de la investigación*, Mc Graw-Hill.

- Ibarra-Castanedo, C., Bendada, A. & Maldague, X. 2007. Thermographic Image Processing for NDT. In: ESTRUCTURALES, A. A. D. E. N. D. Y. (ed.) IV Conferencia Panamericana de Ensayos No Destructivos. Buenos Aires, Argentina.
- Ibarra-Castanedo, C., González, D., Klein, M., Pilla, M., Vallerand, S. & Maldague, X. 2004. Infrared image processing and data analysis. *Infrared Physics & Technology*, 46, 75-83.
- Jiang, L., Zhan, W. & Loew, M. H. 2010. Modeling static and dynamic thermography of the human breast under elastic deformation. *IOPscience*, 187-202.
- 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.
- Joro, R., Lääperi, A.-L., Soimakallio, S., Järvenpää, R., Kuukasjärvi, T., Toivonen, T., Saaristo, R. & Dastidar, P. 2008. Dynamic infrared imaging in identification of breast cancer tissue with combined image processing and frequency analysis. *Journal of Medical Engineering & Technology*, 32, 325-334.
- 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.
- Knaul, F. M., Nigenda, G., Lozano, R., Arreola-Ornelas, H., Langer, A. & Frenk, J. 2009. Scielo. Cáncer de mama en México: una prioridad apremiante. [Online], 51. Available: <http://www.scielo.org.mx/pdf/spm/v51s2/v51s2a26.pdf>.
- 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 Association*, 11, 152-157.
- Laufer, S., Solomon, S. B. & Rubinsky, B. 2012. Tissue characterization using electrical impedance spectroscopy data: a linear algebra approach. *Physiological Measurement*, 33, 997-1013.

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.

Marino, A. A., Morris, D. M., Schwalke, M. A., Iliiev, 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.

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.

Martín, M. 2002. Técnicas clásicas de segmentación de imagen [Online]. Laboratorio de procesamiento de imagen. Available: <http://poseidon.tel.uva.es/~carlos/ltif10001/segmenclasica.pdf> [Accessed August 20 2012].

MATHWORKS. 2013. MATLAB GUI [Online]. Available: <http://www.mathworks.com/discovery/matlab-gui.html> [Accessed 19 de Enero 2013].

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.

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.

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

Montgomery, D. C. 2005. *Introduction to Statistical Quality Control*, Arizona State University.

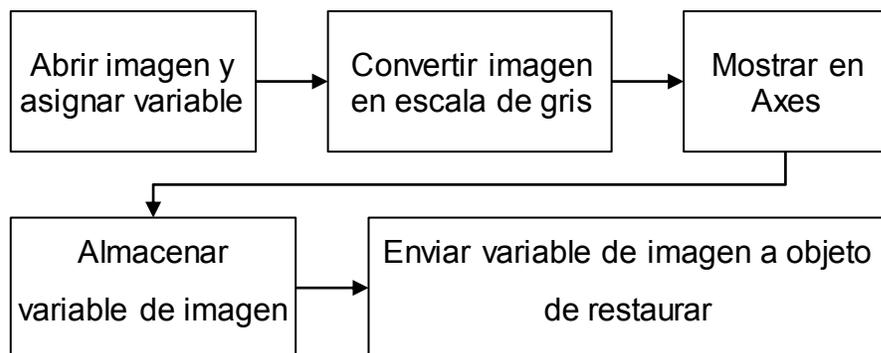
- Montgomery, D. C. & Runger, G. C. 2003. Applied Statistics and Probability for Engineers, John Wiley & Sons, Inc.
- Nakayama, H., Kawase, T., Okuda, K., Wolff, L. F. & Yoshie, H. 2011. In-vivo near-infrared optical imaging of growing osteosarcoma cell lesions xenografted in mice: dual-channel quantitative evaluation of volume and mineralization. *Acta Radiologica*, 52, 978-988.
- 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.
- 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.
- Ng, E. Y.-K. & Kee, E. C. 2008. Advanced integrated technique in breast cancer thermography. *Journal of Medical Engineering & Technology*, 32, 103-114.
- 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> [Accessed 20 sept 2012].
- 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.
- 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.
- 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.

