

**PONTIFICIA UNIVERSIDAD JAVERIANA  
FACULTAD DE INGENIERÍA  
PROGRAMA DE MAESTRÍA EN INGENIERÍA ELECTRÓNICA**



**DETECCIÓN DE CARCINOMAS EN LAS IMÁGENES DE MAMOGRAFÍA  
MEDIANTE TECNICAS DE PROCESAMIENTO DE IMÁGENES Y  
RECONOCIMIENTO DE PATRONES**

**TRABAJO DE GRADO – MODALIDAD PROFUNDIZACIÓN**

**Para obtener el título de:**

**MAESTRO EN CIENCIAS EN LA INGENIERÍA ELECTRÓNICA**

**Presenta:**

**ING. GERARDO RAMÍREZ DE LA ESPRIELLA**

**Director:**

**ING. CARLOS ALBERTO PARRA. Ph.D.**

**BOGOTA, ENERO DEL 2011**

## DEDICATORIA

Al Poder Superior por sus regalos y oportunidad de volver a la Universidad a estudiar después de tantos años. A mi esposa Martha Idally, a mis hijos: María del Mar y Mateo a quienes les robe mucho tiempo en estos años, a mis padres: Inés y Álvaro (+pepd), Hermaños, y en especial a Tita mi hermana mayor quien ha vivido y sufrido esta enfermedad.

A mi abuelo Darío de la Espriella, mis tíos: Darío de la Espriella Cerón y Octavio de la Espriella Cerón quiénes sufrieron esta enfermedad y murieron por ella (+ qepd).

**Y** A todas las mujeres que han sufrido por falta de información y diagnóstico temprano así como a las asociaciones de ayudas para el tratamiento del cáncer.



## AGRADECIMIENTOS

Un especial agradecimiento a mi Director y profesor en las asignaturas de procesamiento de imágenes y reconocimiento de patrones: **Ing. Carlos Alberto parra Ph.D.**

A todos los profesores que durante el desarrollo de mis estudios de maestría me ilustraron.

Al Instituto Nacional de Cancerología de Colombia [www.cancer.goc.co](http://www.cancer.goc.co) en donde pude acercarme al conocimiento de la problemática compleja de esta enfermedad y en donde pude aprender parte del proceso del tratamiento y ayudas tecnológicas a la misma.

Al Ing. Samuel Oporto Díaz MSc. Quien trabajo en el tema y por su aporte en algunas funciones y algoritmos.

A **Prisma Technologies Inc. Ltda.** Por su apoyo económico en el financiamiento de este proyecto y estudios.

## INDICE GENERAL

Pág.

RESUMEN

ABSTRACT

INDICE DE FIGURAS Y TABLAS

### **INTRODUCCION**

1. INTRODUCCION

1.1 OBJETIVO

1.2 JUSTIFICACION

1.3 OBJETIVOS ESPECIFICOS

1.4 ORGANIZACIÓN DEL DOCUMENTO

1.5 REFERENCIA

### **MODELO DE SOLUCION**

2.1 CALCIFICACIONES

2.2 MODELO DE LA SOLUCION

2.3 BASE DE DATOS DE TRABAJO

2.4 RESUMEN

2.5 REFERENCIAS

### **PREPROCESAMIENTO**

3.1 IMAGEN DE TRABAJO

3.2 FILTRO DE ENTRADA

3.3 CORTE AUTOMATICO

3.4 RESUMEN CAPITULO

3.5 REFERENCIAS

## **EXTRACION DE CARACTERISTICAS**

- 4.1 TRANSFORMADA TOP HAT
- 4.1 FILTRO DE AJUSTE (ENHANCED)
- 4.3 BINARIZACION
- 4.4 RESUMEN CAPITULO
- 4.5 REFERENCIA

## **SELECCIÓN DE CARACTERISTICAS**

- 5.1 CALCIFICACIONES
- 5.2 TECNICAS DE SELECCIÓN
- 5.3 DISCRIMINADORES PROPUESTOS
- 5.4 RESUMEN
- 5.6 REFERENCIAS

## **CLASIFICACIÓN Y DIAGNÓSTICO**

- 6.1 GENERALIDADES

## **EXPERIMENTACION Y RESULTADOS**

- 7.1 PLATAFORMA DE TRABAJO
- 7.2 BASE DE DATOS
  - 7.2.1 MUESTRAS DE TRABAJO
- 7.3 EXPERIMENTACION EN EL PREPROCESAMIENTO
  - 7.3.1 FILTROS PROBADOS
  - 7.3.2 REDUCCION DEL ROI
    - 7.3.2.1 CORTE MANUAL
    - 7.3.2.2 CORTE AUTOMATICO
- 7.4 EXPERIMENTACION EN LA EXTRACION DE CALCIFICACIONES
  - 7.4.1 TECNICAS PROBADAS
  - 7.4.2 BINARIZACION

## 7.5 EXPERIMENTACION EN LA SELECCION DE CALCIFICACIONES

### 7.5.1 DESCRITORES SEGÚN BIRADS

### 7.5.2 TECNICAS PROPUESTAS

### 7.5.3 VECTOR DE CARACTERISTICAS

## 7.6 EXPERIMENTACION EN LA CLASIFICACIÓN Y EL DIAGNÓSTICO

### 7.6.1 TECNICAS PROPUESTAS

### 7.6.2 DISEÑO DEL ARBOL

### 7.6.3 RESULTADOS EN LA CLASIFICACIÓN

## 7.7 RESULTADO EN EL DIAGNÓSTICO

## **CONCLUSIONES**

### 8.1 FLUJOGRAMA FINAL

### 8.2 CONCLUSIONES

### 8.3 TRABAJO FUTURO

## **GLOSARIO**

## **ANEXOS**

1. ENFERMEDAD DEL CANCER: Historia, Patología, epidemiología, Diagnóstico, imagenología, BI-RADS.
2. TECNICAS DE SEGMENTACIÓN
3. TECNICAS DE CLASIFICACIÓN
4. TRABAJOS DESARROLLADOS PREVIOS – ESTADO DEL ARTE
5. Programa en MATLAB,
6. Referencia MIAS

## RESUMEN

Una de las principales enfermedades que ocasiona la muerte en el mundo es el CANCER y dentro de los diferentes tipos de cáncer el cáncer de Mama ocupa la segunda posición en la población mundial y la primera en las mujeres [8,10] . Dentro del proceso de diagnóstico se utilizan diferentes técnicas y la más utilizada es la mamografía, la cual consiste en una técnica no invasiva mediante la exposición del seno a un haz leve de rayos X generando una imagen digital en dos dimensiones la cual requiere de un análisis efectuado por personal especializado (radiólogos) quienes buscan dos tipos de información de carcinoma: Tumores (masas) y calcificaciones. Las calcificaciones o acumulaciones de calcio que se forman en los ductos de las glándulas mamarias acurde a su tamaño, formas y etiologías ofrecen diferentes propiedades en el diagnóstico de agrupaciones o clúster aislados y pequeños con una alta probabilidad de ser Malignas o Benignas

Este proyecto de tesis pretende aplicar algunas técnicas de procesamiento de imágenes y reconocimiento de patrones en la detección de carcinomas pequeños o **Calcificaciones** siguiendo el procesos de diagnóstico BI-RADS desarrollado por la Asociación Americana de radiología ARS. El trabajo desarrollado se basa en el modelamiento de una serie de algoritmos de procesamiento que actúan sobre imágenes utilizadas por la comunidad científica denominada: MIAS (*Mammographic Image Analysis Society*) la cual tiene 322 imágenes referenciadas, el procedimiento desarrolla una detección automática mediante los vectores morfológicos de **forma** y **tamaño** y la conformación de un vector de características basado en formas geométricas como lo son: área, área conexa, perímetro, solidez y excentricidad para la clasificación y diagnóstico de clasificaciones malignas y benignas.

## ABSTRACT

The use of CAD (Computer Added Detections) it's a one possible second option for radiology reader in early detections of mammography exams. This project used the protocol BIRADS for diagnosis's using image processing and classifications using the MIAS data base of 22 image of microcalcifications, this job search micro calcifications for diagnostic between benign and malign. The methodology used form and size descriptors for feature extraction, selections and classifications. The vector of feature used morphological descriptor how: area, convex area, perim, solidest and eccentricity. The project implement of classifications was decision tree, using solidest and eccentricity for node of decisions. In segmentations the morphological edge detector Top Hat has a good result.



## INDICE DE FIGURAS Y TABLAS

### **Figuras**

**pág.**

- Figura 1. Distribución de las micro calcificaciones de acuerdo a los Descriptores: forma, tamaño y agrupamientos.
- Figura 2. Flujo grama de la solución
- Figura 3. Distribución de las imágenes según MIAS, target de trabajo de 25 Imágenes
- Figura 4. Diagrama del sistema de investigación del grupo de BioIngenium
- Figura 5. Etapas en el preprocesamiento
- Figura 6. Vista de la calcificación en la imagen mdb227
- Figura 7. Aplicación de filtros a la imagen mdb227: A) imagen original, B) conversión a grises 8 Bits, C) filtro mediaño 3X3, D) filtro mediaño 5X5
- Figura 8. Resultado del corte automático en imagen mdb239, A) imagen original, B) imagen cortada.
- Figura 9. Resultado de la binarización de la imagen. A) imagen original, B) imagen binaria
- Figura 10. Resultado del proceso del corte automático
- Figura 11. Flujo grama del procedimiento de extracción de características
- Figura 12. Transformada top hat
- Figura 13. Operador ITF
- Figura 14. Ajustes de imagen con low\_in=0; low\_out=0; high\_out=1 y Variamos high\_in en: A) 0.05, B)0.1 , C)0.25 , D) 0.5
- Figura 15. Función de transferencia aplicada como filtro de enhancemet mediante La transformada de intensidad de grises.
- Figura 16. Ejemplos de aplicar el operador de binarización, usando los diagramas De la figura 48 y con un umbral de binarización de 0.275.
- Figura 17. Ejemplo grafico de cómo actúa un sistema de selección de característica, Acordes con un operador discriminador.
- Figura 18. Ejemplos del procedimiento visual que hace el especialista, imagen mdb219, Procesada y bina rizada aumentada en el ROI de la posible calcificación.
- Figura 19. Diferentes variables que son útiles en técnicas de selección de elementos En imágenes digitales.
- Figura 20. Calculo de perímetro y área en un objeto pixelado
- Figura 21. Ejemplo del diámetro máximo y mínimo en un objeto no circular
- Figura 22. Área convexa de una figura irregular
- Figura 23. Ejemplos de solidez en algunos objetos: A)0.9 B)0.85 C) 0.84 D)0.75 E)1.00
- Figura 24. Flujo grama de la etapa de clasificación y diagnóstico
- Figura 25. Población de imágenes del experimento.
- Figura 26. Exploración de la imagen mdb219 y su centroide de agrupaciones de Calcificaciones
- Figura 27. Ejemplo de calcificaciones localizadas en la imagen mdb219, total 11

## **Figuras**

**pág.**

- Figura 28. Bloques del pre-procesamiento
- Figura 29. Vista del tamaño del recorte manual 450x1024
- Figura 30. Resumen visual de la etapa de pre procesamiento sobre la imagen mdb227
- Figura 31. Flujo grama de la etapa de extracción de características
- Figura 32. Resultado de aplicar extracción mediante Sobel y Tophat en las imágenes de pruebas.
- Figura 33. Resultado de extracción sobre imágenes de prueba.
- Figura 34. Resultado de extracción sobre imágenes de prueba Malignas
- Figura 35. Mediante el análisis del tamaño y forma se etiquetan las Calcificaciones
- Figura 36. Áreas y rangos de caracterización
- Figura 37. Algoritmo de separación de calcificaciones pequeñas y grandes, para su Estudio posterior acorde con el discriminador de formas
- Figura 38. Imagen de salida de la etapa de selección de características, Imagen mdb227 se localizan 2 hallazgos posibles benignos y 3 hallazgos Posibles malignos.
- Figura 39. Imagen de salida de la etapa de selección de características, Imagen mdb222 se localizan 2 hallazgos posibles benignos y 1 hallazgo Posible maligno.
- Figura 40. Algoritmo propuesto para la clasificación
- Figura 41. Diseno del rango de los limites de decision en el clasificador
- Figura 42. Árbol de decisiones para el clasificador.
- Figura 43. Imagen de entrada y salida del sistema, imagen mdb227
- Figura 44. Flujo grama final del algoritmo implementado

## **ANEXO 1**

## **Figuras**

**pág.**

- Figura A1-1. División celular
- Figura A1-2. Papiro de Edwin Smith , El martirio de Santa Águeda
- Figura A1-3. Proceso y ciclo de crecimiento de un carcinoma
- Figura A1-4. Localizaciones de focos de cáncer
- Figura A1-5. Anatomía del seno
- Figura A1-6. Carcinoma lobulillar en situ
- Figura A1-7. Carcinomas ductal infiltrante:
- Figura A1-8. Incidencia de cáncer de mama ajustada a la edad en los EEUU
- Figura A1-9. Incidencia del cáncer de mama por edad en 4 países
- Figura A1-9.1. Etiología de las masas de seno según la edad
- Figura A1-10. Tipos de diagnóstico temprano para detección de carcinomas
- Figura A1-11. Dimensiones de carcinomas, tumor y micro carcinomas

## **Figuras**

**pág.**

- Figura A1-12. Esquema de clasificación del cáncer de mama según: tamaños, Áreas, densidad
- Figura A1-13. Relación entre verdaderos positivos (V+), verdaderos negativos (V-), Falsos positivos (F+) y falsos negativos (F-).
- Figura A1-14. Interconexión de sistemas de información con equipos de imágenes Conformando la tecnología de imagenología
- Figura A1-15. Elementos de un equipo mamógrafo
- Figura A1-16. Imágenes de equipos mamografos, estación de lectura
- Figura A1-17. Hallazgos en la mamografía
- Figura A1-18. Diferentes tipos de carcinomas de pequeño tamaño
- Figura A1-19. Flujo grama del proceso de tratamiento de la enfermedad de cáncer De mama
- Figura A1-20. Resultados del BIRADS según resultados objetivos: diagnóstico incierto, No existe ningún carcinoma, diagnóstico benigno, diagnóstico maligno.

## **ANEXO 2**

## **Figuras**

**pág.**

- Figura A2-1. Umbralización: (a) histograma de intensidades de grises en la imagen Mostrando los posibles umbrales (b) imagen en escala de grises
- Figura A2-2 Aplicación de la técnica de umbral sencilla: A) imagen original, B) umbral >0.25. C) umbral >0.5 D) umbral >0.75 E) umbral >8
- Figura A2-3 Modelo para detección de borde unidimensional y Bidimensional
- Figura A2-4 Ejemplo de un sistema para detección de bordes
- Figura A2-5 Diagrama del gradiente de una señal  $f(x)$ .
- Figura A2-6 Matriz  $z$ , en la que cada elemento tiene un nivel diferente de gris.
- Figura A2-7 Tabla de operadores según valores índices  $Z_i$  bajo las técnicas de Gradientes. Prewitt, Sobel, Kirsch, Roberts.
- Figura A2-8 Tablas operador laplaciano
- Figura A2-9 Ejemplos de aplicar operadores a una imagen. A) Roberts, B) Prewitt, C) Sobel
- Figura A2-10 Ejemplo del operador morfológico de dilatación
- Figura A2-11 Ejemplo del operador morfológico de erosión
- Figura A2-12 Otro ejemplo de cómo la erosión y dilatación separa elementos de Una imagen
- Figura A2-13 Ejemplo del operador morfológico de apertura
- Figura A2-14 Ejemplo del operador morfológico de cierre

## ANEXO 3

### **Figuras**

**pág.**

- Figura A3-1. Neurona y sus conexiones sinápticas
- Figura A3-2 Diagrama básico de una red RNA
- Figura A3-3 Diagrama de entrenamiento de una RNA
- Figura A3-4 Esquemas de redes neuronales
- Figura A3-5 Modelo de un Perceptron
- Figura A3-6 Modelo de una RNA - MLP
- Figura A3-7 Modelo de una RNA Hopfiled
- Figura A3-8 Modelo de una RNA función de base radial RBF
- Figura A3-9 Diagrama de función de decisión en lógica difusa
- Figura A3-10 Diagrama de mutación según algoritmos genéticos.
- Figura A3-11 Ejemplo de una red usando el algoritmo de *Bayes*.
- Figura A3-12 Tabla de probabilidad condicional en una red de bayes.
- Figura A3-13 Ejemplo de un árbol de decisiones.

## TABLAS

### **Tablas**

**pág.**

- Tabla 1. Resumen de imágenes según hallazgos y tejidos en MIAS.
- Tabla 2. Resumen de las calcificaciones en la base de datos MIAS
- Tabla 3. Algunas publicaciones de trabajos sobre el tema de procesamiento. De imágenes de mamografía en diferentes técnicas de procesamientos.
- Tabla 4. Imágenes escogidas para el experimento.
- Tabla 5. Tabla resumen de calcificaciones por imágenes
- Tabla 6. Parametrización en la etapa de segmentación
- Tabla 7. Resultados de prueba del clasificado sobre imagen mdb227
- Tabla 8. Resultados de las imágenes de entrenamiento
- Tabla 9. Resultados de las imágenes de pruebas
- Tabla 10. Eficiencia, FP, FN en los resultados de 20 imágenes en pruebas.
- Tabla 11. Porcentaje de error en los resultados
- Tabla 12. Resultado de una prueba diagnóstica y la ausencia o no de la enfermedad
- Tabla 13. Resumen de los índices de diagnóstico de las pruebas del experimento

### **Tabla de anexos**

- Tabla A1-1. Clasificación de BI-RADS Vs. Valor predictivo positivo VPP
- Tabla A3-1 Funciones de activación más utilizadas en RNA.



## INTRODUCCIÓN

### 1.1 INTRODUCCIÓN

En el mundo el cáncer de mama es un problema prioritario de salud pública, por ser la primera causa de muerte en la población femenina [1]. La incidencia del cáncer de mama se ha incrementado en las últimas décadas por diversos factores como son: cambios en los niveles de actividad física, la edad del primer parto, subida en los niveles del stress, baja calidad en los niveles de vida; aunque la incidencia es menor en países en desarrollo, las mujeres de países de menores ingresos, como los de América Latina tienen una menor supervivencia, puesto que un alto porcentaje de casos se diagnostican en estadios tardíos de la enfermedad [2]. Teniendo en cuenta el carácter multifactorial e incierto de la etiogenesis de la enfermedad se hace énfasis en la prevención secundaria antes que la primaria, es decir en la promoción del *diagnóstico temprano* de la enfermedad: el diagnóstico temprano junto con el tratamiento oportuno determina la supervivencia de las mujeres. En Estados Unidos la supervivencia de las pacientes a quienes se les diagnostica la enfermedad en estados (I y II) es del 89% en cinco años, mientras que la supervivencia de las mujeres con diagnóstico tardío (III, IV) es del 35%. En Colombia aun no se tienen las cifras reales debido a que hasta ahora se está implementado el plan de epidemiología en el Instituto Nacional de Cancerología, sin embargo los casos por año son muy elevados representando en las mujeres la segunda causa de muerte después del cáncer de cuello uterino.

La importancia en el diagnóstico temprano es la base de la prevención, la enfermedad del cáncer es una enfermedad muy compleja y en el caso de las mujeres requiere de auscultamiento cada año después de los 40 años y dos exámenes después de los 50 años. Este examen denominado análisis de mamografía está basado en la lectura de una imagen adquirida a través de un sistema de rayos X, la lectura e interpretación la efectúa un médico especializado en radiología de mama y requiere de una estación de trabajo de alta resolución en la cual se puedan visualizar masas y carcinomas de diferentes formas, tamaños, morfologías, tonalidades de grises y texturas. El principal problema está en la lectura dada por el radiólogo al denominar si una imagen presenta o no algún rasgo maligno o benigno con el cual en forma temprana se diagnostique y se trate. Tiempos típicos de lecturas de una doble imagen, la cual involucra los dos senos (izquierdo y derecho), no es menor a veinte minutos; un falso positivo ocasiona costos en el tratamiento y un falso negativo ocasiona dejar pasar la enfermedad y que esta madure y crezca, posiblemente ocasionando metástasis.

Una práctica usual en el protocolo de diagnóstico del radiólogo es una segunda lectura efectuada por un colega, esta de igual forma ocasiona mayores costos en el diagnóstico. Por ello se ha venido trabajando en desarrollar sistemas expertos de ayuda en la detección asistida denominada CAD (Computer Aided Detections) y es ahí en donde este trabajo de grado pretende profundizar en estudiar técnicas de procesamiento de imágenes médicas orientadas a la detección temprana de

carcinomas en imágenes del seno (imágenes de mamografía). La comunidad científica ha venido trabajando en este tema desde 1986 pero en la década de los 90 publicaciones de Betal [3], Zhao [4], Papadopou [5] y muchos más han aportado a esta temática, compañías multinacionales de igual forma han logrado desarrollar soluciones aprobadas por la FDA [6]:

- Kodak: se basa en un sistema que agrupa una solución completa en el tema de imagenología, desde la unidad que lee la imagen o mamógrafo, el sistema de almacenamiento o sistema PACS (picture archive communications systems), estaciones de lectura y diagnóstico, sistema CAD (computer aided detections).
- Hologic: [7] SecurView™ DX sistema especializado en la lectura y el diagnóstico automático e inteligente de mamografías.