- Peña, D. 2002. Análisis de datos multivariantes, Madrid, McGraw-Hill/Interamericana.
- Pérez López, C. 2005. Muestreo Estadístico. Conceptos y problemas resueltos, Madrid.
- 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.
- 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.
- 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.
- Valera, R. E. 2010. Impacto psicológico de un falso positivo a la mamografía en un grupo de mujeres participantes de un programa de cribado poblacional de cáncer de mama. *Psicooncología*, 7, 207.

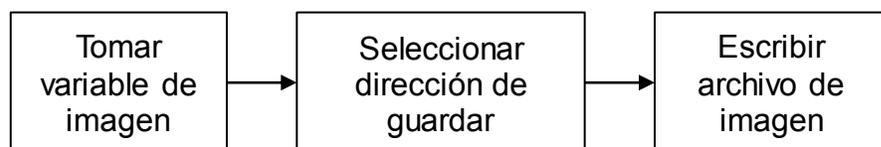
## 8. ANEXOS.

Algoritmos de la GUI:

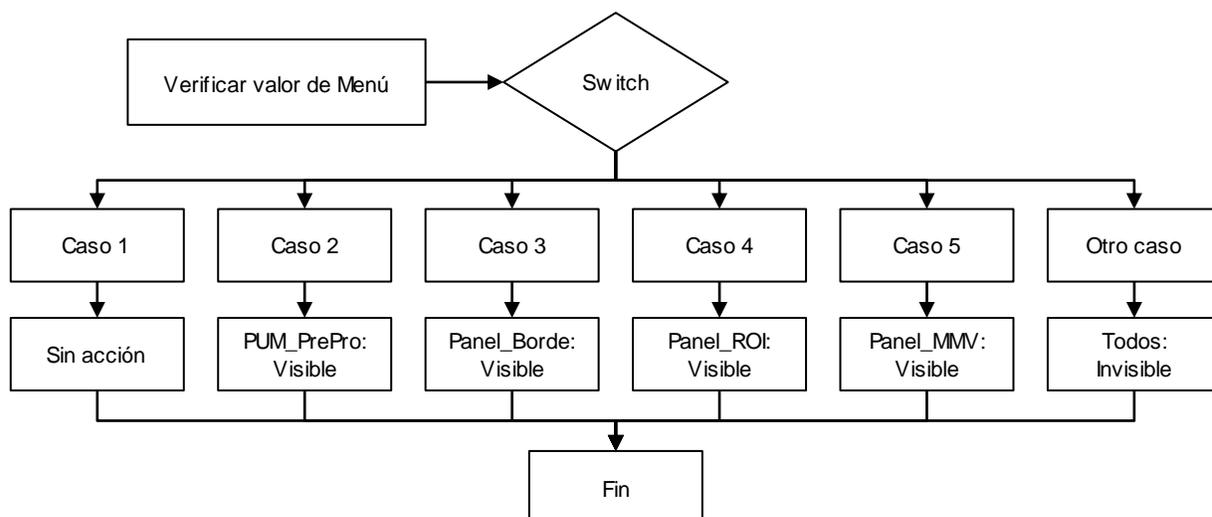
### Abrir imágenes.



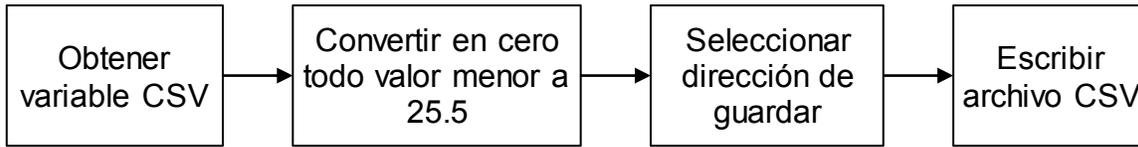
### Grabar imagen.



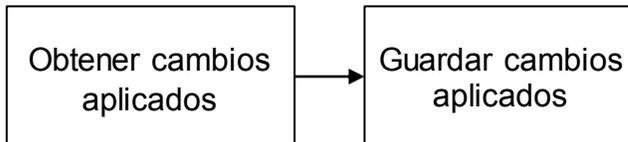
### Selección de acciones.



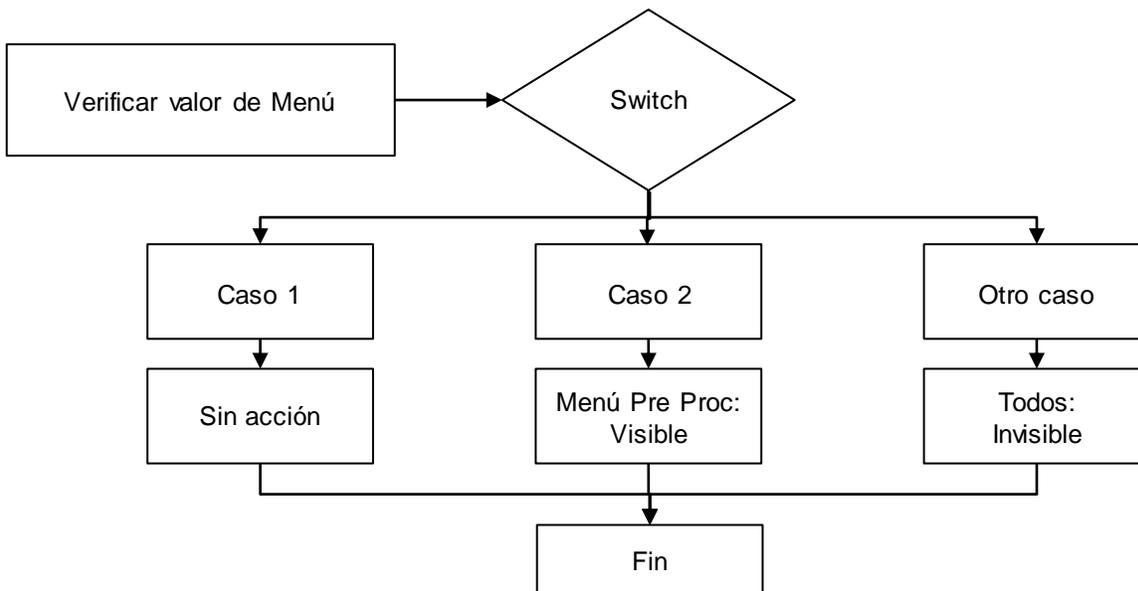
**Guardar archivo con datos SCV.**



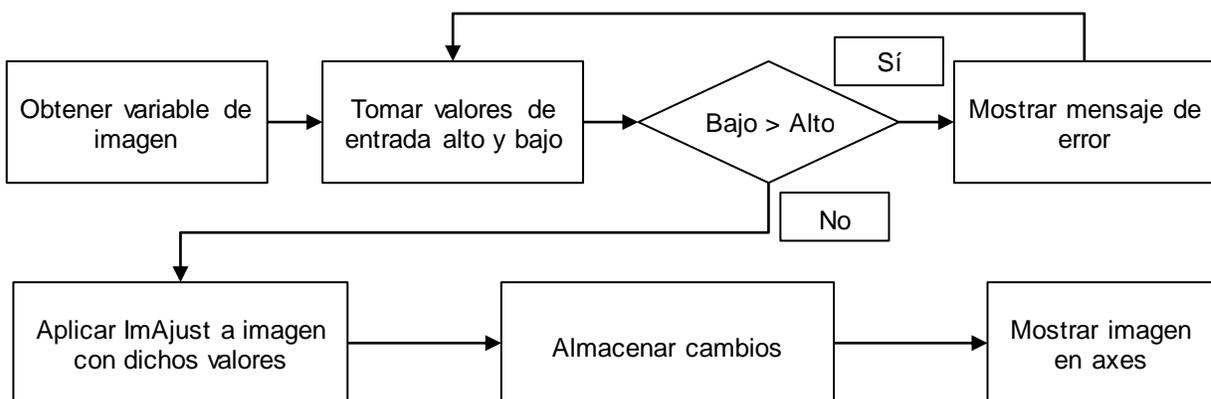