## 1.2 OBJETIVO

Evaluar los descriptores morfológicos como son: Formas y Tamaños en forma automática como ayuda al protocolo BI-RADS en la detección de carcinomas en imágenes de mamografía digital

## 1.3 JUSTIFICACIÓN

Actualmente en Colombia son pocos los cupos en las especializaciones y residencias de radiología que ofrecen las Universidades en las diferentes facultades de medicina, es el caso del Hospital San Ignacio de la Pontificia Universidad Javeriana tan solo admiten a cuatro (4) estudiantes por semestre en la especialización de Radiología, con una duración de especialización de cuatro años. Cada día hay más número de casos por día y de diferentes especialidades, pocos de estos estudiantes se especializan en mamografía que es la técnica especializada en el cáncer de mama. Si bien una lectura para obtener un diagnóstico inicial puede durar entre 20 a 40 minutos por paciente en el día son muchos casos por diagnosticar. Un medico radiólogo utiliza como instrumento en el diagnóstico inicial la lectura de la imagen basado en una estación de trabajo con dos monitores de alta resolución en la gama de grises, además de un software de visualización en la cual puede hacer ciertos procesamientos locales como ayuda en: ampliar zonas de interés, efectuar filtrajes y en pocos casos técnicas de segmentación ( sobel, canny, Laplace), lo cierto es que después de un trabajo largo en el día la percepción visual no es la misma a la de la mañana y es posible no detectar visualmente puntos críticos que en este caso pueden ser relacionados a carcinomas. Para ellos la comunidad médica ha implementado protocolos de doble lectura y de compartir el diagnóstico ante dudas o no claridades en la visualización de la imagen entre colegas pero los costos del proceso de detección temprana se vuelve más costos al que actualmente es.

Es ahí en donde nace el proceso del diagnóstico el termino de falsos positivos (FP) y falsos negativos (FN), un falso positivo FP ocasiona continuar con la profundización en el diagnóstico cuando no es requerido, esto incrementa costos, además de exámenes innecesarios como lo son Biopsias, Ecografías o Resonancias del tipo MRI, contrario a un FN en el cual se descarta un posible carcinoma y puede continuar el procesos de crecimiento e incluso repoblamiento en metástasis de la enfermedad. Bien en el tiempo en la cual la mujer vuelva el siguiente año a su examen de rutina este carcinoma ya puede estar en estado crítico (haber crecido). Por ello la detección temprana es de suma importancia en el proceso de control y erradicación de la enfermedad, y un instrumento de apoyo en el diagnóstico puede ser un sistema inteligente de detección temprana o CAD (Computer Aided Detections).

Por ello un proyecto que aporte en el tema de detección temprana es una buena justificación y ayuda a los especialistas en reducir tiempos de lectura así como orientar en la localización de parecidos en la imagen., reducir los FP y FN.

Este proyecto solo actúa como un modelo aplicado a imágenes de referenciamiento, sin embargo en la industria biomédica de equipamientos de apoyo en lectura ya existen soluciones de sistemas en línea que interactúan en tiempo real con el especialista, sin embargo son soluciones muy costosas y cerradas no asequibles a ser masificadas sobre todo en los países de bajos recursos y en donde el índice de enfermedad por cápita es muy alto, caso el de Colombia.

#### **1.4 OBJETIVOS ESPECÍFICOS**

Los objetivos específicos del trabajo esperados son:

- Desarrollar las etapas de procesamiento de imágenes aplicadas al cáncer de mama (procesamiento de imágenes médicas) aplicando análisis morfológico: tamaños y formas dentro de un proceso automático tomando imágenes de una base de datos referenciada.
- Evaluar la detección según los descriptores (formas, tamaños) de los posibles carcinomas en un conjunto de datos previamente clasificados.
- Evaluar la efectividad de los descriptores para seguir al protocolo BI-RADS en el proceso de diagnóstico
- Trabajar sobre base de datos de imágenes referenciadas como lo son: MIAS, DDSM.



## 1.5 ORGANIZACIÓN DEL DOCUMENTO

A continuación hacemos una descripción breve de los contenidos en cada capítulo del documento.

Capítulo 1: Introducción, objetivos y alcances del proyecto.

Capítulo 2: Modelo de la solución planteada, base de datos de referenciamiento a usar, trabajos acordes con el tema en la comunidad científica.

Capítulo 3: Pre procesamiento, imágenes de trabajo, filtraje inicial, técnicas de corte.

Capítulo 4: Extracción de características: técnicas de segmentación, filtros de ajuste y binarización.

Capítulo 5: Selección de características, técnicas propuesta, formas geométricas y vector de características

Capítulo 6: Clasificación – concepto a usar.

Capítulo 7: Experimentación y resultados del trabajo. Se exponen cada etapa del flujo grama propuesta en el experimento.

Capítulo 8: Conclusiones del trabajo y trabajo futuro.

Glosario

Anexos:

- ENFERMEDAD DEL CANCER: Historia, Patología, epidemiología, Diagnóstico, imagenología, BI-RADS.
- TECNICAS DE SEGMENTACIÓN
- TECNICAS DE CLASIFICACIÓN
- TRABAJOS DESARROLLADOS PREVIOS – ESTADO DEL ARTE
- Programa en MATLAB,
- Referencia MIAS

## 1.5 REFERENCIAS

1. [1] ferlay j, bray f, pisani P. DM. GLOBOCAM 2000: cáncer incidence, mortality and prevalence worldwide.
2. [2] Carolina Wiesner, coordinadora del grupo de planeación y gestión del programa de prevención del Instituto nacional de Cancerología, Revista colombiana de cancerología, Volumen 11, pág. 14, 2007.
3. [3] D. Betal, N Roberts, segmentations and numerical analysis of microcalcifications on mammograms using mathematical morphology. The British Journal of radiology, (70), 1997
4. [4] D. Zhao, M. Shridhar, and D.G. Daut. Morphology on detections of calcifications in mammograms, IEEE international conference on acoustic speech, and signal processing ICASSP-92, volume 3, pag 129-132, 1992.
5. [5] A. Papadopoulosa, D.I. Fotiadisb, and A. Likasb. An automatic microcalcifications detections system based on a hybrid neural network classifier. Artificial Intelligence in medicine, (25):149-167, 2002.
6. [6] FDA es la Asociación Americana de Administración de Alimentos y Medicamentos, es la agencia del gobierno de los Estados Unidos responsable de la regulación de alimentos (tanto para seres humanos como para animales), suplementos alimenticios, medicamentos (humanos y veterinarios), cosméticos, *aparatos médicos (humanos y animales)*, productos biológicos y productos eméticos, 2010.
7. [7] Hologic, Inc. is a leading developer, manufacturer and supplier of premium diagnostic products, medical imaging systems, and surgical products dedicated to serving the healthcare needs of women throughout the world, Hologic suite of products spans nine key areas: mammography and breast biopsy; radiation treatment for early-stage breast cancer; cervical cancer screening; treatment for menorrhagia; permanent, 2010.

## 2.0 MODELO DE SOLUCIÓN

De acuerdo a lo presentado en el **anexo 1**. Continuamos definiendo las características de las calcificaciones con miras a cuantificar las herramientas a utilizar en el proyecto.

### 2.1 CALCIFICACIONES

Debemos especificar la morfología de las calcificaciones para efectos de definir los límites de búsqueda en los algoritmos a desarrollar en este trabajo y así poder seleccionar los parámetros que se utilizarán en el vector de características basado en los descriptores de forma y tamaño. Recordemos que en este trabajo solo se trabajaran carcinomas péquenos o **microcalcificaciones** por ser los de mayor dificultad en la lectura y diagnóstico.

Las calcificaciones son pequeños depósitos de mineral de calcio que se presentan como puntos blancos en la mamografía. Muchas mujeres tienen al menos una calcificación que puede ser observada. Las calcificaciones no se pueden sentir a la palpación y se encuentran de preferencia en una mamografía. Las calcificaciones se analizan según su tamaño, forma, número, y distribución (ver figura 1). Algunas reglas generalmente aceptadas en relación a las calcificaciones, son las siguientes [1, 2, 3, 4] :

1. Las calcificaciones grandes, redondas u ovales y uniformes en su tamaño, tienen mayor probabilidad de estar asociadas a procesos benignos, pero las calcificaciones pequeñas, irregulares, polimórficas, con ramificaciones y de morfología heterogénea están asociadas con procesos malignos.
2. Las calcificaciones aisladas tienen poca probabilidad de ser malignas, pero cuando se presentan agrupadas tiene alta probabilidad de ser malignas.
3. Las calcificaciones de menos de 0.5 milímetros de diámetro tienen alta probabilidad de ser malignas.
4. Las calcificaciones de más de 1 milímetro de diámetro tienen alta probabilidad de ser benignas.
5. Las microcalcificaciones aparecen como pequeñas manchas o puntos blancos en la imagen. Se estudia su forma y tamaño para tomar decisiones sobre su etiología, su diámetro es menor o igual a 1.0 mm. Tienen cierta probabilidad de ser malignas en especial si forman grupos de tres o más en áreas menores a 1.0 cm<sup>2</sup>. (Ver figura A1-11 y A1-12 del anexo 1).
6. Las macrocalcificaciones aparecen como puntos blancos grandes en la imagen, son depósitos de calcio voluminosos. Pueden corresponder a cambios degenerativos en la mama tales como envejecimiento de arterias, lesiones

anteriores o inflamaciones. Por lo general son benignas, cerca del 50% de las mujeres mayores de 50 años la tienen, su tamaño es de más de 1.0 mm. Los límites mencionados son referenciales.

La ACR distingue tres (3) categorías de calcificaciones (figura A1-18 en el anexo 1):

- Típicamente benignas once (11) tipos:
  - Cutáneas o de piel
  - Vasculares
  - Grosera
  - En vara larga
  - Redondeada (>1mm)
  - Puntuada (< 0.5 mm)
  - Esférica o centrolucida
  - Aro o en cascara de huevo
  - Calcificación láctea
  - Sutura
  - Distrofica
- Intermedias: Amorfas o indistintas
- Probablemente malignas: Pleomorfas o Heterogéneas , Finas y/o ramificadas

En resumen podemos visualizar la clasificación de las microcalcificaciones acordes con los descriptores así:

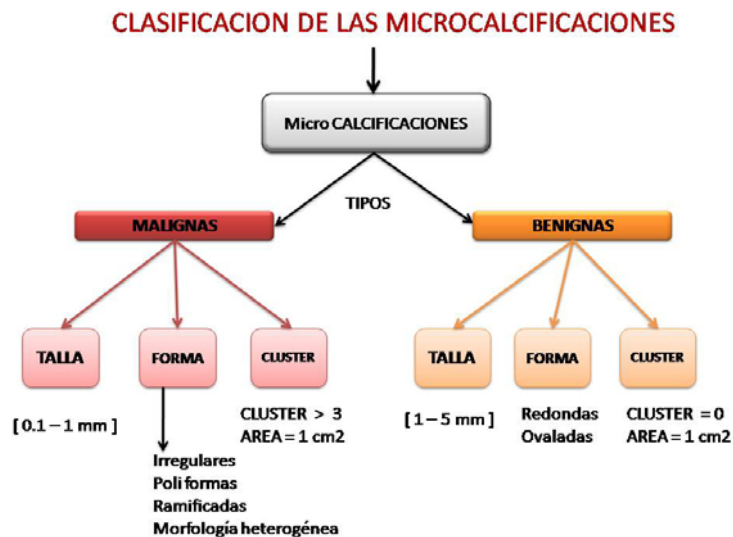


Figura 1. Distribución de las microcalcificaciones de acuerdo a los descriptores: forma, tamaño y agrupamiento

## 2.2 MODELO DE LA SOLUCION

El modelo de solución es el procedimiento que se siguió para resolver los objetivos planteados en el anteproyecto como problemas a resolver. La cual se divide en dos secciones: una de sintonización y entrenamiento.

1. Procesamiento de la imagen: incluye las etapas de lectura de la base de datos, preprocesamiento, extracción de características.
2. Reconocimiento de patrones: incluye las etapas de selección de características, clasificación y análisis.
3. Sintonización y entrenamiento del sistema.

Los pasos son sistemáticos ordenados acorde con el flujo grama que se muestra en la figura 2:

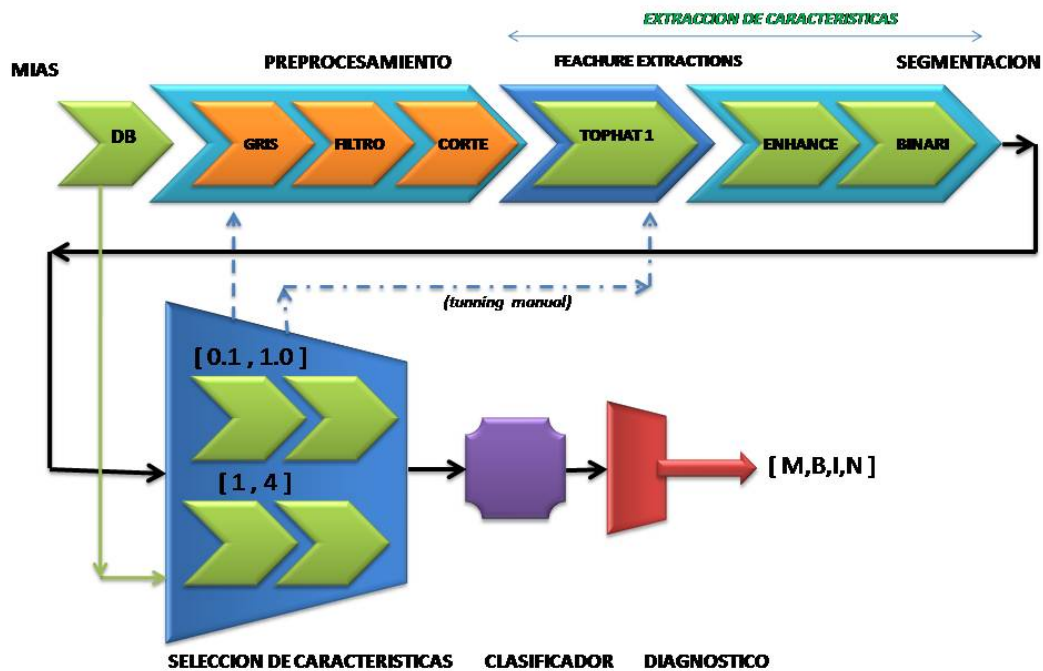


Figura 2. Flujo grama de la solución

- **Pre procesamiento:** en esta etapa se lee la imagen de la base de datos, se convierte a la gama de grises en 256 niveles, se limpia mediante el filtro mediaño para quitar ruido de alta frecuencia y finalmente se hace un corte automático de la imagen. El objetivo es tener una imagen con un data limpio y maximizando el ROI (región de interés) de la imagen a procesar y clasificar.

- **Extracción de Características:** se exponen las diferentes técnicas de segmentación con miras a presentar la técnica TOP HAT con la cual se experimento, se hace un filtro de ajuste de la gama de grises de la imagen mediante técnicas de Enhancement y se binariza la imagen para efectos de tener un realce total de puntos brillas que se estudiaran como posibles calcificaciones.
- **Selección de características:** en esta etapa se procesa la imagen en búsqueda de de las calcificaciones acorde con su tamaño y forma mediante las diferentes técnicas de morfología en magnitudes geométricas: área, perímetro, centroide, circularidad. Así poder tener la data requerida para la siguiente etapa de clasificación y posterior diagnóstico.
- **Clasificación:** esta etapa expone las diferentes técnicas sobre clasificación de patrones y para ello se trabajara con el data de los descriptores tamaño y forma ya seleccionados en la etapa anterior (selección de características), se utilizará la técnica de árbol de decisiones acorde con el formato BIRADS para lograra el diagnóstico esperado.
- **Diagnóstico:** esta etapa seleccionará las 4 opciones esperadas del proceso: BIRADS: 0, 1-2, 3, 4.

### 2.3 BASE DE DATOS DE TRABAJO

En el desarrollo de todo trabajo que realiza análisis y reconocimiento de patrones lo más importante es tener una base de datos bien documentada, en el procesamiento de imágenes medicas el mayor problema con el que se encuentra el investigador es obtener imágenes reales, por derechos de autor toda imagen debe ser autorizada por el paciente quien es el dueño de su diagnóstico y de su historia clínica, esta incluye las imágenes que de hecho acorde con el formato DICOM quedan marcadas con el nombre y datos principales del paciente. Por ello se ha recurrido a bases de datos de dominio público que la comunidad científica ha ido documentando y referenciando para trabajos y estandarización de las investigaciones.

Las dos bases de datos de mayor uso en el área de imágenes de mamografía son:

1. **DDSM:**[5] Digital Database for Screening Mammography es una serie de imágenes acumuladas por la Universidad del Sur de la Florida (USF) preparadas para la comunidad de investigadores. La base de datos tiene 2620 casos (imágenes) disponibles en 43 volúmenes. Estas imágenes fueron obtenidas en el MassaChusetts General Hospital, Wake Forest University School of medicine, Sacred Heart Hospital. Las imágenes dentro de su referenciamiento usan el BIRADS, las imágenes tiene una resolución de 8 y 16

bits en formato .PGM y .TIFF, en este trabajo no utilizaremos esta base de datos.

2. **MIAS.** Esta base de datos fue desarrollada por (The Mammographic Image Analysis Society) en el Reino Unido [6] con miras a ser utilizadas por la comunidad como referencias para estudios de análisis y procesamiento de imágenes digitales. La base de datos está compuesta por 322 imágenes (tabla 1):

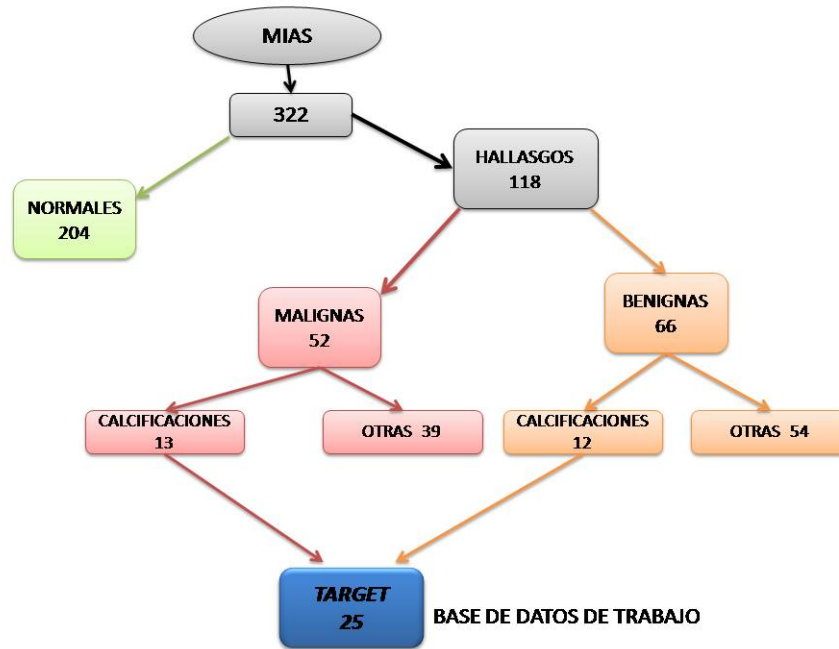


Figura 3. Distribución de las imágenes según MIAS, target de trabajo de 25 imágenes.

Total imágenes:	322 (ver figura 3)
Imágenes Normales:	204
Imágenes con hallazgos:	118
Benignas:	66
Malignas:	52

Los hallazgos que se pueden encontrar son calcificaciones (25 casos), masas circunscritas (20 casos), masas espiculadas (21 casos), masas mal definidas (15 casos), distorsión arquitectural (20 casos) y asimetría (17 casos). De las 322 imágenes, 25 contienen al menos una calcificación, la cuales era nuestro target de trabajo en este proyecto, de estas doce (12) son benignas y trece (13) son malignas., en la tabla 1 se resumen los diferentes tipos de casos.

Las imágenes fueren digitalizadas a 50 upixel y a 200 upixel y están en formato PGM 1024 x 1024 .PGM. En este trabajo se usan imágenes de 200 u/pixel. La base de datos de alta resolución (50u/pixel) puede ser adquirida y la base de datos de baja resolución puede ser bajada desde el servidor de la MIAS [5] .

Hallazgo	Tipo de tejido								
	Graso		G.Glandular		Denso		Totales		
	B	M	B	M	B	M	B	M	Total
calcificaciones	2	4	5	4	5	5	12	13	25
masas circunscritas	7	2	6	2	3	0	16	4	20
masas espiculadas	2	4	4	3	6	2	12	9	21
masas mal definidas	4	4	3	2	1	1	8	7	15
distorsión arquitectural	4	2	2	4	4	4	10	10	20
asimetrías	2	3	3	2	3	4	8	9	17
normal	66		64		74		204		204
subtotal	87	19	87	17	96	16	270	52	322
totales	106		104		112		322		322

Tabla 1 [1]. Resumen de imágenes según hallazgos y tejidos en MIAS.

El formato que MIAS documenta y referencia la base de datos de la siguiente forma:

**Primera columna:** nombre de referencia de la imagen.

**Segunda columna:** tejido de fondo:

F Grueso o grasoso

G Grueso-granular

D Denso-granular

**Tercera columna:** Clase de anormalidad presente:

CALC Calcificación

CIRC Masa bien definida/circuns

SPIC Masas de contorno irregular

MISC Otras masas definidas (malas)

ARCH Distorsión de arquitectura

ASYM Asimetría

NORM Normal



**Cuarta columna:** severidad de la ańormalidad.

B Benigna

M Maligna

**5th, 6th columna:** x,y coordenada del centro de la ańormalidad.

**7th columna:** Radio aproximado (in pixel) del circulo que encierra la ańormalidad.

#### FORMATO REFERENCIAMIENTO

1	2	3	4	5	6	7
=====						
mdb001	G	CIRC	B	535	425	197
mdb002	G	CIRC	B	522	280	69
mdb003	D	NORM				
mdb004	D	NORM				
mdb005	F	CIRC	B	477	133	30
mdb005	F	CIRC	B	500	168	26
mdb006	F	NORM				
mdb007	G	NORM				
mdb008	G	NORM				
mdb009	F	NORM				
mdb010	F	CIRC	B	525	425	33
mdb011	F	NORM				
mdb012	F	CIRC	B	471	458	40
mdb013	G	MISC	B	667	365	31

El resumen acorde con las calcificaciones en la base de datos la podemos ver en la siguiente tabla (tabla 2), es realmente las imágenes de referenciamiento que se utilizarán en este proyecto.