**Aplicar acciones.**

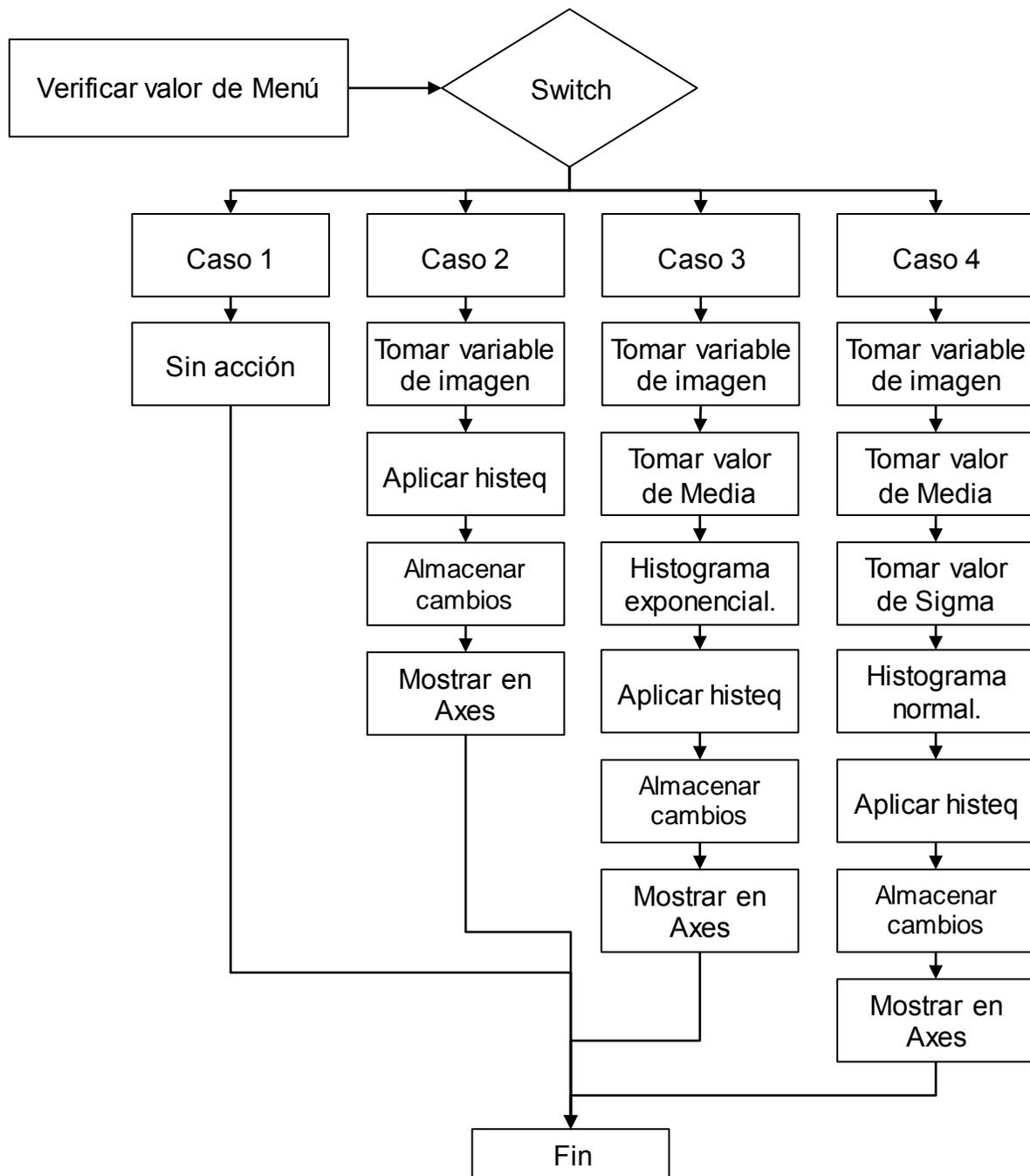


**Selección de pre procesamiento.**

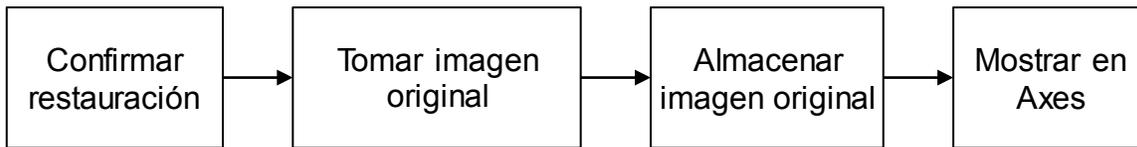


**Función ImAdjust.**

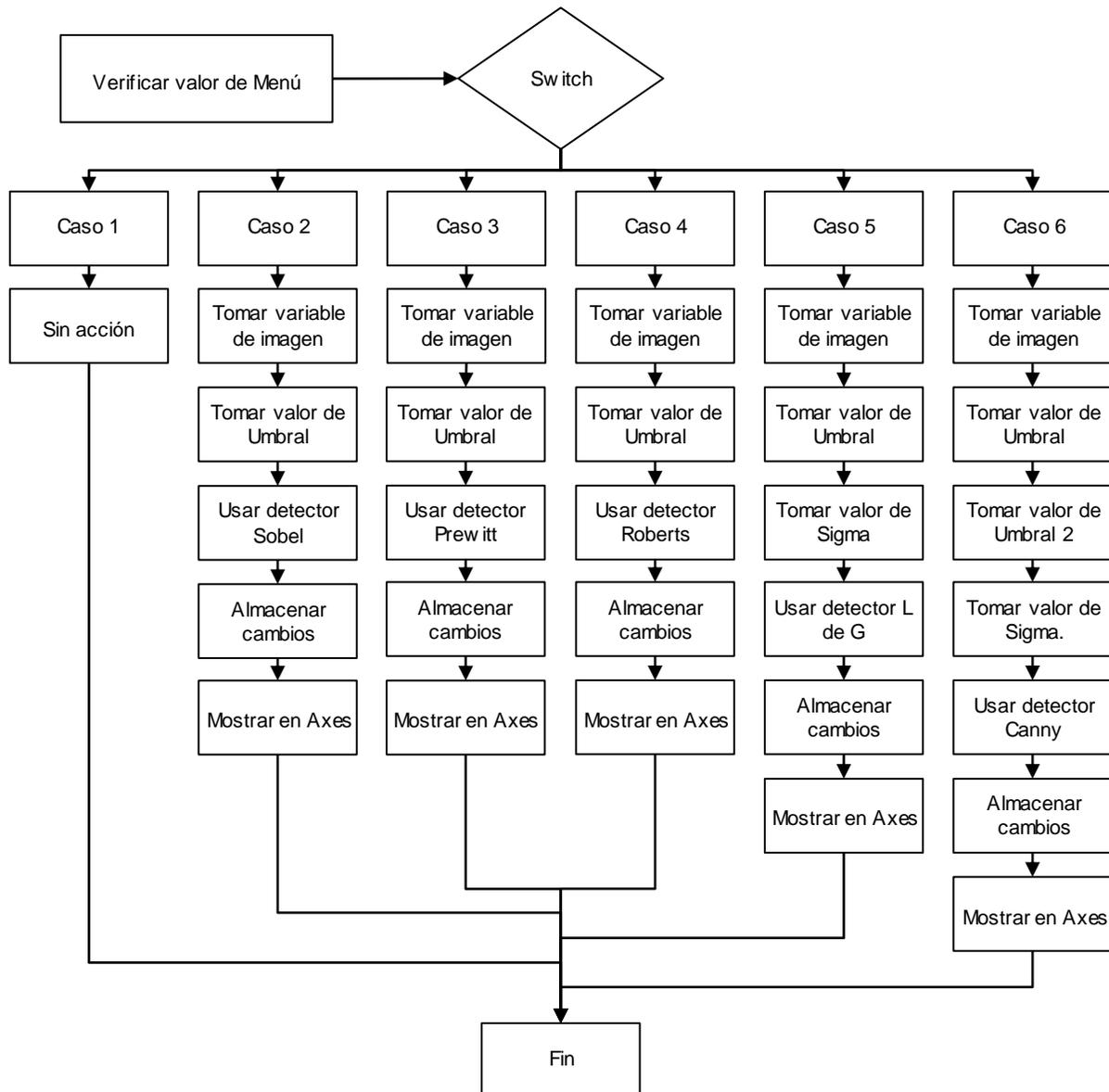


**Función ImHist.**

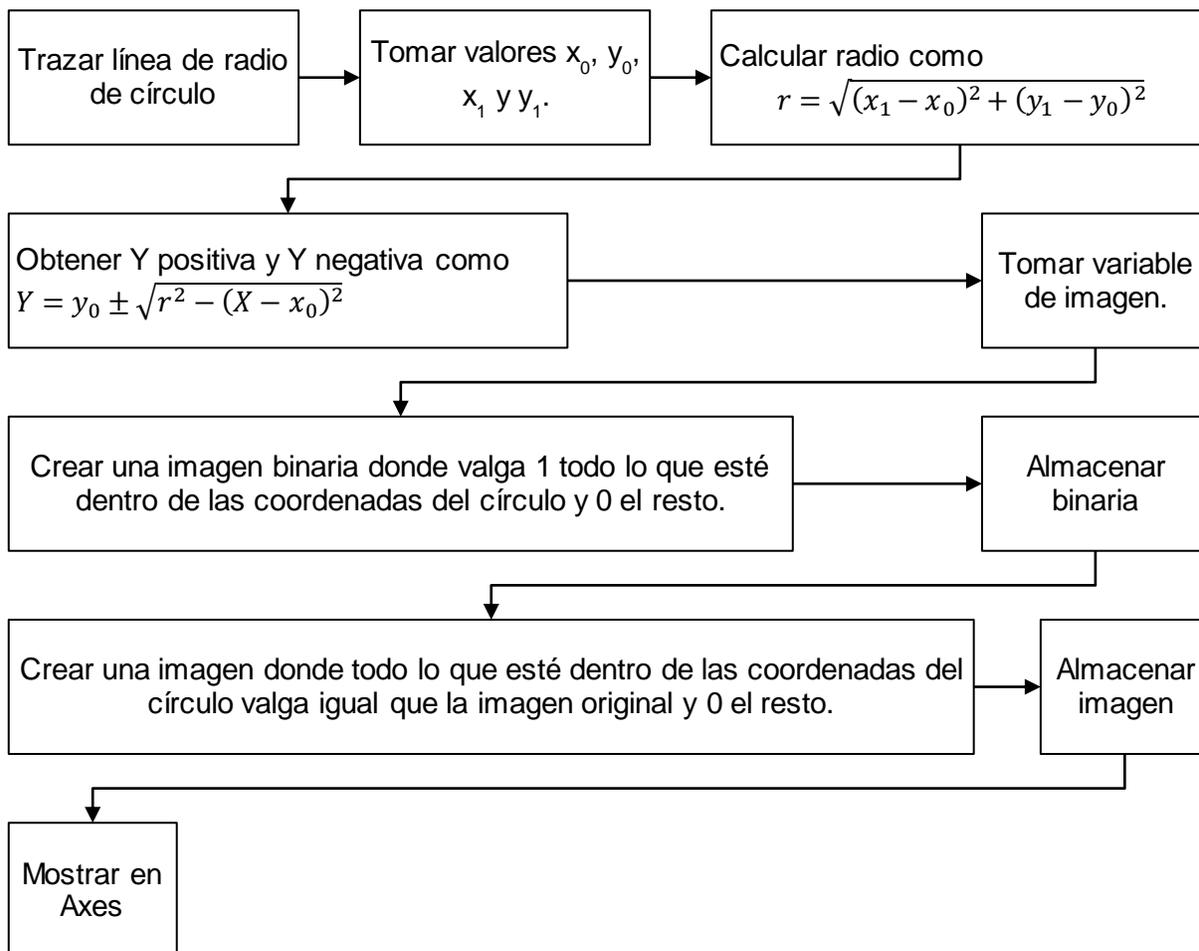
**Restaurar original.