No	Imagen	Tejido	Lesión	Dx	X	Y	Radio
1	mdb209	G	CALC	M	647	503	87
2	mdb211	G	CALC	M	680	327	13
3	mdb212	G	CALC	(B)	687	882	10
4	mdb213	G	CALC	M	547	520	45
5	mdb214	G	CALC	(B)	582	916	11
6	mdb216	D	CALC	M			
7	mdb218	G	CALC	B	519	629	8
8	mdb219	G	CALC	B	546	756	29
9	mdb222	D	CALC	B	398	427	17
10	mdb223	D	CALC	B	523	482	29
	mdb223	D	CALC	B	591	529	6
11	mdb226	D	CALC	B	287	610	7
	mdb226	D	CALC	B	329	550	25
	mdb226	D	CALC	B	531	721	8
12	mdb227	G	CALC	B	504	467	9
13	mdb231	F	CALC	M	603	538	44
14	mdb233	G	CALC	M			
15	mdb236	D	CALC	B	276	824	14
16	mdb238	F	CALC	M	522	553	17
17	mdb239	D	CALC	M	645	755	40
	mdb239	D	CALC	M	567	808	25
18	mdb240	D	CALC	(B)	643	614	23
19	mdb241	D	CALC	M	453	678	38
20	mdb245	F	CALC	M			
21	mdb248	F	CALC	B	378	601	10
22	mdb249	D	CALC	M	544	508	48
	mdb249	D	CALC	M	575	639	64
23	mdb252	F	CALC	B	439	367	23
24	mdb253	D	CALC	M	733	564	28
25	mdb256	F	CALC	M	400	484	37

Tabla 2 [1, Pág. 135.] Resumen de las calcificaciones en la base de datos MIAS

## 2.4 RESUMEN CAPITULO

En este capítulo revisamos la metodología de trabajo acorde con un modelo propuesto, revisamos también la presentación de la base de datos que se utilizará **MIAS** como fuente de referenciamiento y material de entrada al programa. De igual forma se reviso la literatura y aportes que se han hecho sobre el tema, los diferentes grupos de investigación que están trabajando en la misma problemática, resaltamos la del grupo de la Universidad nacional de Colombia quien ya ha tiene un producto definido y en proceso de maduración.

## 2.5 REFERENCIAS

[1] Samuel Oporto Días, detección automática de agrupaciones de microcalcificaciones en mamografías digitalizadas, tesis, Instituto tecnológico de monterrey, Diciembre 2004.

[2] Dr. Don Antonio Albarracín Marín-Blazquez, Microcalcificaciones mamarias: Evaluación y Manejo, Hospital general Universitario de Murcia Reina Sofía. <http://scgd3murcia.iespana.es/scgd3murcia/microcalcificacionesmama...>

[3] Kopans, Daniel B., Breast imaging. ISBN: 0-397-51302-x. paginas: 300-323, Lippincott-Raven Publishers, 1998.

[4] Winchester, Hudis, Norton. Breast Cancer. 2d editios, ISBN: 1-55009-272-3, 2006.

[5] DDSM, South Florida University.

[6] MIAS, <http://peipa.essex.ac.uk/info/mias.html>

[7] Mislav Grgic, Kresimir Delac & Jelena Bozek, University of Zagreb, Faculty of Electrical Engineering and Computing - Department of Wireless Communications, Video Communications Laboratory: <http://www.mammoimage.org/general-info/>  
Grupo de investigacion: Mammographic Image Analysis Research Community (scientific and technical discussions on mammographic image analysis, digital mammography, computer-aided detection and diagnosis - CAD).

[8] grupo de investigación BioIngenium de la Universidad nacional de Colombia, director: Eduardo Romero, investigador Líder: Fabián Narváez.

<http://www.bioingenium.unal.edu.co/pagpro.php?idp=mamografias&lang=es&linea=3>

[9] <http://spie.org/>

### 3.0 PREPROCESAMIENTO

La etapa de preprocesamiento es la primera etapa en la metodología en un sistema de procesamiento y reconocimiento de imágenes, su principal objetivo es lograr una limpieza de la imagen, eliminando elementos, filtrando señales no deseables también denominadas ruido que por lo regular toda imagen fuente trae dentro del proceso de captura o de migración de etapas previas. En nuestro caso es eliminar efectos que puedan distorsionar en la imagen la claridad en las calcificaciones en la imagen de mamografía, de igual forma reducir el área de trabajo, se estima que más del 50% de la imagen original no es información relevante de la mama que si puede generar tiempos y costos en el procesamiento.

El flujo grama de en esta etapa de pre procesamiento (ver figura 5) está conformado por tres procedimientos: filtraje de la imagen, imagen binaria y corte automático. A esta etapa ingresa la imagen I en formato .pgm tal como se tiene de las 25 imágenes dichas en la sección 2.3 y la salida es una imagen recortada y filtrada en la gama de gris reduciendo los umbrales de saturación.

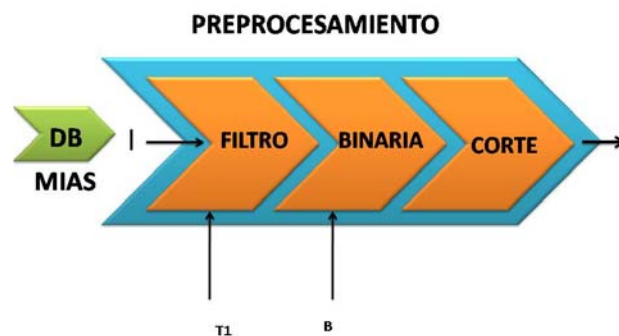


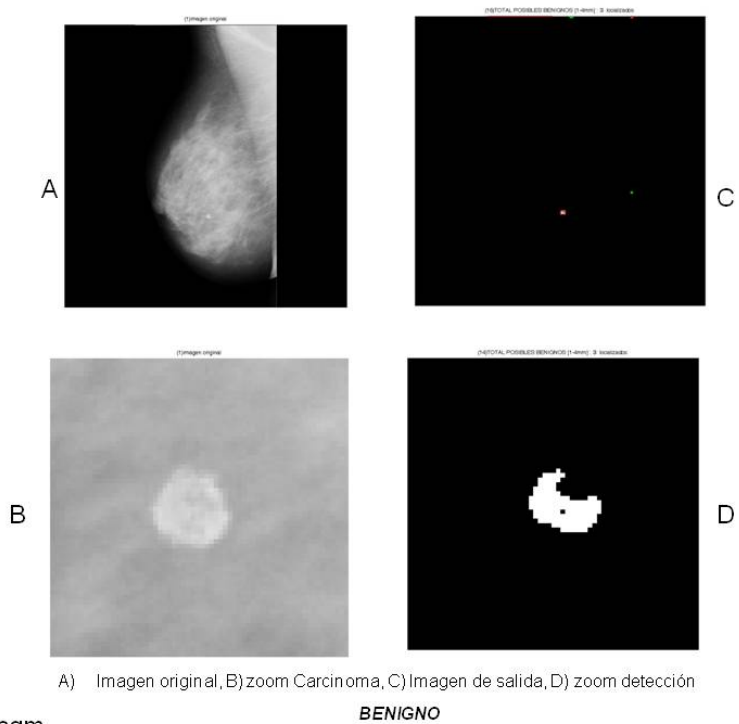
Figura 5. Etapas en el preprocesamiento.

### 3.1 IMAGEN DE TRABAJO

Como se menciona en la sección 2.3 y en la tabla 1 las imágenes son de 1024 x 1024, durante el trabajo del proyecto se utilizaron visualizadores de imágenes (viewers) con miras a localizar y familiarizarse con la información que la base de datos da sobre el referenciamiento de cada imagen según el formato BIRADS que tiene cada una de ellas. En la figura 6 se aprecia la localización de la calcificación de la imagen mdb227.

1            2 3        4 5        6 7  
 =====  
 mdb227 G CALC B 504 467 9

Imagen:            227  
 Tipo de tejido:    Grueso-Granular  
 Añormalidad:     calcificación  
 Diagnóstico:      Benigno  
 Coordenadas:     [ 504 , 467 ]  
 Radio:             9



mdb227.pgm

Figura 6. Vista de la calcificación en la imagen mdb227

## 3.2 FILTRO DE ENTRADA

Debido al nivel de ruido de alta frecuencia que puede presentar las imágenes de mamografía se requiere el uso de un filtro que se acople a reducir las señales de alta frecuencia, para ello se utilizó el filtro mediaño el cual es un filtro no lineal que tiene una muy buena propiedad al no eliminar las características significativas de la imagen.

### 3.2.1 CONVERSION A ESCALA DE GRISES

La entrada de la imagen el proceso está dado por la lectura de la base de datos MIAS, en este caso se lee la imagen mdb227 y se almacena en la variable im00 (imagen 00):

```
im00 = imread('mdb227.pgm');
```

La imagen original puede venir con una gama alta de grises, no es conveniente trabajar más de 256 niveles de gris si bien en este trabajo no se está reconociendo ni procesando el descriptor de textura de la imagen por ello se convierte a 8 bits la imagen lo cual es equivalentemente hablar de 256 niveles de gris por cada pixel. Bajo MatLab el tool box sería:

```
im01 = uint8 (im00);
```

### 3.2.2 FILTRO MEDIANA

Debido al nivel de ruido de pequeña magnitud del tipo salt y pimienta ocasionado por la digitalización de las imágenes del mamógrafo al CR se requiere pasar la imagen por un filtro del tipo mediaño

Se usa una máscara de 3 x 3 la cual se centra en cada pixel de la imagen, reemplazando el pixel central por la media de los nueve pixeles que cubren la máscara. El tamaño de la máscara permite que se mantenga las características propias de la imagen y a la vez se eliminen las altas frecuencias o ruido que pueda traer la imagen en el proceso de digitalización del equipo de Rayos X. en la figura 7 se muestra el proceso de usar ventana de 5x5 y 3x3.

```
fml = im01 ;  
f1 = medfilt2 (fml , [3 3] , 'symmetric');
```

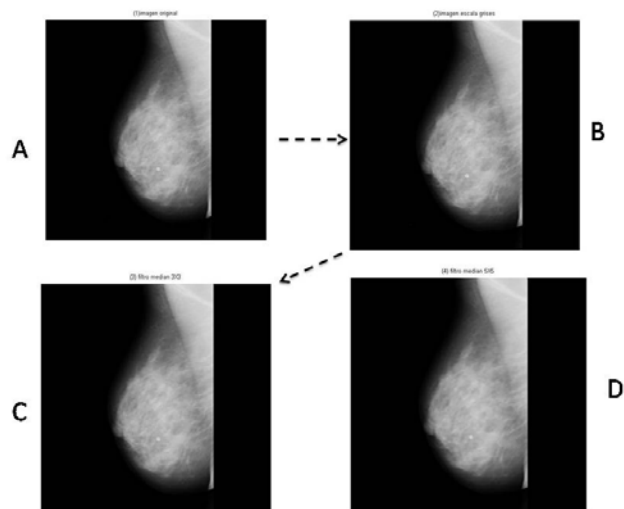


Figura 7. Aplicación de filtros a la imagen mdb227: A) imagen original, B) conversión a grises 8 Bits, C) filtro mediano 3X3, D) filtro mediano 5X5

### 3.3 CORTE DE IMAGEN

Las imágenes análogas tomadas en los mamógrafos análogos, dentro del proceso de captura de la imagen y grabada en el film, quedan plasmadas varias marcas que en el proceso de análisis de la imagen es ruido en la detección de características. Es típico visualizar marcas del nombre de la persona, marca del lado de la mama, espurias de bordes de la imagen sobre el lado del tórax.

El corte automático (figura 8) es un procedimiento que permite reducir información no relevante de la imagen filtrándola y acotando el tamaño de la misma, en otras palabras resaltando la región de interés máxima ROI. El ruido o señales no deseables se cambian por oscuro total o puntos negativos, dejando una imagen mejorada la cual permitirá una mejor eficiencia en costos de procesamiento y tiempos de ejecución del algoritmo.

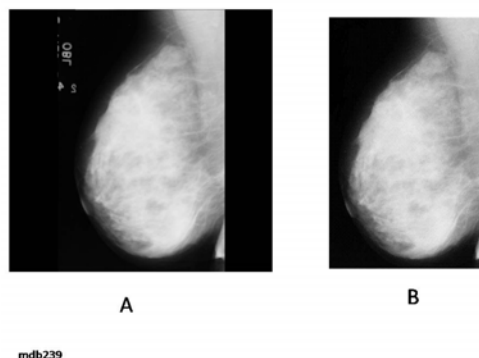


Figura 8. Resultado del corte automático en imagen mdb239, A) imagen original, B) imagen cortada.



El corte de la imagen se puede lograr mediante un proceso manual mediante el uso de una herramienta viewer de imágenes como lo es: *paint*, *XnView*, *CS4 photoshop* entre muchos otros.

Mediante un procedimiento automático se puede lograr efectuando los siguientes los siguientes pasos:

- Binarización de la imagen
- Corte automático

### **Binarización:**

El algoritmo para la binarización sigue el siguiente procedimiento [1]: sobre la imagen de entrada al proceso la cual ya viene del filtro mediana como *fm1* la pasamos a ser la imagen *I*:

$I = fm1;$

El algoritmo debe procesar solo la imagen que corresponde a la mama, que es el área ROI de trabajo, se creará una imagen binaria  $I'$  (ver figura 9) definida como el espejo de la imagen *I*. en esta imagen espejo cada pixel representa una ventana de 16 x 16 Si el promedio de gris de la ventana es mayor o igual que *promedio mínimo* su valor es uno (1), en caso contrario es cero (0). El procedimiento se ejecuta realizando los siguientes pasos (ver Pseudocódigo 4.1):

Pseudocódigo 4.1:

*I*: imagen de entrada al proceso  
Lado ventana  
Promedio Mínimo

***Procedure: [B] = imagen binaria (I, lado ventana, promedio mínimo)***

1.  $B = \text{crear imagen binaria}(\text{filas } I / \text{lado ventana, columnas } I / \text{lado ventana})$
2. desde  $i = 1$  hasta  $i < \text{filas } I * \text{columnas } I$
3. promedio = *promedio gris*( $I[i]$ )
4. si (promedio)  $\geq$  promedio mínimo) entonces  $B(i) = 1$
5. si (promedio)  $<$  promedio mínimo) entonces  $B(i) = 0$
6. fin desde

1) Se genera una imagen espejo de la imagen original *I*, esta será la nueva imagen *B*. En donde cada pixel representa una ventana de 16 X 16 de la imagen original, el tamaño de la imagen *B* es:

$B = [\text{filas imagen/lado ventana, columnas imagen/lado ventana}]$

2) Se calcula el promedio de niveles de gris de cada ventana y si este es mayor que el umbral elegido, se coloca 1 en  $B$  en la posición correspondiente (ver la figura 9). No existe diferencia significativa en los resultados si es que se eligen ventanas de  $8 \times 8$  o de  $16 \times 16$ . El umbral se ha elegido mediante el desarrollo de experimentos; valores bajos del umbral puede incluir regiones sin interés, valores altos del umbral puede eliminar algunas regiones de interés, en especial los bordes de la mama; en este caso el umbral es 15 (256 niveles de gris) valor sacado mediante prueba y error.

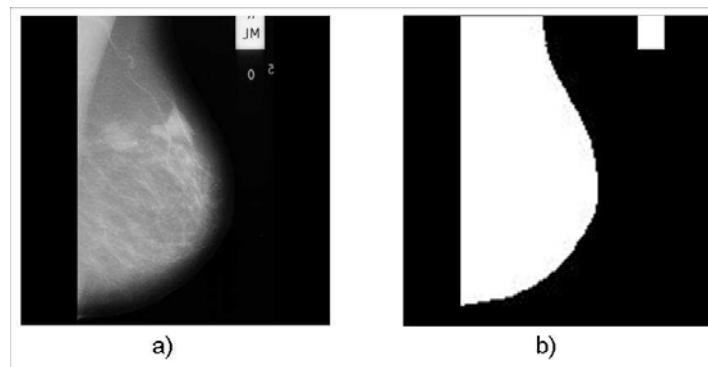


Figura 9 [1]. Resultado de la binarización de la imagen. A) Imagen original, B) imagen binaria.

### **Corte Automático:**

El procedimiento se ejecuta realizando los siguientes pasos (ver Pseudocódigo 4.2):

- 1) Para eliminar elementos aislados en  $B$ , se calcula su centroide, bajo el supuesto de que la mama contiene a su centroide (la mama es un objeto convexo).
- 2) A partir del centroide se buscan todos los píxeles conectados en vecindad  $V_4$  (Cuando los píxeles adyacentes tienen un lado en común). Este último procedimiento permite extraer solamente la región conectada que contiene a la mama (figura 9), obteniéndose la imagen binaria  $B'$ .
- 3) Para reconstruir la imagen de la mama  $I$ , se recorre la imagen binaria  $B'$ . Si el píxel visitado es 1, se copia la ventana correspondiente de  $I$  a  $I'$ , quedando el resto de las regiones en cero.
- 4) para el corte vertical, se acumula verticalmente todos los píxeles en  $B'$ , obteniéndose un vector donde cada posición indica cuántos píxeles existe en la columna correspondiente, a continuación se hace una búsqueda desde los extremos, se extrae la posición donde se ubica el primer píxel con valor diferente de cero tanto a la izquierda como a la derecha (figura 10). A continuación se ejecuta el corte considerando las regiones entre estas dos posiciones. La misma operación se sigue

para el corte horizontal. Este último procedimiento se aplica tanto a la imagen binaria ( $B'$ ) como a la imagen original ( $I$ ), obteniéndose finalmente los resultados como se muestran en la figura 8.

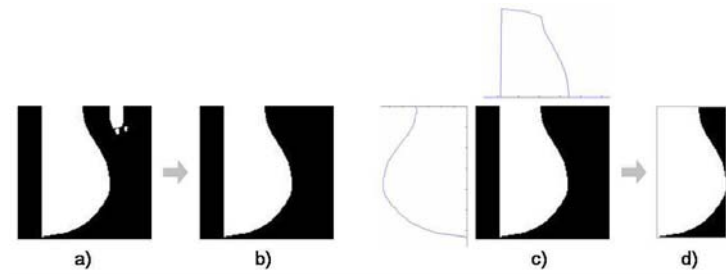


Figura 10 [1]. Resultado del proceso del corte automático: A) imagen binaria espejo, B) imagen binaria sin marcas externas a la mama, C) vistas de los extremos para activar cortes, D) Imagen recortada binaria (horizontal y vertical).

Pseudocódigo 4.2:

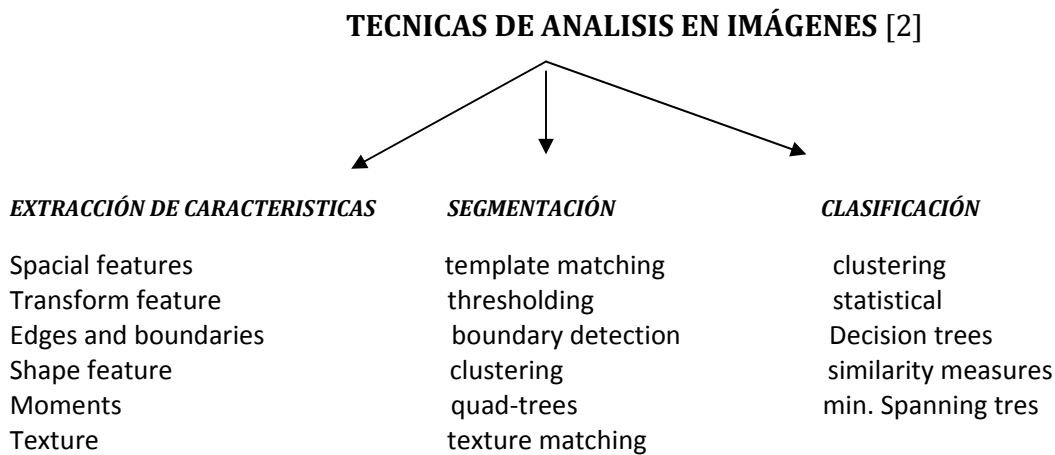
Procedure:  $[I', B'] = \text{corte\_automatico}(I, B)$

1.  $CM = \text{centroide}(B)$
2.  $B' = \text{pixeles\_conectados}(B, CM)$
3. desde  $i = 1$  hasta  $i < \text{pixeles } B$
4.     si  $(B'[i]) = 1$
5.          $I'[i] = I[i]$
6.     fin si
7. fin desde
8.  $acumV = \text{acumular\_pixeles\_vertical}(B')$
9.  $minC = \text{index\_diferente\_cero\_izquierda}(acumV)$
10.  $maxC = \text{index\_diferente\_cero\_derecha}(acumV)$
11.  $I' = \text{cortar\_verticalmente}(I', minC, maxC)$
12.  $B' = \text{cortar\_verticalmente}(B', minC, maxC)$
13.  $acumH = \text{acumular\_pixeles\_horizontal}(B')$
14.  $minF = \text{index\_diferente\_cero\_izquierda}(acumH)$
15.  $maxF = \text{index\_diferente\_cero\_derecha}(acumH)$
16.  $I' = \text{cortar\_horizontalmente}(I', minF, maxF)$
17.  $B' = \text{cortar\_horizontalmente}(B', minF, maxF)$

### 3.4 RESUMEN CAPITULO

En este capítulo hemos desarrollado las etapas del pre procesamiento de la imagen, a pesar que las imágenes de la base de datos MIAS, escogida para este trabajo; viene en un alto porcentaje limpia en su proceso de digitalización. Hemos querido desarrollar un procedimiento estándar para que cualquier imagen que pueda tener distorsiones y ruido se beneficie de las técnicas de filtraje y corte automático. El corte de imagen es un procedimiento muy importante porque recorta información no relevante y permite que la región de interés (mama) quede en su mayoría en la imagen a procesar; así se entrega una imagen más compacta a la siguiente etapa del proceso global del proyecto (la Segmentación).

En los capítulos siguientes: 4, 5, 6 veremos las diferentes técnicas en el análisis de la imagen acorde con este flujo grama:



### 3.5 REFERENCIAS

[1] Samuel Oporto Días, detección automática de agrupaciones de microcalcificaciones en mamografías digitalizadas, tesis, Instituto tecnológico de monterrey, Diciembre 2004. Paginas: 41 - 44

[2] Anil K. Jain, Fundamentals of digital image processing, Prentice-Hall international, Inc. 1989.

#### 4. EXTRACCIÓN DE CARACTERÍSTICAS

En este capítulo desarrollaremos las diferentes técnicas de segmentación, con miras a localizar las señales de interés para posteriormente ser separadas del conjunto y ser valoradas. Tal vez en esta sección es en donde la ciencia del procesamiento de imágenes más ha trabajado y aportado técnicas a la comunidad.

La segmentación o detección de objetos es el proceso por el cual los píxeles pertenecientes a una imagen son divididos en clases, normalmente dos: objetos de interés y fondo. El fin último de la segmentación es el reconocimiento de objetos y en muchos casos su cuantificación. Objetos de interés pueden ser los núcleos en una imagen histológica de un tejido, microorganismos en una muestra de plancton, las hojas de una planta, un tumor y/o calcificaciones en una imagen de mamografía, en la figura 26, observamos el flujo grama de las etapas que conforman este procedimiento desarrollado en este proyecto.



Figura 11. Flujo grama del procedimiento de extracción de características

En la figura 11 se presentan las etapas típicas de un procedimiento a usar en la extracción de características, básicamente se divide en dos secciones: la segmentación y de nuevo un procesamiento de la señal en donde se convierte a imagen binaria para su trabajo más fácil y de poca carga computacional. En el anexo 2 se expande las diferentes técnicas de segmentación que se utilizaron en el desarrollo del este proyecto y que permitieron escoger a la transformada Top hat.

#### 4.1 TRANSFORMADA TOP HAT

Consisten en combinar filtros de máximo y mínimo y a la imagen resultante sustraerle la imagen original o de la imagen original sustraer la resultante. Este tipo de filtros son de gran utilidad (figura 12, uso en detección de calcificaciones) para uniformizar el fondo de las imágenes y resaltar objetos de interés (calcificaciones).

$$\text{Black Tophat} = (\text{Max} + \text{Min}) - \text{Original} \quad (5.12)$$

$$\text{White Tophat} = \text{Original} - (\text{Min} + \text{Max}) \quad (5.13)$$

El filtro Máximo consiste en asignar al píxel de referencia el valor más alto de los encontrados en la máscara, por lo que su efecto es ensanchar las zonas claras y adelgazar las oscuras. El Mínimo hace justo lo contrario, seleccionando el valor más bajo de los encontrados en la máscara y su efecto es reducir las zonas claras y ampliar las oscuras.

Los filtros máximo y mínimo; Son útiles para modificar máscaras. El filtro *Máximo* extiende las áreas en blanco y reduce las áreas en negro. El filtro *Mínimo* extiende las áreas en negro y reduce las áreas en blanco. Al igual que el filtro Mediana, los filtros Máximo y Mínimo se centran en píxeles individuales en la selección. Los filtros Máximo y Mínimo reemplazan, dentro del radio especificado, el valor de brillo del píxel actual por el valor de brillo mayor o menor de los píxeles circundantes.

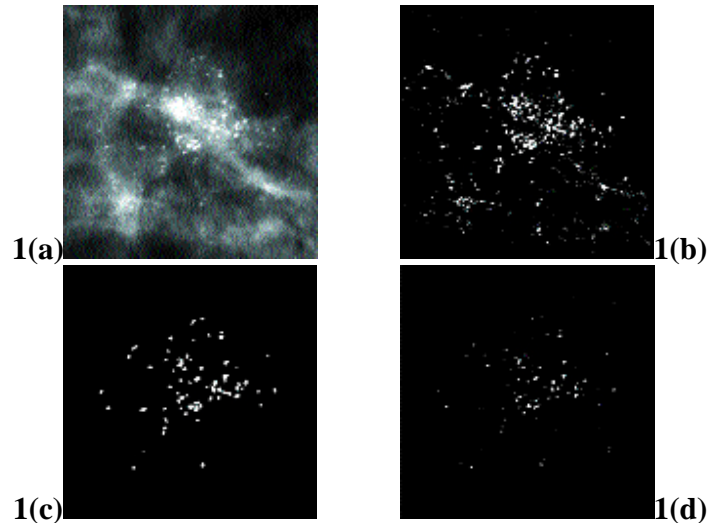
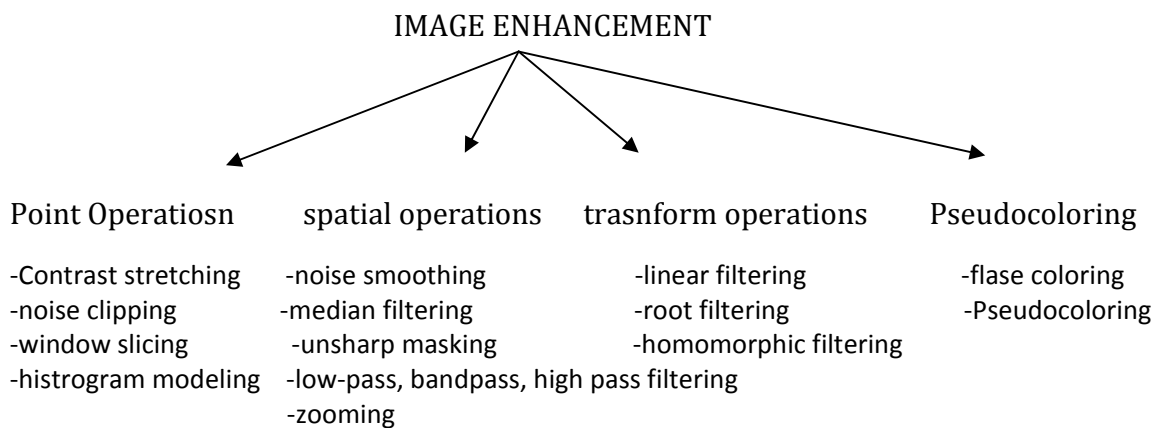


Figura. 12 [10]. Transformada top hat: (a) región de interés mostrando un agrupación de microcalcificaciones malignas; (b) resultado de la transformación top hat, (c) resultado de la diferencia de gaussianas, (d) resultado de la operación de engrosamiento condicional.

## 4.2 FILTRO DE AJUSTE (ENHANCEMENT)

Una vez la imagen ha sido procesada por el operador top hat se requiere hacer un post procesamiento (retoque) a la imagen para ser preparada en el proceso de binarización y así dejar solamente los blancos como el conjunto de elementos posibles de ser señales sospechosas a detectar y clasificar (data real de trabajo). En el siguiente diagrama se muestra las diferentes técnicas de enhancement:



Las técnicas de retoque de imágenes son técnicas matemáticas diseñadas para mejorar la calidad de la imagen, el resultado es una nueva imagen que muestra características de acuerdo a los criterios que se le definen.

En este trabajo se utilizo la técnica de operador espacial en donde:

$$g(x,y) = T [f(x,y)] \quad (5.14)$$

En donde,  $f(x,y)$  es la imagen de entrada y  $g(x,y)$  es la imagen de salida (figura 15),  $T$  es el operador sobre  $f$ . la principal acción para definir una vecindad sobre los puntos  $(x,y)$  de la imagen, el centro de la región es movido pixel a pixel.



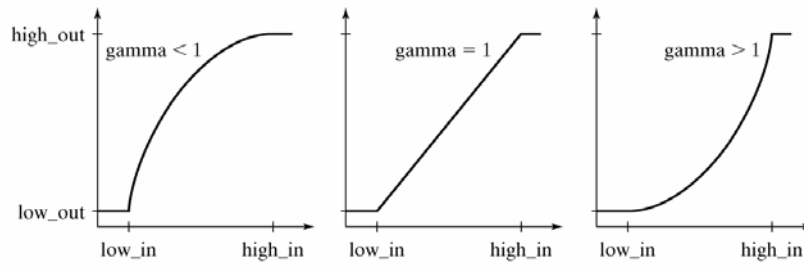


Figura 13 [3] operador ITF: diferentes valores de gama para la función de transformación de intensidad ITF.

El operador espacial que se estudia es el operador simple: transformada de función de intensidad, esta actúa sobre una función del tipo:

$$S = T(r) \quad (5.15)$$

En donde  $r$  es la intensidad de  $f$  y  $s$  la intensidad de  $g$ . en cualquier punto  $(x,y)$  de la imagen. La figura 13 ilustra la aplicación de la función *imadjust* en matlab.:

$$g = \text{imadjust}(f, [\text{low\_in } \text{high\_in}], [\text{low\_out } \text{high\_out}], \text{gamma});$$

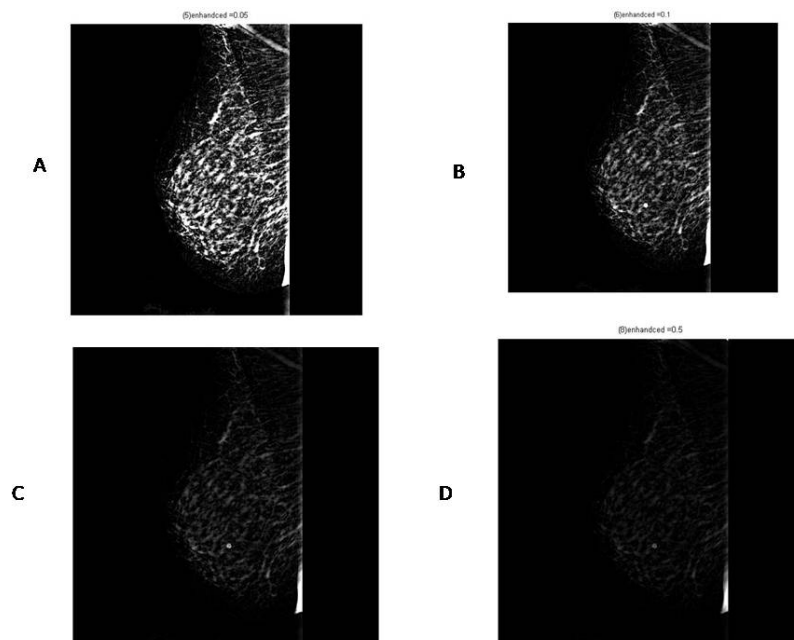


Figura 14. Ajustes de imagen con  $\text{low\_in}=0$ ;  $\text{low\_out}=0$ ;  $\text{high\_out}=1$  y variamos  $\text{high\_in}$  en: A) 0.05, B)0.1 , C)0.25 , D) 0.5

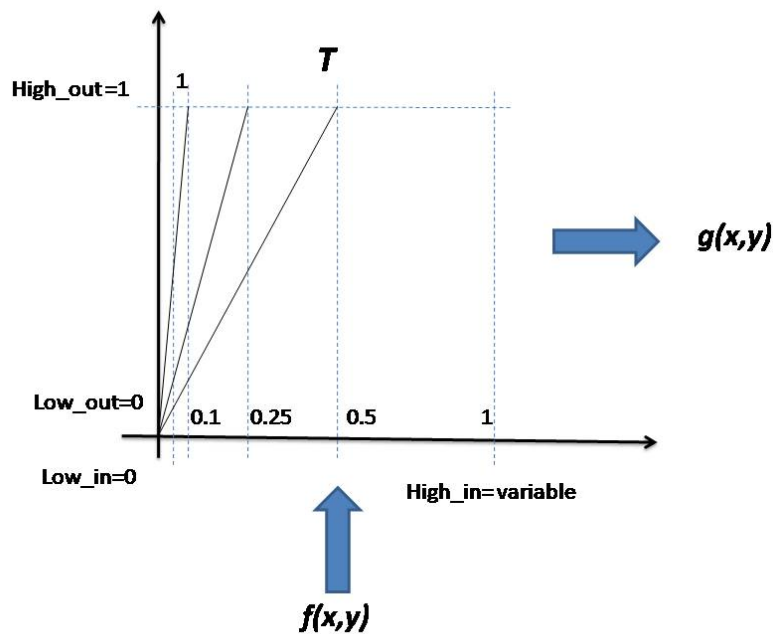


Figura 15. Funcion de trasferencia aplicada como filtro de enhancemet mediante la transformada de intensidad de grises.

En la figura 15 se muestra la función de trasferencia utilizada aplicando el concepto de transformada de intensidad y como la función de entrada es modificada por la rampa variando valores de **high\_in**, en el capítulo 9 se explicaran los valores que mediante pruebas de error y nitidez se definieron ser los de mejor comportamiento.

### 4.3 BINARIZACIÓN

La binarizacion es un procedimiento en la cual se convierte un pixel en solo dos valores 0 (negro) y 1 (Blanco) acorde con un umbral de decisión. Esto permite reducir en el conjunto de elementos seleccionados en el proceso de segmentación reducirlos a solo datos de interés, en otras palabras es un mecanismo de aproximación a la etapa de selección de características.

Hay muchos métodos de binarizar una imagen, unos dinámicos y otros estáticos, el de uso dinámico está basado en el histograma de la imagen, tomando el promedio del mismo, el otro es asumiendo un nivel de umbral el cual se calcula mediante prueba y error. En nuestro caso se utilizo el procedimiento de binarizacion bajo un umbral, en la figura 30 se aprecia ejemplos aplicados a las imágenes de la figura 16 con valor de umbral de 0.275.

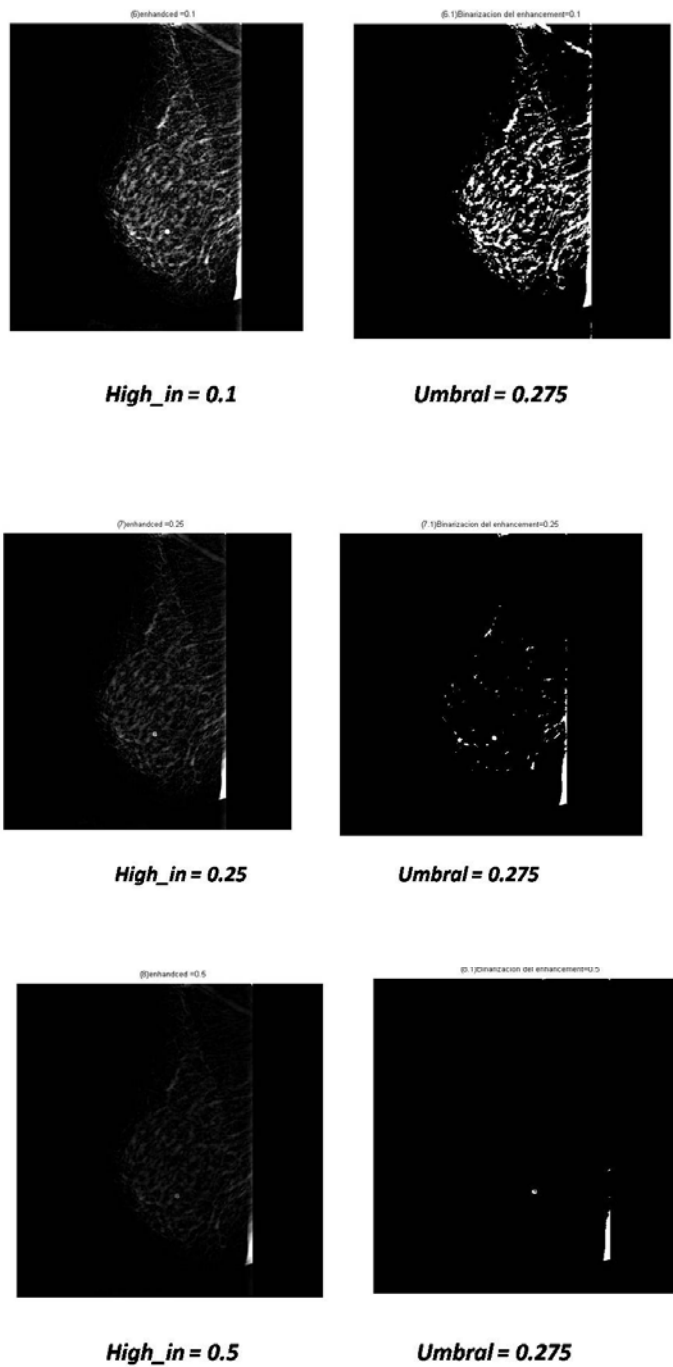


Figura 16. Ejemplos de aplicar el operador de binarización, usando los diagramas de la figura 48 y con un umbral de binarización de 0.275.

#### 4.4 RESUMEN CAPITULO

El proceso de extracción de características puede desarrollarse de multiplex formas hemos resumido varias de ellas, resaltando la importancia de la segmentación y refuerzo de la imagen resultante, y terminado con la binarización de la misma. En general el procedimiento de *segmentación* es una de las etapas más importantes en el procesamiento de imágenes, si la segmentación no es la acertada la selección y clasificación tendrán una alta probabilidad de incertidumbre. Por ello dentro de este proyecto se ha trabajado mediante el uso de los operadores morfológicos en especial el top hat que permite hacer un muy buena selección de los brillos altos en una imagen que es el caso de las calcificaciones en las imágenes de mamografía digital.

## 4.5 REFERENCIAS

- [1] Alasdair MacAndrew, Introduction to digital image processing, Thomson. 2004.
- [2] Anil K. Jain, Fundamentals of digital image processing, Prentice-Hall international, Inc. 1989.
- [3] Rafael C. Gonzales, Richard F. Woods, Steven L. Eddins, Digital Image processing using MatLab. Pearson, 2004.
- [4] José Augusto Moreno, clasificación de mamografías usando la función de base radial de rango tipo M, Instituto técnico nacional, Tesis para optar MSc. México, 2007.
- [5] Ernesto Coto, métodos de segmentación de imágenes médicas, Universidad central de Venezuela. [ecoto@strix.ciens.ucv.edu.ve](mailto:ecoto@strix.ciens.ucv.edu.ve)
- [6] Segmentación por umbralización – método OTSU, Universidad Nacional de Quilmes, Argentina. Vision artificial: <http://iaci.unq.edu.ar/materias/vision/archivos/apuntes/Segmentaci%C3%B3n%20por%20umbralizaci%C3%B3n%20-%20M%20-%20A9todo%20de%20Otsu.pdf>
- [7] Rubén Medina, Jesús Bellera, bases del procesamiento de imágenes médicas, Universidad de Los Andes, Facultad de Ingeniería, Grupo de Ingeniería Biomédica de la ULA (GIBULA, Venezuela).
- [8] [http://www.lluisvives.com/servlet/SirveObras/57915842105571617400080/008591\\_3.pdf](http://www.lluisvives.com/servlet/SirveObras/57915842105571617400080/008591_3.pdf)
- [9] [http://www.tsc.uc3m.es/imagineria/Curso\\_ProcesadoMorfologico/Contenido/Operaciones/OperacionesMorfologicas.html](http://www.tsc.uc3m.es/imagineria/Curso_ProcesadoMorfologico/Contenido/Operaciones/OperacionesMorfologicas.html)
- [10] [http://www.informedica.com.ar/antiguos/2.asp?id\\_nota=103](http://www.informedica.com.ar/antiguos/2.asp?id_nota=103)

## 5 SELECCIÓN DE CARACTERÍSTICAS

En los capítulos anteriores hemos venido procesando la imagen digital de mamografía, hasta tener un subconjunto de elementos posible de interés, sin embargo solamente en forma visual pudiéramos reconocer cuales serian los subconjuntos que se seleccionarían para ser calificados o clasificados como posibles éxitos dentro de esta población. En otras palabras en estos siguientes capítulos es en donde la parte de **computer vision** entra a ser importante, en donde se requiere una inteligencia artificial que permita seleccionar, clasificar y discriminar acorde con lo que estamos buscando.

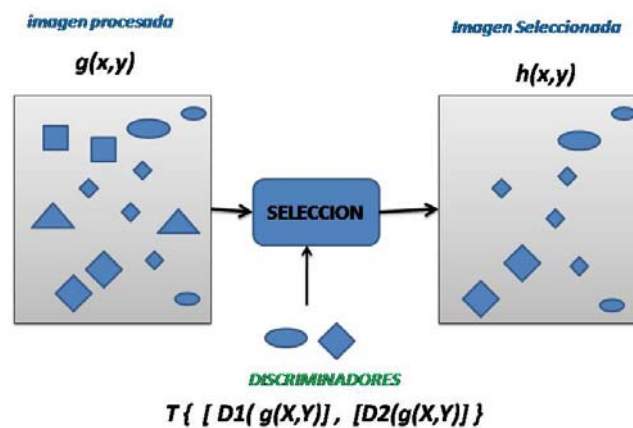


Figura 17. Ejemplo grafico de cómo actúa un sistema de selección de característica, acordes con un operador discriminador.

En la figura 17 la imagen procesada en etapas anteriores  $g(x,y)$  es precedida por un procesamiento de análisis de selección basada en discriminadores  $T$  como  $D_i$  ( $i=, 1, \dots, n$ ) discriminadores usualmente denominados vector de características. El vector de características es la base de datos se entrada para la etapa del clasificador en comparación a las diferentes técnicas de clasificadores supervisados, no supervisados y entrenados. En este trabajo solo se trabajaran dos discriminadores: tamaño y forma los cuales en la grafica aparecen en forma visual como un ovalo y un diamante para señalar  $D1$  y  $D2$  como discriminadores de formas y tamaño.

## 5.1 CALCIFICACIONES

En la sección 2.1 describimos las características de las calcificaciones, la figura 18 nos muestra en forma explicativa y visual como un radiólogo busca en una imagen los pequeños brillos los cuales tiene una probabilidad alta de ser calcificaciones.