**



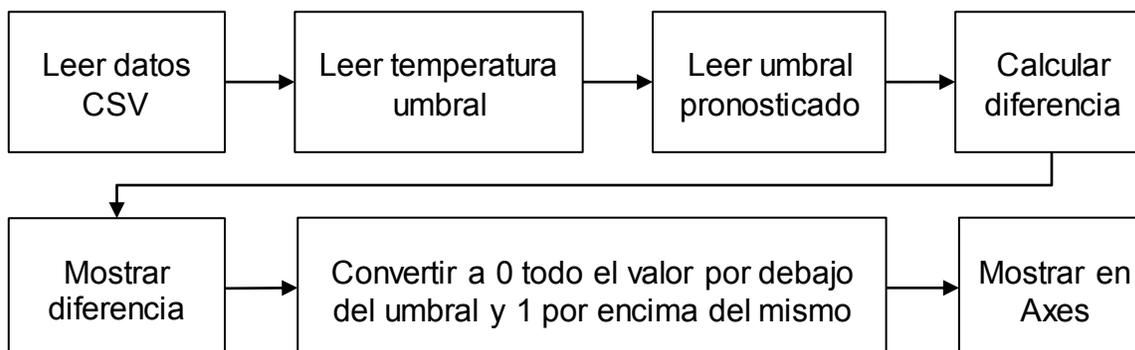
**Función Detector De Bordes.**



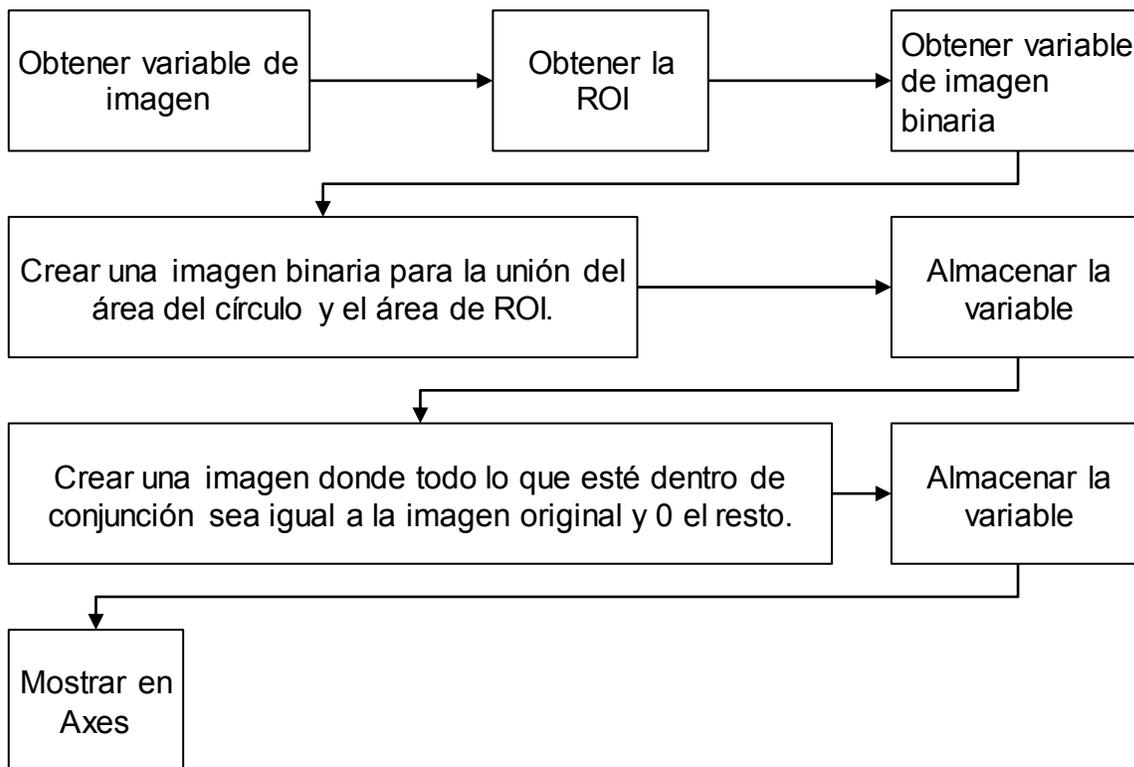
### Dibujar círculo de ROI.



### Binarización.



**Dibujar Poli línea de ROI.**



**Abrir archivo CSV, filtrar archivo y obtener umbral pronosticado.**