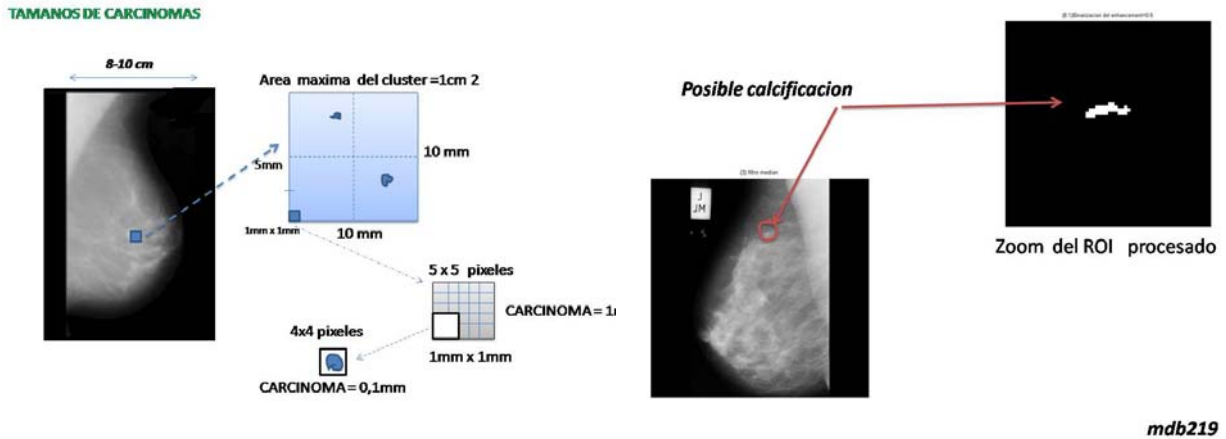


Figura 18. Ejemplos del procedimiento visual que hace el especialista, imagen mdb219, procesada y binarizada aumentada en el ROI de la posible calcificación.

En la siguiente sección veremos las diferentes técnicas descriptoras que permiten hacer una selección de características.

## 5.2 TECNICAS DE SELECCIÓN

Acorde con las siguientes características en una imagen es posible seleccionar objetos o elementos de interés (figura 19):

- Texturas [1]: son una serie de propiedades que tiene una imagen como granosidad, porosidad y dirección de las fibras. La textura es muy utilizada en imágenes satelitales para el reconocimiento de mapas, carretes, ciudades. La textura posee cualidades ópticas y táctiles,
- Formas: por los diferentes aspectos morfológicos que tiene un objeto dentro de una imagen, en ellas podemos resaltar: círculos, óvalos, estrellados, cuadrados, sin forma específica.
- Tamaños: por las diferentes en aspectos como el área y perímetro que tiene iguales figuras o similares, aspectos de magnitud geométrica se explican a continuación en la siguiente sección.

- **Color:** acorde con las diferentes tonalidades de color es posible utilizar también estas como una variable para selección de elementos u objetos.



Figura 19. Diferentes variables que son útiles en técnicas de selección de elementos en imágenes digitales.

De igual forma técnicas más complejas puede ser la mezcla de varias de las anteriores. Pero la que más se utiliza en selección y clasificación de imágenes medicas sobre todo imágenes de grises son las que se basan en aspectos geométricos y que vamos a explicar a continuación.

### ***Magnitudes geométricas***

[2] En procesamiento de imágenes el cálculo de magnitudes geométricas se refiere a la evaluación de características asociadas con objetos extraídos ya sea con propósito de reconocimiento o para elaborar una base de datos. Estos algoritmos son aplicados en imágenes médicas para varias aplicaciones: evaluación de tumores, calcificaciones, conteo de plaquetas, análisis en patología, análisis de exámenes en laboratorio clínico, análisis de tejidos. [3] La medición de objetos está relacionada con el procesamiento gráfico en donde los límites y los objetos son delimitadas por los usuarios en forma manual (enseñanza y sintonización) aunque también se usa métodos de segmentación previos a este proceso.

- **Área:** Por lo general para el cálculo de este parámetro se computa la suma de los pixel que integran el objeto localizado( figura 20), hay que definir el límite del objeto o la región en donde se encuentra el objeto así como el rango de valores de pixel que se deberán contar:

$$\text{Área} = \sum_{i=1}^N \sum_{j=1}^M \text{seg}(i,j) \quad (5.1)$$



En donde  $seg(i,j)$  es un pixel segmentado (procesado) y  $Ni$  y  $Mj$  corresponden al lado de la ventana que contiene la señal u objeto a calcular el área.

- **Perímetro:** es el número de pixeles que están en la frontera del objeto segmentado, o el número de lados adyacentes a la frontera. El cálculo se basa en el diámetro vertical, horizontal y diagonal del objeto a calcular:

$$P = N_h + N_v + N_d (\sqrt{2}) \quad (5.2)$$

$N_h$ : número de pixeles horizontales

$N_v$ : numero de pixeles verticales

$N_d$ : número de pixeles diagonales

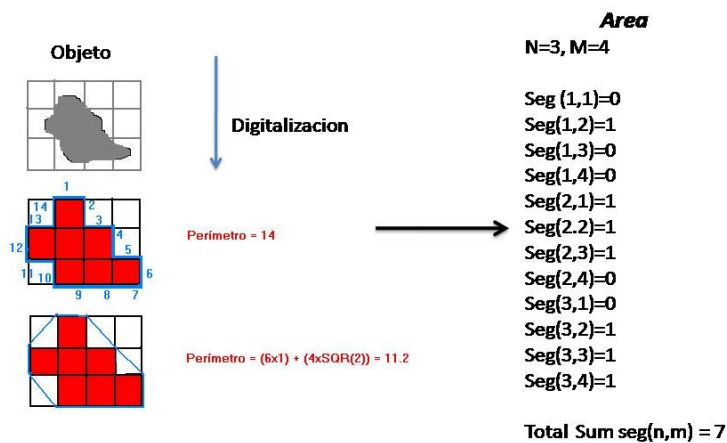


Figura 20. Calculo de perímetro y área en un objeto pixelado.

- **Centroide:** se refiere a la ubicación del centro geográfico de un objeto (centro de masa). El centroide se expresa en coordenadas  $x,y$  y se calculo mediante la expresión:

$$X_c = \frac{1}{\text{area}} \sum_{i=1}^N X_i \quad (5.3)$$

$$Y_c = \frac{1}{\text{area}} \sum_{i=1}^N Y_i \quad (5.4)$$

- **Diámetro máximo:** [4] calcula el valor del diámetro máximo de una señal (objeto), es la distancia entre los extremos más lejanos del borde del objeto. Para un círculo el diámetro es dos veces el radio. Este valor es encontrado calculando la distancia entre todos los puntos del perímetro y buscando un par de puntos  $(x_1, y_1)$  y  $(x_2, y_2)$  con la mayor distancia, está dada por la fórmula:

$$\text{Diámetro\_maximo} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (5.5)$$

- **Diámetro mínimo:** calcula el valor del diámetro mínimo de una señal (objeto), este valor se calcula buscando la mayor distancia que existe entre los pares de puntos del perímetro que formen una línea perpendicular con el diámetro máximo (figura 21).

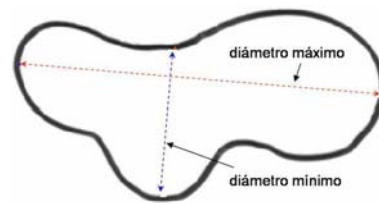


Figura 21 [4]. Ejemplo del diámetro máximo y mínimo en un objeto no circular.

- **Circularidad:** es un indicador en cuanto me informa la forma de la señal, este indicador para un círculo es 1, y para un cuadrado es 0.78 y para una fibra es 0.0. la fórmula que lo define está dada por:

$$\text{Factor de forma} = 4 \pi \frac{\text{area}}{\text{perimetro}^2} \quad (5.5)$$

- **Elongación:** es el cociente entre el diámetro mínimo y el diámetro máximo, este valor puede estar entre [ 0 y 1] . es 1 si el objeto se convierte en un cuadrado, y 0 si se vuelve una línea.

$$\text{Elongación} = \frac{\text{diámetro mínimo}}{\text{diámetro máximo}} \quad (5.6)$$

- **Relación de aspecto:** es el cociente entre el diámetro máximo y el diámetro mínimo:

$$\text{Relación de aspecto} = \frac{\text{diámetro máximo}}{\text{diámetro mínimo}} \quad (5.7)$$

- **Compacidad:** es la medida que nos indica que tan aglutinada o compacta es una señal u objeto, se calcula de estas formas:

$$\text{Compacida1} = \frac{\text{perímetro}^2}{\text{area}} \quad (5.8)$$

$$\text{Compacida2} = \frac{\text{perímetro}^2}{4 \pi \text{ area}} \quad (5.9)$$

- **Área convexa:** es el área que ocupa un polígono u objeto no uniforme que contiene toda la señal (figura 22).

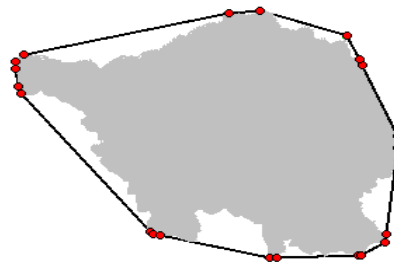


Figura 22. [4] área convexa de una figura irregular.

- **Solidez:** es una medida de densidad de una señal u objeto, se obtiene como la relación entre el área del objeto y el área convexa que la contiene. Valores cercanos a 1 significan que el objeto es bien sólido (calcificación benigna) valores menores a 1 significan que el elemento tiene bordes irregulares (calcificación maligna):

$$\text{Solidez} = \frac{\text{area}}{\text{area convexa}} \quad (5.10)$$

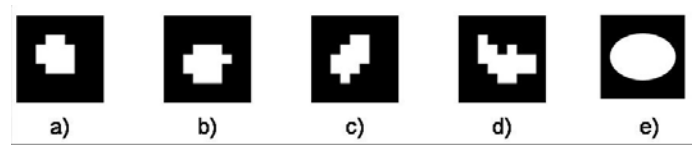


Figura 23. [4] Ejemplos de solidez en algunos objetos: A)0.9 B)0.85 C) 0.84 D)0.75 E)1.00

- **Agrupamiento:** es el proceso de contra y etiquetara objetos dentro de una imagen. Los grupos o ***clúster*** que se obtiene mediante el agrupamiento están constituidos por pixeles y están relacionados entre sí por una característica. Las técnicas que se utilizan para este propósito son la K medias que se enunciará en el siguiente capítulo.
- **Excentricidad:** mide la forma del objeto, rango entre [ 0 1], siendo cero objetos totalmente circulares y uno objetos lineales.

### 5.3 DISCRIMINADORES PROPUESTOS

Acorde con los discriminadores que se han escogido para este trabajo: forma y tamaño los discriminadores que mejor se comportarían serian:

#### ***Formas:***

- Solidez
- Circularidad
- Elongación
- compacidad

#### ***Tamaño:***

- Área.
- perímetro

## 5.4 RESUMEN

En este capítulo vimos diferentes técnicas útiles para el proceso de selección de objetos o señales en una imagen acorde con los descriptores. Para este trabajo las magnitudes geométricas son las más útiles en identificar índices que nos permitan en el proceso de selección localizar formas y tamaños, acordes con el siguiente paso. En el cual se clasifican las señales procesadas para dictar un resultado.

Hasta esta parte del documento hemos avanzado en explicar como la imagen se va procesando y seleccionando áreas de interés, ahora teniendo la técnica para evaluar las posibles señales u objetos encontrados seleccionados usando las técnicas de forma, como son: solidez, circularidad entre otras, las cuales nos identificarán que tan cercana es una calcificación a ser benigna y con las técnicas de tamaño: área y perímetro nos identificarán que tan cercana es una calcificación a ser maligna.

## 5.6 REFERENCIAS

[1] <http://www.analisisfotografia.uji.es/root/analisis/metod/6c.htm>

[2] harley R. Myler., Arthur weeks., "the pocket handbook of image processing algorithm in C. Prencite hall. 1993.

[3] José agosto Moreno, clasificación de mamografías usando la función de base radial de rango tipo M, Instituto técnico nacional, Tesis para optar MSc. México, 2007.

[4] Samuel Oporto Días, detección automática de agrupaciones de microcalcificaciones en mamografías digitalizadas, tesis, Instituto tecnológico de monterrey, Diciembre 2004. Paginas: 67-72

## 6 CLASIFICACIÓN Y DIAGNÓSTICO

La etapa de clasificación y diagnóstico es la última etapa en el ciclo de un sistema inteligente asistido por programa dentro del procesamiento y reconocimiento de patrones en las imágenes medicas. Se considera la etapa más compleja por su alta incidencia matemática y probabilística, los últimos quince años se han desarrollado nuevas técnicas de clasificación y se ha experimentado en la detección y reconocimiento de calcificaciones en mamografías. En el anexo 3 se profundizan las técnicas más utilizadas.

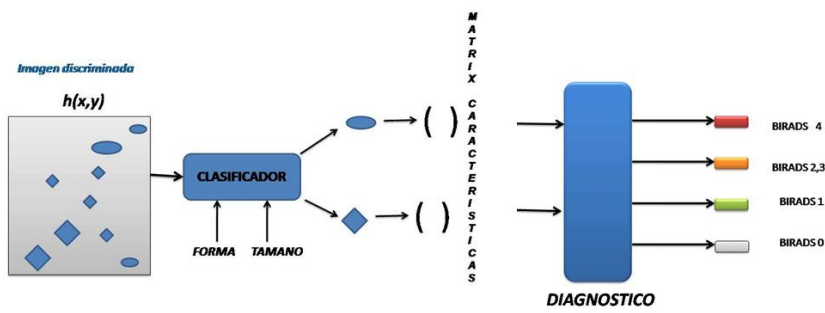


Figura 24. Flujo grama de la etapa de clasificación y diagnóstico.

En la figura 24 observamos en forma sencilla como es el proceso de clasificación y resultado del proceso de clasificación de una imagen entregada como salida de la etapa anterior de selección. La etapa de selección nos entrega una imagen con señales u objetos específicos acordes con los parámetros de forma y tamaño, en este trabajo se consideraran las formas circulares y ovaladas, así como tamaños pequeños de estas formas, esta imagen es la entrada al sistema de clasificación el cual procesa basado en la técnica que se escogerá y cuyo resultado será una matriz de valores como resultado del clasificador, estos valores son pasados por un proceso de selección acorde con el diagnóstico del protocolo BIRADS propuesto.

## EXPERIMENTACIÓN Y RESULTADOS



## 7.1 PLATAFORMA DE TRABAJO

El desarrollo del proyecto se usó diferentes herramientas de software sobre una plataforma computacional con las siguientes características:

- Estación de trabajo: configurada con procesador Intel Dual core i3-540, con 4Gb en memoria Ram y disco d 500Gb SATA, en la parte de aceleración y video se tenía una tarjeta Nvidia G-force 9400GT con 1GB en RAM con salida de doble monitor. Se trabajo con dos monitores de 22" con una resolución de 1900 x 1600 pixeles. El objetivo de esta plataforma era de ofrecer una máquina rápida y que permitiera el despliegue en dos monitores, así como poder terne en cada monitor una maquina virtual independiente de la otra para efectos de pruebas y trabajos en las imágenes.
- Sistema operativo: se trabajo bajo **Windows Xp**, service pack 3 como sistema operativo base, y bajo virtualBox se subieron versiones de Linux **Ubuntu** para efectos de pruebas de software de libre uso en imágenes, así como herramientas Dicom especiales en imágenes medicas.
- Plataforma de software de desarrollo: la simulación principal se hizo sobre **MatLab R2009a**, sin embargo se utilizaron herramientas de visualización y procesamiento básico como lo son: imageMagik, Gimp, iTk, ParaView, XnView para recortes, edición de imágenes de pruebas, filtros y tools de segmentación.

## 7.2 BASE DE DATOS

Como lo mencionamos en la sección 2.3 la base de datos que se escogió para este proyecto fue la MIAS (The Mammographic Image Analysis Society), a base de datos está compuesta por 322 (ver figura 23):

Total imágenes:	322
Imágenes Normales:	204
Imágenes con hallazgos:	118
Benignas:	66
Malignas:	52

Los hallazgos que se pueden encontrar para el caso de calcificaciones son: 25 casos discriminados así (ver tabla 1):

Tipo tejido	imagen referenciada
=====	
Tejido graso:	
Benignas: 2	mdb248, mdb252
Malignas: 4	mdb245, mdb256, mdb231, mdb238
Tejido Graso Glandular:	
Benignas: 5	mdb212, mdb214, mdb218, mdb219, mdb27
Malignas: 4	mdb209, mdb211, mdb213, mdb233
Tejido denso:	
Benignas: 5	mdb222, mdb223, mdb226, mdb236, mdb240
Malignas: 5	mdb216, mdb239, mdb241, mdb249, mdb253

### 7.2.1 MUESTRAS DE TRABAJO

De todos los 25 casos denominados la población u elementos de la población solo escogemos algunos para el proceso del proyecto.

- Muestras de entrenamiento: son las imágenes que se escogen para el entrenamiento del sistema de clasificación
- Muestras de pruebas: son las imágenes que se escogen para el proceso de verificación del sistema entrenado
- Muestras de simulación: son imágenes de simulación artificiales, creadas a maño y diferentes a las de la base de datos y que nos servirán para probar el funcionamiento del sistema.

Las imágenes de entrenamiento deben ser diferentes a las de pruebas. Acorde con el protocolo BIRADS debemos escoger imágenes que permitan los 5 estados: BIRADS 0, 1,2,3,4 ; las 25 imágenes están enmarcadas en la lista de malignas y benignas, pero las de incertidumbre (BIRADS 0) y normales BIRADS 1 las escogeremos de las otras imágenes de MIAS acorde con el anexo 1, así completaremos la población de imágenes para el desarrollo del experimento.

Se escogen las siguientes cantidades de imágenes (tabla 4) para el desarrollo del experimento (ver figura 25):

BIRADS 0	2 imágenes de entrenamiento, 2 imágenes de pruebas
BIRADS 1	2 imágenes de entrenamiento, 2 imágenes de pruebas
BIRADS 2,3	4 imágenes de entrenamiento, 4 imágenes de pruebas
BIRADS 4	4 imágenes de entrenamiento, 4 imágenes de pruebas

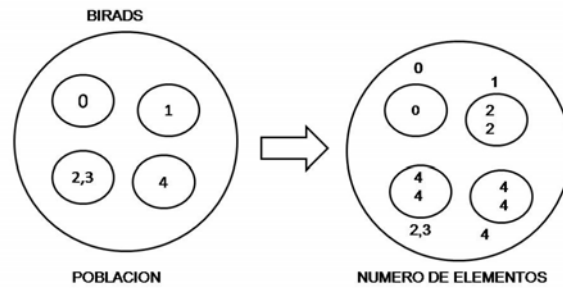


Figura 25. Población de imágenes del experimento

<b>CLASE</b>	<b>ENTRENAMIENTO</b>	<b>PRUEBAS</b>
BIRADS 0 (inciertas)	No hay imágenes	No hay imágenes
BIRADS 1 (normales)	mdb014, mdb018	mdb022, mdb024
BIRADS 2,3 Malignas	mdb209 , mdb211, mdb213 mdb233	Mdb238, mdb239, mdb241, mdb245
BIRADS 4 Benignas	mdb227, mdb222, mdb223, mdb226	mdb218, mdb236, mdb248, mdb252

Tabla 4 , imágenes escogidas para el experimento

La base de datos MIAS no tiene imágenes que cumplan con el valor de BIRAD 0, por ello no se incluirán muestras en este experimento.

La base de datos MIAS así como en los sistemas de vistas de imágenes medicas (viewers) las coordenadas se posiciona en el eje (x,y) teniendo como origen la esquina superior derecha (0,0).

La siguiente imagen mdb219 según la referencia es:

mdb219 G CALC B 546 756 29

Nos representa un centroide en las coordenadas (546,756) con 29 pixeles de radio. En la región se interés se encuentra un agrupamiento de calcificaciones (ver figuras 26, 27)



Figura 26. Exploración de la imagen mdb219 y su centroide de agrupaciones de calcificaciones.

imagen	X	Y
219	526	767
219	526	738
219	528	744
219	531	743
219	534	760
219	539	742
219	544	751
219	544	742
219	545	742
219	548	766
219	549	742



Figura 27. Ejemplo de calcificaciones localizadas en la imagen mdb219, total 11

El análisis de las 25 imágenes con marcación de calcificación de la base de datos MIAS nos arroja los siguientes resultados, debido a que la simple información entregada en las 7 columnas no es totalmente completa.

- Las imágenes mdb212 y mdb214 la región marcada por la base de datos cae fuera de la mama, por lo tanto estas imágenes no son consideradas como imágenes con calcificaciones.
- La imagen mdb240, tampoco tiene ningún hallazgo, por lo tanto no es considerada
- En las imágenes: mdb216, mdb233, mdb245 no se entrega información sobre el centroide ni del radio, estas aparecen como \*Note 3\*.
- En las imágenes: mdb223, mdb226, mdb239, mdb249 se aprecian más de una región de interés.

Quedando un total de 22 imágenes de trabajo. En la tabla 7 se han resumidos los hallazgos visuales sobre la base de datos de referencia, localizándose un total de 252 calcificaciones sobre 27 **clúster** de agrupamientos.

No	Imagen	Diagnóstico	Agrupamientos	Calcificaciones
1	mdb209	M	1	19
2	mdb211	M	1	2
3	mdb212	(B)	0	
4	mdb213	M	1	15
5	mdb214	(B)	0	
6	mdb216	M	1	21
7	mdb218	B	1	3
8	mdb219	B	1	13
9	mdb222	B	1	3
10	mdb223	B	2	6
11	mdb226	B	3	4
12	mdb227	B	1	2
13	mdb231	M	1	2
14	mdb233	M	1	65
15	mdb236	B	1	4
16	mdb238	M	1	5
17	mdb239	M	2	2
18	mdb240	(B)	0	
19	mdb241	M	1	12
20	mdb245	M	1	43
21	mdb248	B	1	4
22	mdb249	M	2	11
23	mdb252	B	1	7
24	mdb253	M	1	5
25	mdb256	M	1	4
total			27	252
promedio				9.3

Tabla 5 [anexo 2]. Tabla resumen de calcificaciones por imágenes

### 7.3 EXPERIMENTACIÓN EN EL PREPROCESAMIENTO

La figura 28. Nos muestra las etapas implementadas en el pre procesamiento acorde con el modelo propuesto:

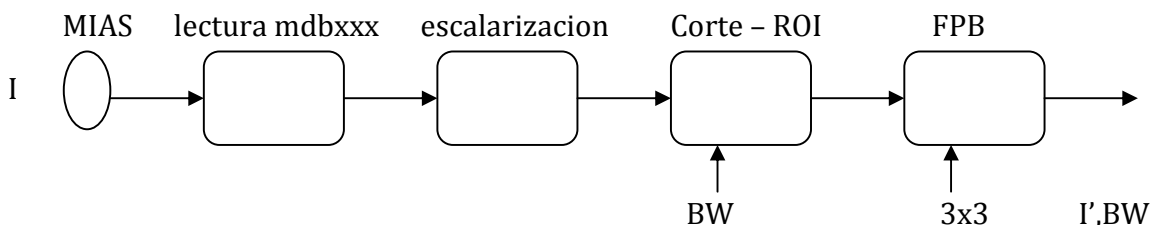


Figura 28. Bloques del pre-procesamiento

En el figura 28, se muestran las etapas que se utilizaron en el diseño del pre-procesamiento del experimento. Describiremos a continuación los filtros probados y sus resultados.

#### 7.3.1 ESCALARIZACIÓN.

La escalarizacion es la conversión de un pixel a formato de 8 Bits permitiendo una resolución de 256 niveles de grises, a pesar que las imágenes escogidas en la base de datos MIAS viene en formato de 1024 x 1024 a 8 bits, para efectos de implementar un programa abierto al uso de otras bases de datos e incluso la entrada de una imagen de un tipo diferente a .pgm se decidió dejar la escalarizacion a 8 bits.

Implementación Matlab:

```
im01 = uint8 (im00);
```

#### 7.3.2 REDUCCIÓN DEL ROI

En el planteamiento de la metodología se había planteado efectuar esta etapa después del filtro mediaño, sin embargo para efectos de reducir procesamiento en el proceso se relocaliza hacer el corte de la mama antes del filtro

##### 7.3.2.1 CORTE MANUAL

Mediante la herramienta de abobe photoshop CS5 se hace el corte en forma manual seleccionando la ventana del ROI (región de interés) de tal forma que se escoge solo el área de la mama en la nueva ventana de trabajo, quedando ventanas dentro del rango de: [ 450-550 , 1024] (ver figura 29).

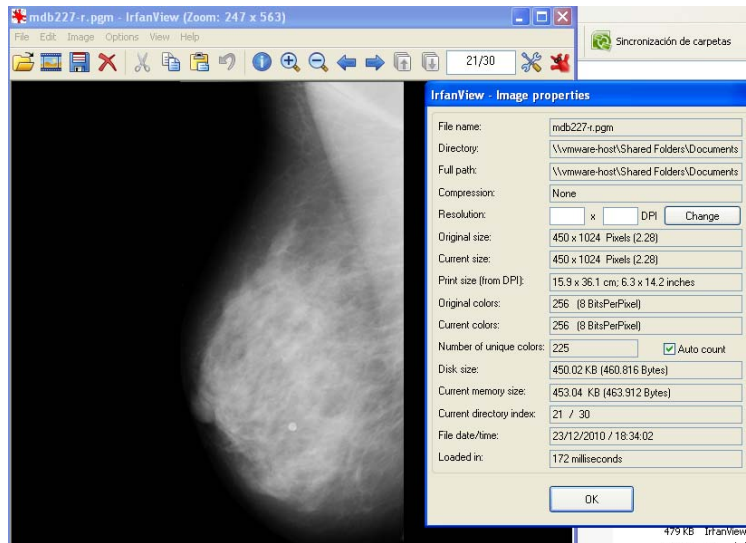


Figura 29. Vista del tamaño del recorte manual 450x1024

### 7.3.2.2 CORTE AUTOMÁTICO

Se desarrollo el algoritmo propuesto en la sección 2.3

### 7.3.3 FILTRO PASA BAJOS

El filtro escogido y probado el cual mostro mejor comportamiento fue el filtro mediana con una ventana de 3x3, se probaron ventanas de 5x5 y 7x7 sin embargo se perdía resolución y nitidez de la imagen. La siguiente línea toma como entrada la imagen im01 y efectúa el filtro 3x3 medianas dejando como una nueva imagen im\_mediana.

```
im_mediana = medfilt2 (im01 , [3 3] , 'symmetric');
```



## RESUMEN ETAPA DE PREPROCESAMIENTO

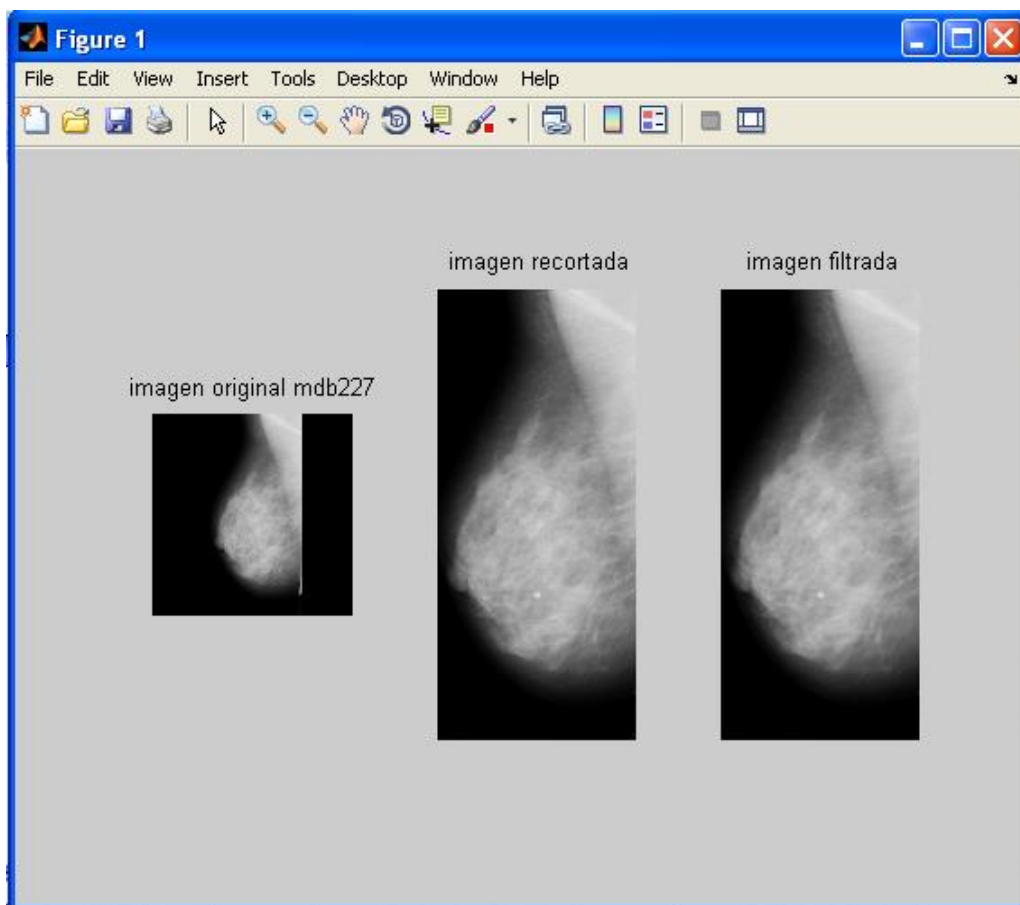


Figura 30. Resumen visual de la etapa de pre procesamiento sobre la imagen mdb227

En la figura 30 se muestra el resultado de la ejecución de la primera etapa en la metodología propuesta: pre procesamiento corrida sobre el programa de Matlab.

## 7.4 EXPERIMENTACIÓN EN LA EXTRACCIÓN DE CALCIFICACIONES

La figura 31. Nos muestra las etapas implementadas en la extracción de características acorde con el modelo propuesto:

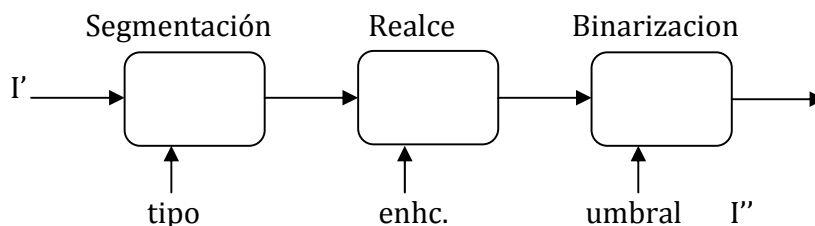


Figura 31. Flujo grama de la etapa de extracción de características

### 7.4.1 TECNICAS PROBADAS

Para la detección de borde y segmentación se utilizaron técnicas según la derivada, en especial la técnica de *Sobel* y la de *Top Hat* vistas en la sección 4.

### 7.4.2 REALCE Y BINARIZACIÓN

En la tabla 6. Se muestran los diferentes valores trabajados para lograr la mejor percepción visual y de comportamiento en la segmentación de la imagen.

<b>TIPO</b>	<b>ENHC.</b>	<b>UMBRAL</b>	<b>RESULTADO</b>	<b>Percepción visual</b>
<b>Sobel</b>	50	0.2	Enhc = 300 Umbral=0.35	Mejor localización de acumulaciones
	100	0.275		
	250	0.35		
	300			
<b>Top hat</b>	.3	0.025	Enhc = 0.425 Umbral= 0.25	Mejor localización de acumulaciones
	0.35	0.275		
	0.4	0.3		
	0.5			

Tabla 6. Parametrización en la etapa de segmentación

En las siguientes figuras observamos los resultados en las imágenes de pruebas: mdb227, mdb222, mdb223, mdb226 (figura 32)

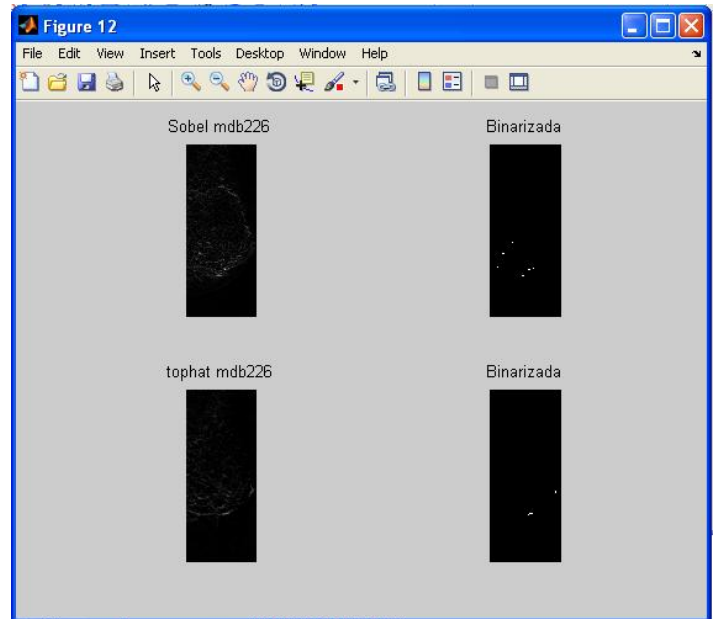
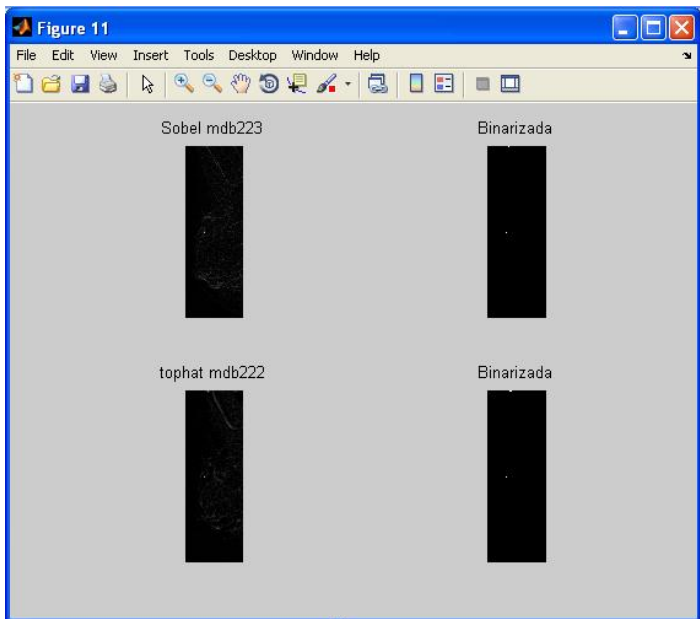
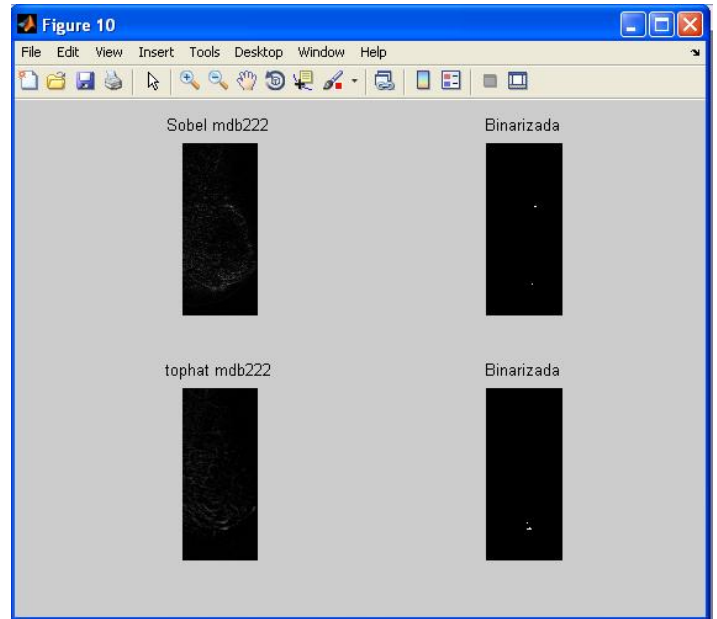
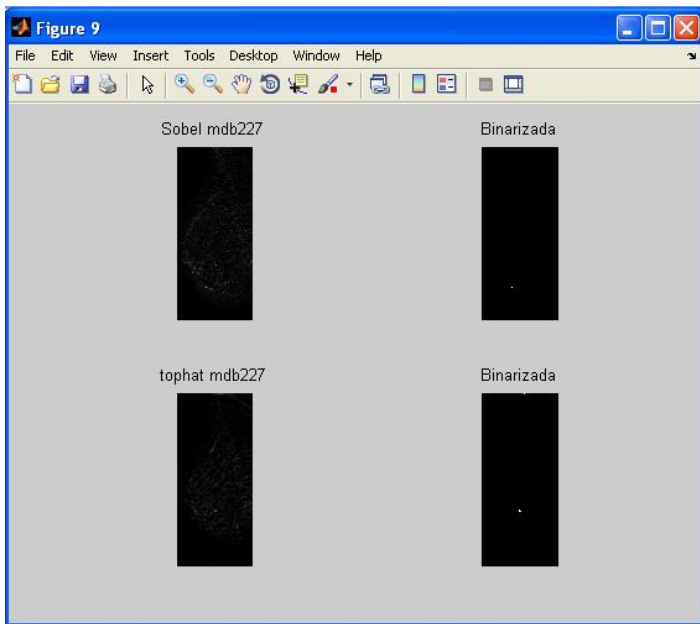


Figura 32. Resultado de aplicar extracción mediante Sobel y Tophat en las imágenes de pruebas.

En este trabajo se decidió utilizar la técnica de **Top Hat** con los siguientes parámetros:

```

realce = 0.5 ;           %nivel de enhacement o realce.
umbralbin_tophat = 0.25 ; %nivel del umbral de binarizacion
r_tophat = 14 ;        %radio de aplicacion de Top hat r=14

```

### 7.4.3 RESUMEN

- Pruebas de las imágenes benignas.

Según la tabla 7 observamos que las calcificaciones por cada imagen son:

Imagen	agrupamiento	calcificaciones
mdb227	1	2
mdb222	1	3
mdb223	2	6
mdb226	3	4

el resultado de la extracción aparece en la figura 33.

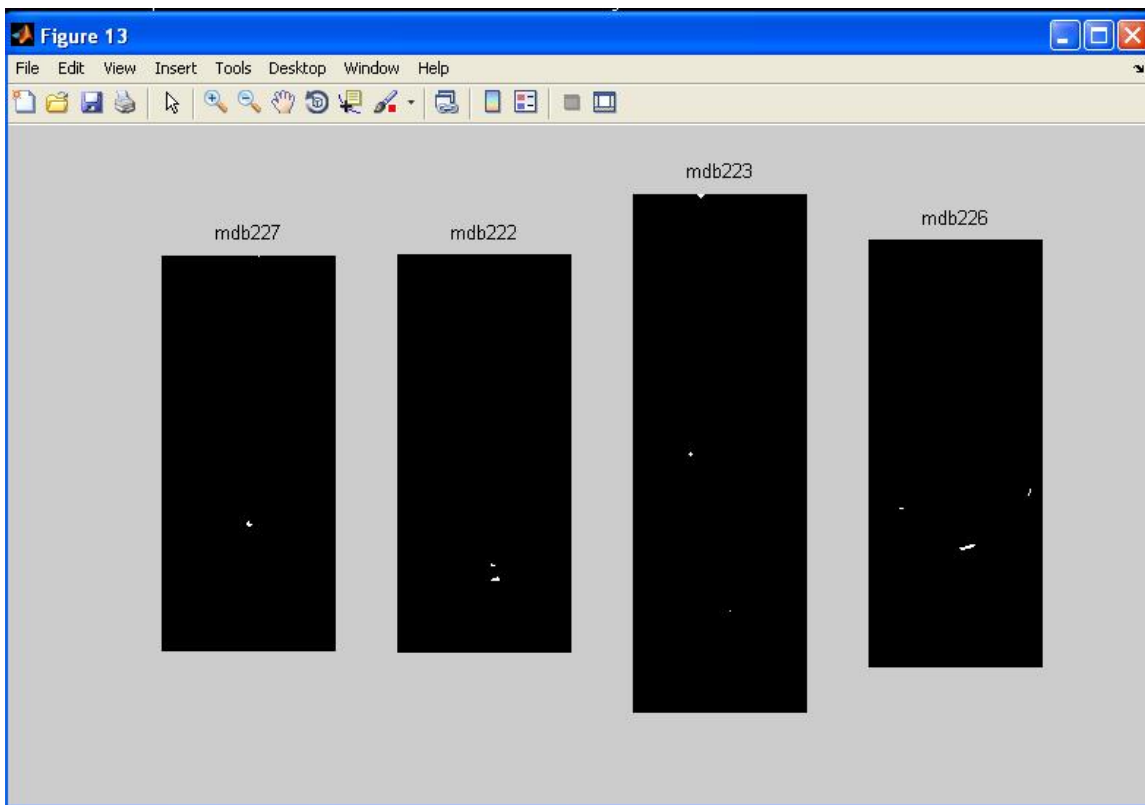


Figura 33. Resultado de extracción sobre imágenes de prueba Benignas

- Pruebas de las imágenes malignas.

Según la tabla 7 observamos que las calcificaciones por cada imagen son:

Imagen	agrupamiento	calcificaciones
Mdb209	1	19
Mdb211	1	2
Mdb213	2	15
mdb233	3	65

el resultado de la extracción aparece en la figura 34.

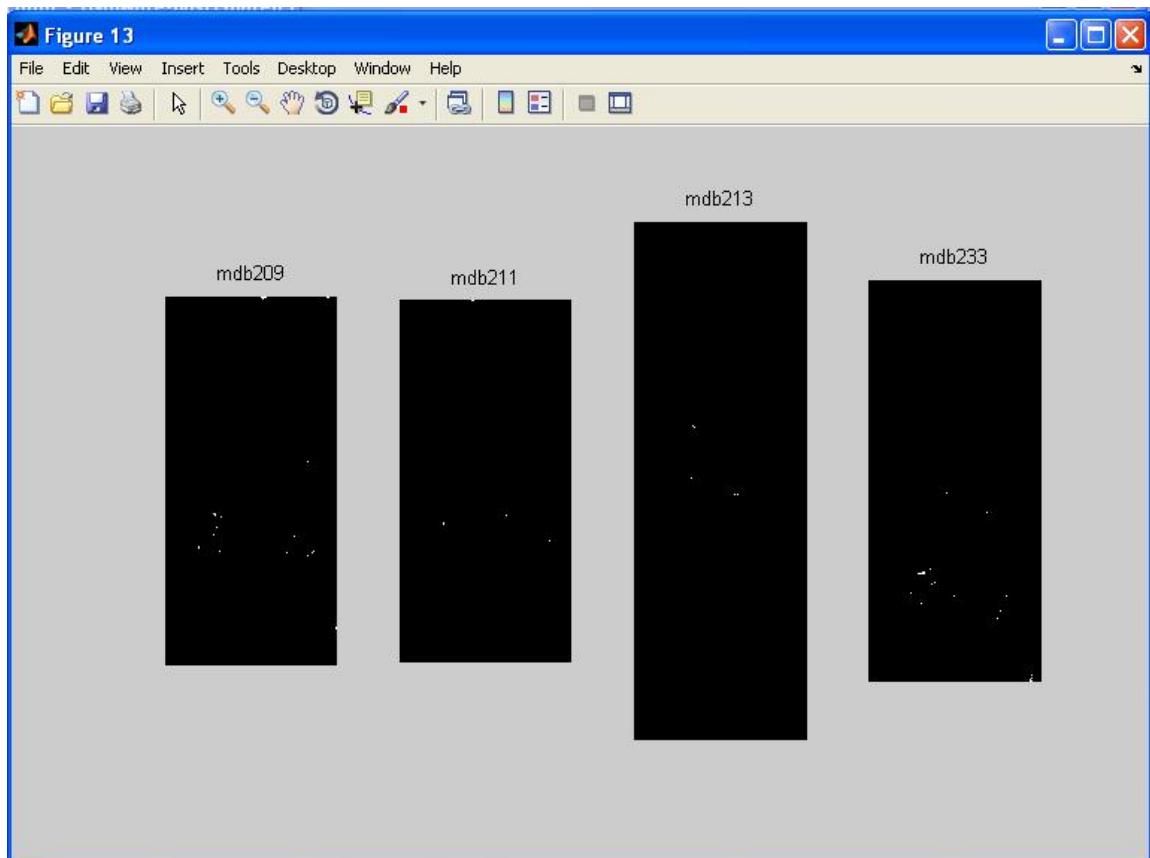


Figura 34. Resultado de extracción sobre imágenes de prueba Malignas

## 7.5 EXPERIMENTACIÓN EN LA SELECCION DE CALCIFICACIONES

Acorde con el tamaño de las imágenes con formato 1024 x 1024 se tiene un total de 1'048.576 puntos por analizar, y si tenemos un total de 22 imágenes, el tiempo de procesamiento sería muy largo y costoso. En la etapa de segmentación se ha reducido a un porcentaje muy alto las regiones de interés con posibles agrupamientos a identificar como calcificaciones.

### 7.5.1 DESCRITORES SEGÚN BIRADS

BIRADS especifica dos descriptores importantes en el diagnóstico de las calcificaciones: tamaños y forma.

- Se define una calcificación benigna aquellas que tiene un tamaño de 1-4 mm con formas circulares, ovaladas y uniformes.
- Las clasificaciones malignas por lo regular son pequeñas en rangos con un diámetro de 0.2 a 1 mm con formas finas, no uniformes con una distribución de grupos, por lo regular se evalúan áreas de 1 cm<sup>2</sup> en el cual se localizan más de 5 señales.

Repetimos las figuras 11 y 21 respectivamente para ilustrar los descriptores según tamaño y formas:

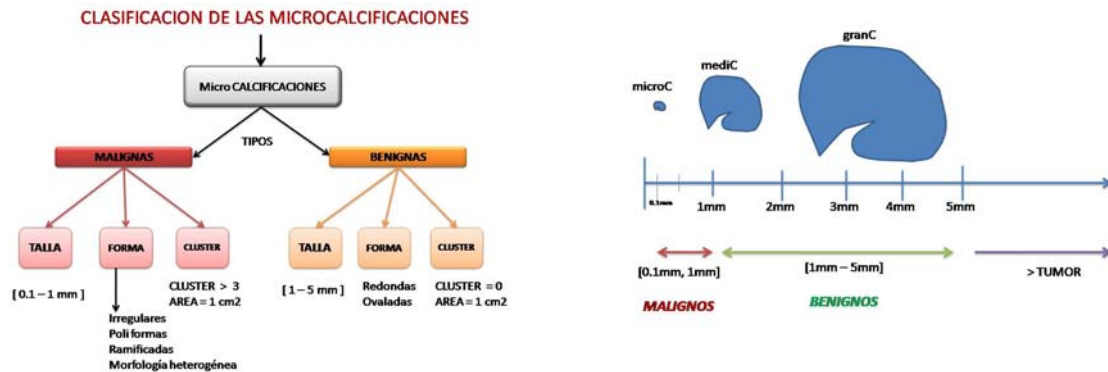


Figura 35. Mediante el análisis del tamaño y forma se etiquetan las calcificaciones.

### 7.5.2 TECNICAS PROPUESTAS

La señal de la imagen en la anterior etapa – segmentación - se tiene depurada en data pura (ver figura 33, 34), debemos determinar el tamaño de las ventanas acorde con la resolución de las imágenes importadas de la base de datos MIAS. El nivel de resolución de las imágenes es de 200u/Pixel, es decir que el lado de cada pixel es de 0.2mm lo cual quiere decir que el objeto más pequeño localizable es de 0.04 mm<sup>2</sup>

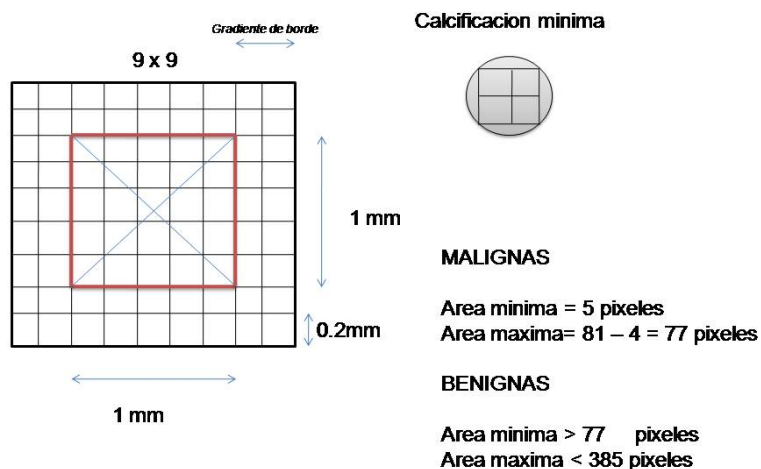


Figura 36. Áreas y rangos de caracterización

En la sección 5.3 vimos algunos de los posibles discriminadores que se podían utilizar para poder etiquetar las señales acorde con los parámetros que nos permitan especificar y clasificar si la señal u objeto es Benigno, maligno o no presenta anormalidad.

Para ello se trabajó con el cálculo del área, perímetro, excentricidad, área convexa y solidez. El procedimiento que se realizó está basado en el diagrama 35 y 36 respectivamente; el cual se basa en dos secciones:

- Separación de objetos Grandes: las calcificaciones grandes están localizadas según el protocolo BIRADS entre los rangos de 1mm a 5 mm lo cual es equivalente a 77 a 385 píxeles acorde con la resolución de las imágenes MIAS. El algoritmo se basa en la búsqueda de objetos comprendidos en este rango  $> 77$  y menores a  $< 385$  respectivamente. Los que no se encuentran en este rango los elimina de la imagen binaria y como resultado la salida es una imagen con ventana de señales u objetos probablemente Benignos.
- Separación de objetos pequeños: las calcificaciones pequeñas están localizadas según el protocolo BIRADS en los rangos de 0.2mm a 1mm lo cual es equivalente a 5 a 77 píxeles respectivamente. El algoritmo implementado se basa en la búsqueda de objetos menores a  $< 77$  píxeles y descartando los mayores a este tamaño (posiblemente Benignos). El resultado es una ventana de imagen binaria que contiene objetos menores a  $< 77$  píxeles, posteriormente se buscan los objetos con un área mayor  $> 5$  píxeles para descartar los más pequeños que ya no se consideran datos de interés.

El algoritmo lo podemos visualizar en el figura 37, en donde observamos las dos rutas de selección. Una vez se han etiquetado en cada ventana de imagen las posibles señales denominadas '*objetos posibles Benignos*' y '*objetos posibles malignos*', se encuentra el vector de características.

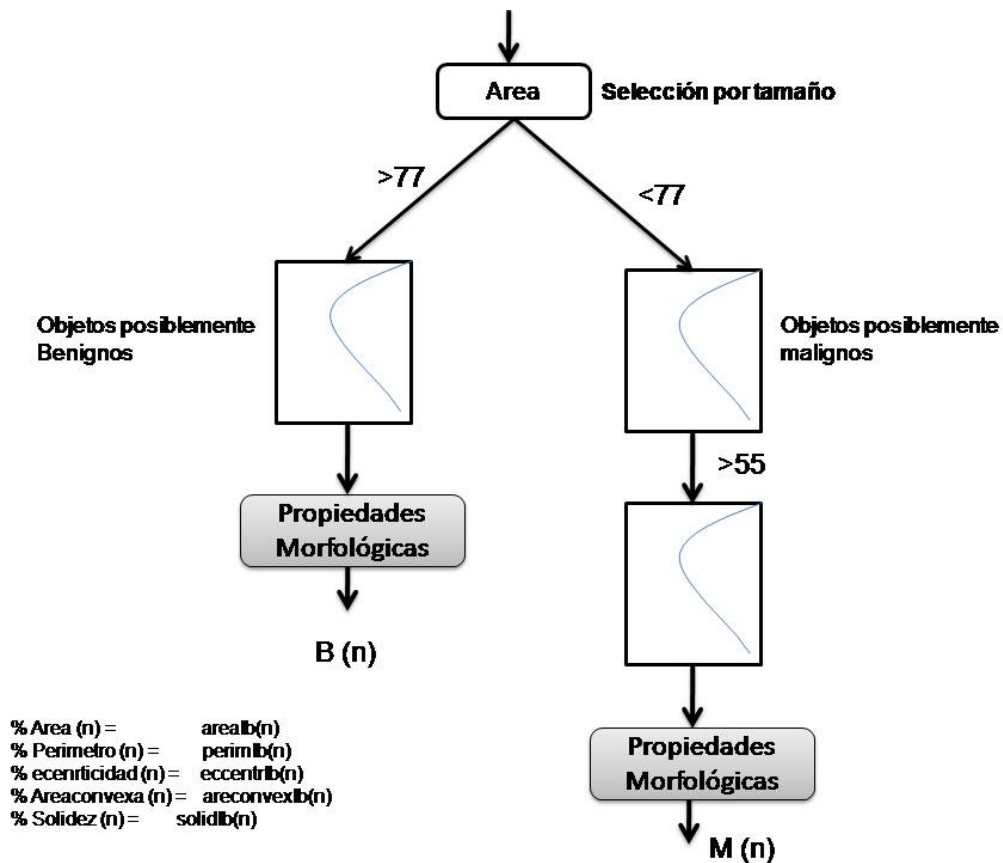


Figura 37. Algoritmo de separación de calcificaciones pequeñas y grandes, para su estudio posterior acorde con el discriminador de formas.

### 7.5.3 VECTOR DE CARACTERÍSTICAS

El vector de características debe permitir identificar los dos discriminadores escogidos para este proyecto: forma y tamaño. Por ello en la sección 5.3 se plantearon algunos de ellos, en nuestra experimentación se trabajó con los siguientes señales:

- Área
- Perímetro
- Excentricidad
- Área convexa
- Solidez.

Cada objeto localizado tiene este vector de características



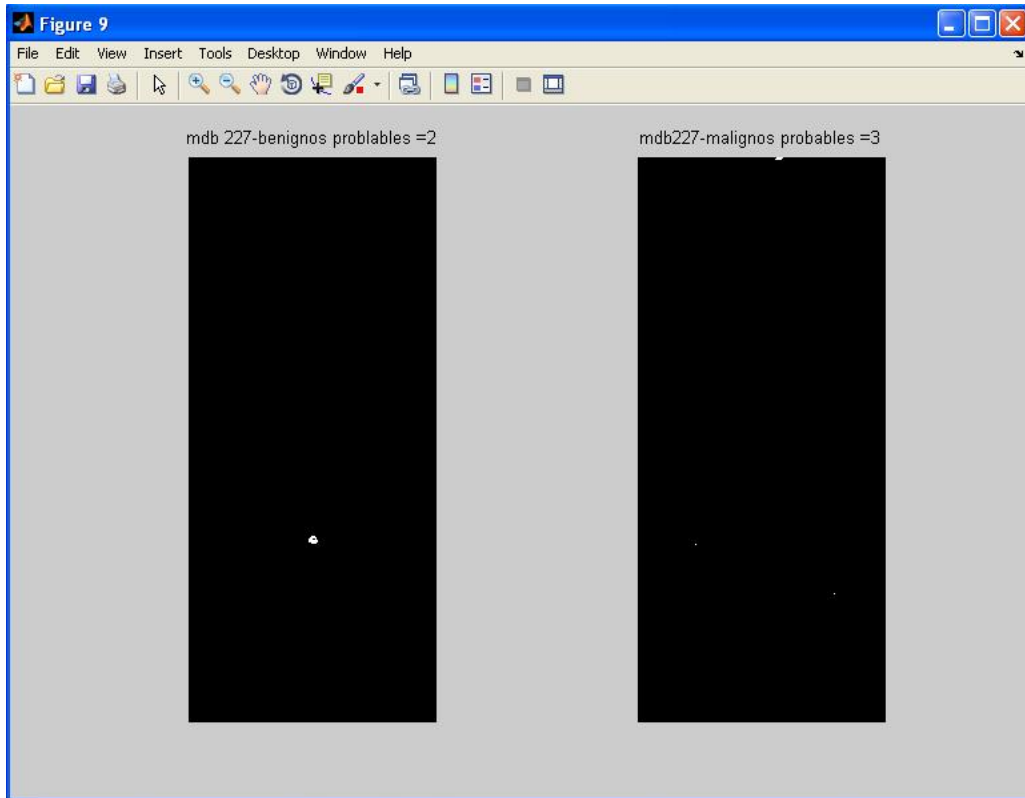


Figura 38. Imagen de salida de la etapa de selección de características, imagen mdb227 se localizan 2 hallazgos posibles benignos y tres hallazgos posibles malignos.

***Resultado (vector características) imagen mdb227***

Total Objetos hallados en Imagen:	7
Objetos etiquetados:	5
Objetos probables a ser benignos:	2
Objetos probables a ser malignos:	3
	Hallazgos benignos:
Vector área:	[192,170]
Vector perímetro:	[382,47.213203435596420]
Vector excentricidad:	[0.999986436539961,0.440598448306397]
Vector área convexa:	[192,178]
Vector solidez:	[1,0.955056179775281]
	Hallazgos malignos:
Vector área:	[5,50,7]
Vector perímetro:	[382,47.213203435596420]
Vector excentricidad:	[382,47.213203435596420]
Vector área convexa:	[5,60,7]
Vector solidez:	[1,0.8333333333333333,1]

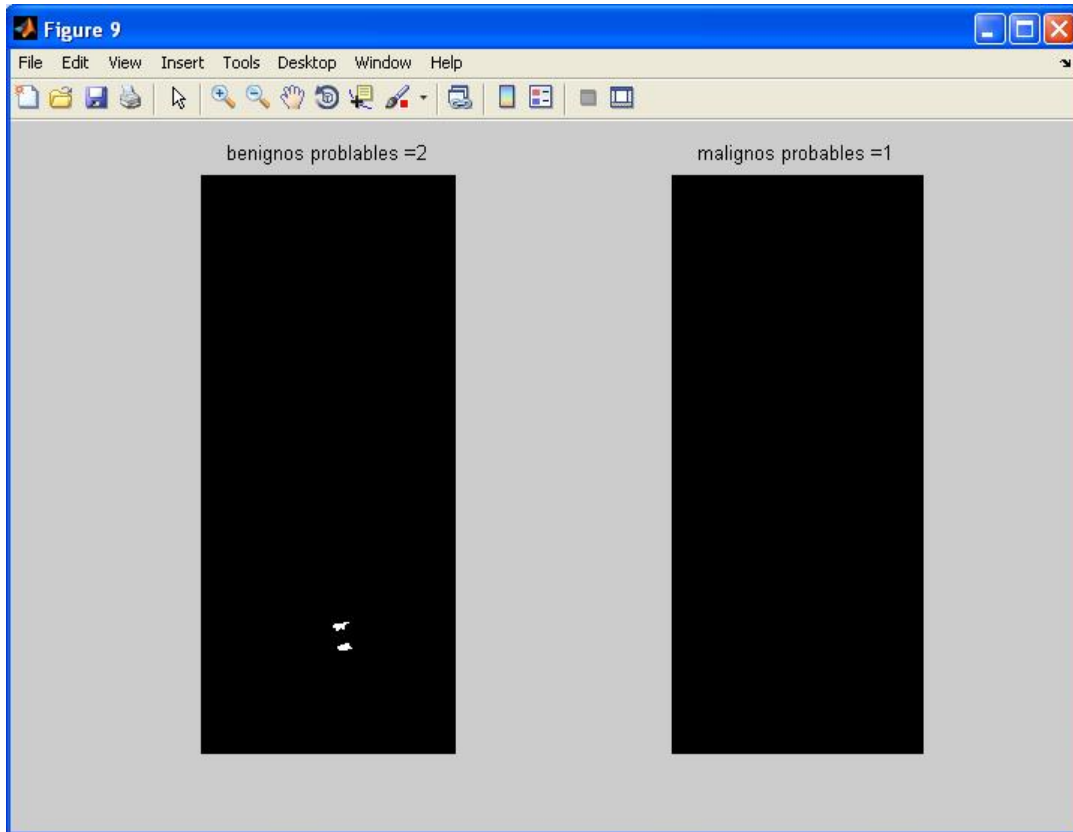


Figura 39. Imagen de salida de la etapa de selección de características, imagen mdb222 se localizan 2 hallazgos posibles benignos y 1 hallazgos posibles malignos.

***Resultado (vector características) imagen mdb222***

Total Objetos hallados en Imagen:	3
Objetos etiquetados:	3
Objetos probables a ser benignos:	2
Objetos probables a ser malignos:	1
	Hallazgos benignos:
Vector área:	[242,235]
Vector perímetro:	[86.911688245431380,66.041630560342600]
Vector excentricidad:	[0.902851068964672,0.872267704356917]
Vector área convexa:	[310,259]
Vector solidez:	[0.780645161290323,0.907335907335907]
	Hallazgos malignos:
Vector área:	[6]
Vector perímetro:	6.828427124746190
Vector excentricidad:	0.8333333333333333
Vector área convexa:	6
Vector solidez:	1

#### 7.5.4 RESUMEN

La selección de características y etiqueteamiento de probables señales conforme al protocolo BIRADS (malignas o benignas) son muy notables por el descriptor del tamaño, y esta medición se basa en el área de las señales etiquetadas, pero no es suficiente esta información para discriminar y clasificarlas como benignas y malignas, por ello la conformación del vector de características basado en técnicas de magnitudes geométricas o formas morfológicas de cada objeto etiquetado nos permiten en la etapa del clasificador realmente seleccionar si estas corresponden a los requisitos del protocolo BIRADS: circularidad, formas irregular, solidez. En la base de referenciamiento MIAS solo manifiesta que las imágenes son benignas o malignas, pero en la realidad las imágenes son conformaciones de carcinomas benignos, malignos tanto de tamaños pequeños (calcificaciones), grandes o masas tumorales en el procesamiento de las imágenes observamos cómo se presentan las combinaciones de estas señales que el sistema detectara y etiqueta pero solo el especialista considerara cual debe ser el diagnóstico medico.

## 7.6 EXPERIMENTACIÓN EN LA CLASIFICACIÓN Y EL DIAGNÓSTICO

### 7.6.1 TECNICAS PROPUESTAS

En el desarrollo del proyecto se trabajaron tres líneas de posibles clasificadores: algoritmos genéticos, técnicas de lógica difusa y arboles de decisión. De las tres la que se trabajo fue la del árbol de decisiones por la forma en el diseño de las ramas y nodos acordes con las opciones de tamaño y formas. En la figura 40 se muestra el algoritmo propuesto para la clasificación, una vez se clasifican por tamaño (área) se conforma el vector de características visto en la sección 7.5.3 con el cual se procede a discriminar las formas regulares e irregulares las cuales nos dirán si las señales etiquetadas son benignas o malignas (figura 40).

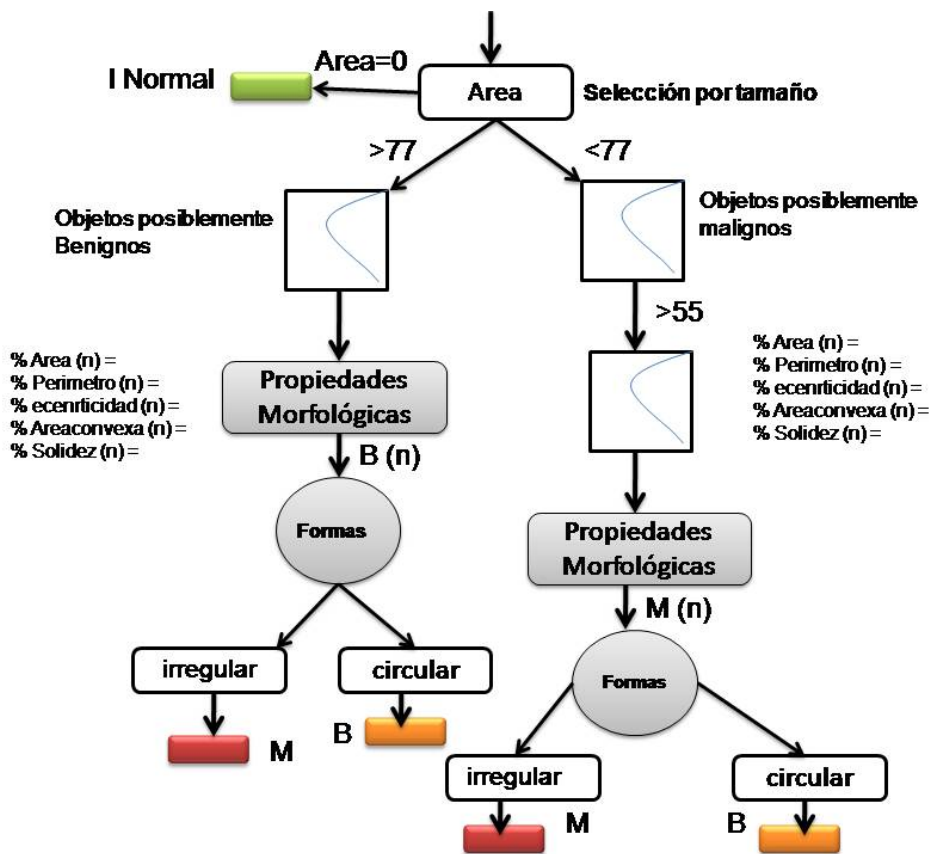


Figura 40. Algoritmo propuesto para la clasificación.

## 7.6.2 DISEÑO DEL ARBOL

Las siguientes son los parámetros adaptivos a considerar para poder identificar las formas en el árbol.

- **Solidez de los objetos:** esta característica geométrica en un objeto nos permite medir que tan sólido es un objeto, este parámetro tiene como rango [1 0] denotando cercano a 1 si el objeto es totalmente sólido y 0 si es irregular.

*Rango de solidez: Benigno: aproximado a Cero*

*Maligno: por debajo de cero*

- **Excentricidad:** esta característica geométrica en un objeto nos permite medir que tan circular o lineal es un objeto, mediante la elipse mide la diferencia entre los dos radios, el parámetro tiene como rango [1 0] denotando 0 para un objeto totalmente circular y 1 para uno lineal.

*Rango de excentricidad: Benigno: cercano a cero*

*Maligno: cercanos a uno*

Acorde con la figura 35, debemos racionalizar al descriptor formal bajo estos dos parámetros, así construimos un árbol de decisiones conformado por 3 niveles y doce ramas importantes de decisión: solidez y excentricidad; para cada (n) de los objetos etiquetados como probablemente malignos o benignos, ver figura 88.

La ruta que más se cumple dentro de BIRADS son:

Maligno: (Sólido irregular & excentricidad lineal)

Benigno: (Sólido regular & excentricidad circular)

Y los rangos que menos se cumplen rangos inciertos, que ameritan una nueva imagen dentro del procedimiento de la mamografía:

Malignos: sólido, circular

Benignos: irregular, lineal

Estas condiciones hacen parte del clasificador.

Para definir los límites de los rangos es importante conocer el comportamiento de dos perfiles de imágenes que tengan similitud a Benigna como lo es la imagen mdb227 y maligna la imagen mdb209. Dentro del proceso de sintonización del clasificador lo más importante adicional a el diseño es los ajustes de los rangos de decisión de los nodos, es por ello que procedemos a hacer esta sintonización.

### ***Imagen mdb227.***

#### *Ventana de selección Benignas:*

Excentricidad:

0,999986576730762 : línea en el corte superior, ruido x corte o propio de la imagen en la toma  
0,419890909963146 : valor menor a 0.5 hacia círculo

Solidez:

1 : solidez completa, ruido en la toma de la imagen  
0,962365591397850 : objeto muy sólido, típico de una calcificación.

Área:

193  
179

Área convexa:

193  
186

El área y área convexa nos muestra como el objeto (1) es una línea ocasionada por el ruido en la toma de la imagen, y el objeto (2) no es un círculo total, sin embargo al comparar la excentricidad vemos que su valor de 0.41 es  $< 0.5$  tendiendo a ser circular.

#### *Ventana de selección Malignas:*

Excentricidad:

0,909268621388100  
0,948432255311706  
0,906022077196367  
0,979795897113271

Los valores tienden a ser lineales.

Solidez:

0,884615384615385  
0,714285714285714  
0,933333333333333  
1

Área:

23  
60  
14  
5

Area convexa:

26

84

15

5

El objeto (4) tiene forma de lineal, muy pequeño y solido, se podria decir que es un ruido en la imagen

### ***Imagen mdb209.***

#### *Ventana de selección Benignas:*

Excentricidad:

0,999991870414406

0,825607431958801

Solidez:

1

0,874213836477987

Area:

248

139

Area convexa:

248

159

El perfil de los dos objetos es casi lineal, los valores de excentricidad son cercaños a 1., luego no son objetos tipicos de una calcificacion en el rango de 77 – 385 pixeles o benignas. Se deben de descartar

#### *Ventana de selección Malignas:*

Excentricidad:

0,594217701544794

0,950581828073644

0,800552514161520

0,678410529355515

0,950217192151204

0,934594196236394

0,765092055676006

0,795822425754222

0,727891344051509

0,632455532033676

0,965715300140749

0,893913908299514

0,978907819428726

0,901281139438573

0,803692864658059

0,989818325811964

0,979373780076978

En el rango de 1-0.8 hay 11 objetos : tienden a ser lineales

En el rango de 0.8-0.5 hay 6 objetos : menor comportamiento lineal.

**Solidez:**

0,935483870967742  
0,823529411764706  
0,914285714285714  
0,911392405063291  
0,782608695652174  
0,661764705882353  
1  
1  
0,9444444444444444  
1  
0,8333333333333333  
0,857142857142857  
0,768292682926829  
1  
1  
0,8000000000000000  
0,882352941176471

En el rango de 1-0.9 hay 9 objetos

En el rango de 0.9 – 0.6 hay 8 objetos

Esto quiere decir que 9 objetos tienden a ser muy sólidos, tan solo 8 son los que tienen perfil típico de una clasificación maligna.

**Area: Areaconvexa:**

29 31  
14 17  
32 35  
72 79  
18 23  
45 68  
5 5  
12 12  
34 36  
6 6  
30 36  
24 28  
63 82  
5 5  
21 21  
36 45  
15 17

Hay 7 objetos no irregulares, los cuales no tienen perfil de ser malignos.



## DISEÑO DEL RANGO DE LOS DESCRIPTORES.

El comportamiento para la clasificación de los objetos *benignos* debe ser bajo estos criterios:

*“Detección de objetos muy circulares y muy solidos”*

El comportamiento para la clasificación de los objetos *malignos* debe ser bajo estos criterios:

*“Detección de objetos muy lineales y poco solidos”*

La figura 41, se muestra como se definen los limites para la decision en los dos (2) nodos de clasificación: solido y excentricidad. Se han separado los limites para la Rama benigna y para la rama maligna, ya que las senales u objetos presentan características diferentes y tambien para poder sintonisar el arbol de deciones en una forma independiente (figura 41).

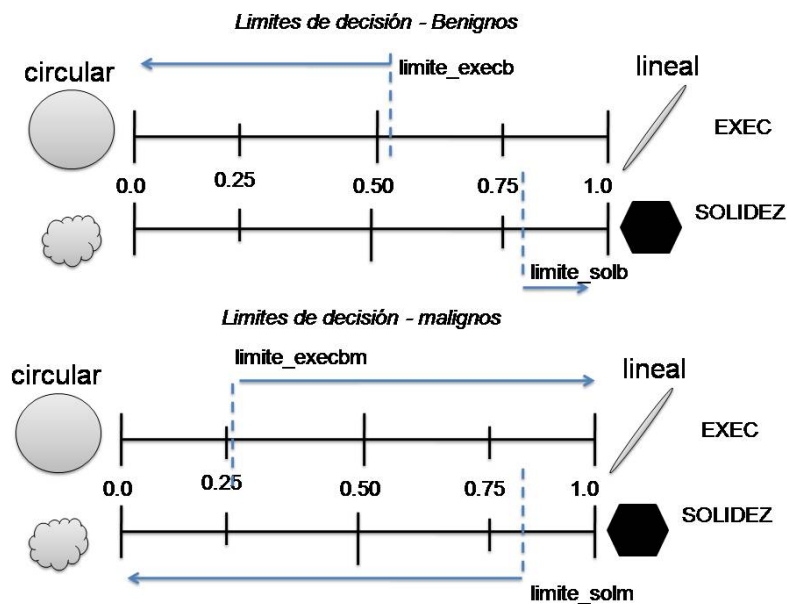


Figura 41. Diseño del rango de los límites de decisión en el clasificador

Una vez hemos diseñado las políticas de los límites procedemos a construir el árbol de decisiones considerando los límites para que los resultados del clasificador sean congruentes con el protocolo BIRADS, así:

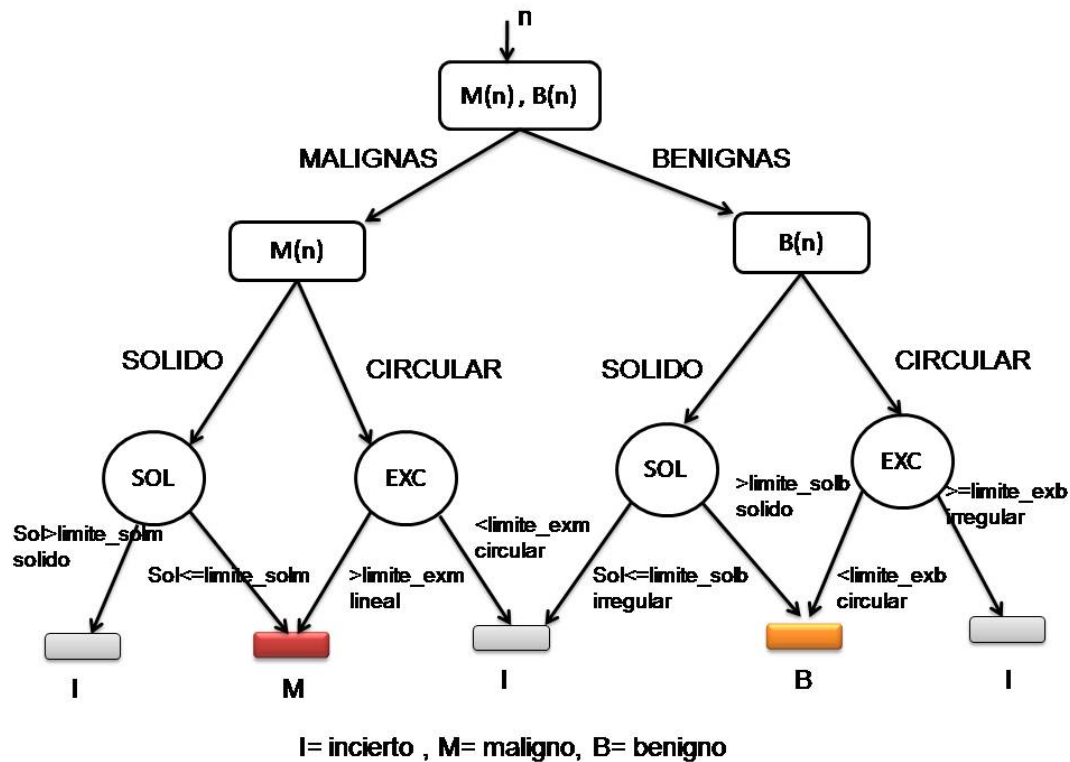


Figura 42. Árbol de decisiones para el clasificador.

El árbol en cada lado tiene dos ramas: Solido y Circular que llegan a un nodo de decisión cada una en la cual se hace la pregunta del comportamiento de las variables Solidez y excentricidad respectivamente (figura 42). En cada nodo de decisión se tiene los limitadores de decisión escogidos y probados manualmente. Las reglas de decisión son:

***If ( Sol < limite\_solM ) & if ( exc > limite\_exeM )      Malignas***

***If ( Sol > limite\_solB ) & if ( exc < limite\_exeB )      Benignas***

Los límites probados en el algoritmo fueron:

```
limite_exeCb = 0.95 ;    %límite de excentricidad objetos benignos
limite_solB  = 0.745 ; %límite de solidez objetos benignos
```

```
limite_exeCm = 0.25 ;    %límite de excentricidad objetos malignos
limite_solM  = 0.935 ; %límite de solidez objetos malignos
```

### 7.6.3 RESULTADOS DE LAS PRUEBAS

Imagen=	mdb227						
<b>DATOS BENIGNOS</b>							
variable		matlab		valor			
total objetos posibles		total_roib		2			
<b>VECTOR CARACTERISTICAS</b>							
descriptor	matlab	n1	n2	n3	n4	n5	n6
Area	arealb(n)	193	179				
Perímetro	perimlb(n)	384	48.04				
Excentricidad	eccentr1b(n)	1	0.41				
Are convexa	areaconvex1b(n)	193	186				
Solidez	solid1b(n)	1	0.96				
<b>CLASIFICADOR</b>							
CIRCULAR	benigno_exec(n)	0	1				
SOLIDO	benigno_sol(n)	1	1				
POSITIVOS	benigno_true(n)	1	1				
INCERTIDUMBRE	benigno_inser(n)	0	0				
TOTAL BENIGNOS(+)	total_benignos	2					
TOTAL B(l)	total_insertid..B	0					
<b>DATOS MALIGNOS</b>							
variable		matlab		valor			
total objetos benignos localizados		total_roim		9			
total objetos reducidos		total_roimin		4			
<b>VECTOR CARACTERISTICAS</b>							
descriptor	matlab	n1	n2	n3	n4	n5	n6
Área	arealm(n)	23	60	14	5		
Perímetro	perimid(n)	17.65	44.38	12.82	8		
Excentricidad	eccentr1m(n)	0.90	0.94	0.90	0.97		
Are convexa	areaconvex1m(n)	26	84	15	5		
Solidez	solid1m(n)	0.88	0.71	0.93	1		
<b>CLASIFICADOR</b>							
CIRCULAR	maligno_exec(n)	1	1	1	1		
SOLIDO	maligno_sol(n)	0	1	0	0		
POSITIVOS	maligno_true(n)	0	1	0	0		
INCERTIDUMBRE	maligno_inser(n)	0	0	0	0		
TOTAL MALIG(+)	total_malignos	1					
TOTAL M(l)	total_insertid..M	0					

Tabla 7. Resultados de prueba del clasificado sobre imagen mdb227

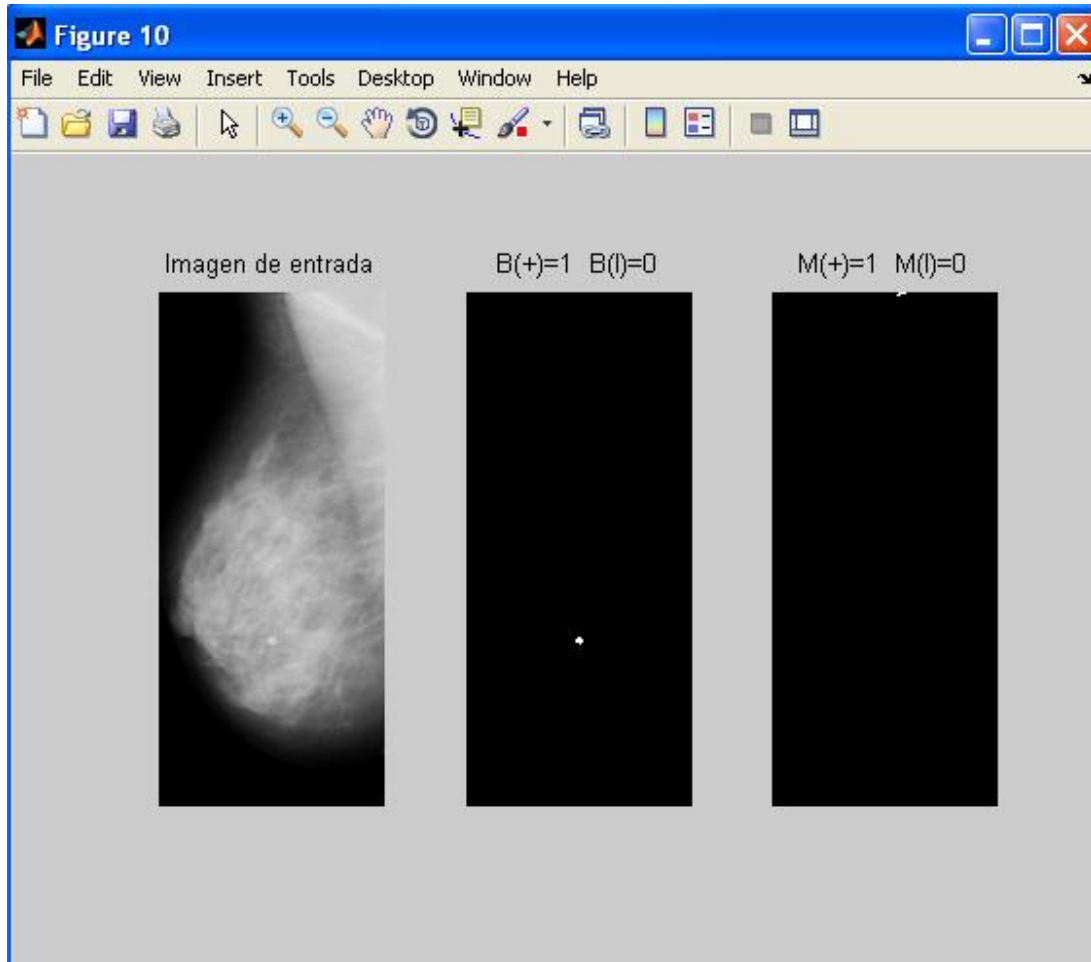


Figura 43. Imagen de entrada y salida del sistema, imagen mdb227

Imagen de prueba del sistema. La entrada es la imagen mdb227 (tabla 7) y se muestran las dos ventanas diferenciando en la segunda figura se muestra las salidas clasificadas de hallazgos benignos, y en la tercera figura las salida de hallazgos malignos.

Esta imagen (figura 43) se caracteriza por tener un comportamiento benigno, según la base MIAS la referencia es:

mdb227 G CALC B 504 467 9

Se observa el resultado en la ventana de benignos, localizando una calcificación con área de 179 píxeles, solidez: 0.96 y excentricidad: 0.41 (ver tabla 9). El hallazgo benigno se localiza en el borde superior y es error en la toma de la mamografía.

IMAGEN	B(+)	B(I)	M(+)	M(I)	VP
mdb227	1	0	3	0	B
mdb222	2	0	0	0	B
mdb223	1	1	2	1	B
mdb226	3	0	4	0	B
mdb209	1	0	10	0	M
mdb211	0	0	3	0	M
mdb213	0	0	2	0	M
mdb233	0	0	5	0	M
mdb014	0	0	4	0	N
mdb018	0	0	1	0	N

Tabla 8. Resultados de las imágenes de entrenamiento

IMAGEN	B(+)	B(I)	M(+)	M(I)	VP
mdb218	0	0	1	0	B
mdb236	0	0	4	0	B
mdb248	0	0	2	0	B
mdb252	0	0	3	0	B
mdb238	1	0	2	0	M
mdb239	0	0	3	0	M
mdb241	0	0	0	0	M
mdb245	0	0	3	0	M
mdb022	0	0	0	0	N
mdb024	0	0	1	0	N

Tabla 9. Resultados de las imágenes de pruebas

- B (+) : objetos localizados como benignos positivos
- B (I) : objetos localizados inciertos en el rango de posibles benignos
- M (+) : objetos localizados como malignos positivos
- M (I) : objetos localizados inciertos en el rango de posibles malignos

VP: valor predictivo según base de datos MIAS, B se referencia como Benigna.

: Valor predictivo según base de datos MIAS, M se referencia como Maligna

: Valor predictivo según base de datos MIAS, N se referencia como Normal

La tabla 8 y 9 respectivamente son los resultados de probar el sistema con las 20 imágenes: 10 de entrenamiento y 10 de pruebas.

Una vez hemos efectuado el proceso de entrenamiento y de pruebas debemos identificar la veracidad de los resultados acorde con la referencia que la base de datos MIAS diagnostica en su información. Por ello debemos identificar en cada resultado del procesamiento y clasificación si esta es un FN o FP.

- Un falso negativo **FN** es cuando existiendo la enfermedad pero en el proceso del diagnóstico no detecta esta, ocasionando que el paciente no sea atendido y permitiendo que la enfermedad continúe su proceso de crecimiento y maduración.
- Un falso Positivo **FP** es cuando no existiendo una calcificación pero en el proceso de diagnóstico se da como detectada. Esto ocasiona tratamientos innecesarios.

La tabla 10 es el resumen del comportamiento del experimento, con los valores de eficiencia y FP, FN.

IMAGEN	B(+)	B(I)	M(+)	M(I)	VP	VN	efic	FP(B)	FN(B)	FP(M)	FN(M)
mdb227	1	0	3	0	B	B	1	-	-	-	3
mdb222	2	0	0	0	B	B	2	-	-	-	-
mdb223	1	1	2	1	B	B	1	-	1	1	2
mdb226	3	0	4	0	B	B	3	-	-	-	4
mdb209	1	0	10	0	M	M	10	1	-	-	-
mdb211	0	0	3	0	M	M	3	-	-	-	-
mdb213	0	0	2	0	M	M	2	-	-	-	-
mdb233	0	0	5	0	M	M	5	-	-	-	-
mdb014	0	0	4	0	N	M	0	-	-	-	4
mdb018	0	0	1	0	N	M	0	-	-	-	1
mdb218	0	0	1	0	B	M	0	-	-	-	1
mdb236	0	0	4	0	B	M	0	-	-	-	4
mdb248	0	0	2	0	B	M	0	-	-	-	2
mdb252	0	0	3	0	B	M	0	-	-	-	3
mdb238	1	0	2	0	M	M	2	1	-	-	-
mdb239	0	0	3	0	M	M	3	-	-	-	-
mdb241	0	0	0	0	M	M	0	-	-	-	-
mdb245	0	0	3	0	M	M	3	-	-	-	-
mdb022	0	0	0	0	N	N	1	-	-	-	-
mdb024	0	0	1	0	N	M	0	-	-	-	1

Tabla 10. Eficiencia, FP, FN en los resultados de 20 imágenes en pruebas.

Con respecto al diagnóstico, clínicamente es muy importante no dejar pasar los FN debido a la prevalencia dejar la posible enfermedad expuesta al crecimiento. Por ello es más importante afinar el sistema a que los FN no se escapen. Por otro lado si se enfatizan los FP esto generar mayores costos en el diagnóstico, pero clínicamente hablando es mejor ante la seguridad en el tamisaje de la enfermedad.

<b>MUESTRAS = 20</b>	<b>RESULTADOS</b>	<b>PORCENTAJES</b>
Hallazgos efectivos	12	60%
Inciertos	2	10%
FP(B)	2	10%
FN(B)	1	5%
FP(M)	1	5%
FN(M)	10	50%
VN	7	35%

Tabla 11. Porcentaje de error en los resultados

**EXPLICACIÓN DE LA TABLA.**

- De las 20 imágenes procesadas el sistema nos identifico acorde con MIAS unas 12 detecciones como positivas, esto quiere decir que el sistema es un 60% eficiente.
- Nos identifico dos (2) imágenes en el rango incierto, lo que genera en el especialista radiólogo su decisión si se requiere realizar de nuevo la toma de la imagen. En incertidumbre estaríamos hablando de un 10% de error en el proceso
- En cuanto a FP(B) se refiere a que a pesar que la imagen anticipadamente es diagnosticada como maligna, esta ha encontrado un signo positivo como benigna. Solo tenemos dos (2) hallazgo entre las 20 muestras, lo que indica que un 10% de error. Es un buen valor para un diagnóstico asistido.
- En cuanto a FP(M) se refiere a que a pesar que la imagen anticipadamente es diagnosticada como maligna, esta ha encontrado un signo positivo como maligna. Solo tenemos un (1) hallazgo entre las 20 muestras, lo que indica que un 5% de error. Es un buen valor para un diagnóstico asistido.
- En cuanto a los FN(M) altos es debido a lo sensible que esta el sistema en la detección de calcificaciones pequeñas, las cuales son las de mayor prevención a ser malignas, y que no se deben dejar escapar, es mejor continuar con los exámenes que dejarlos para los años siguientes. El error es del 50% lo que quiere decir que es muy sensible en los límites del clasificador.
- Los verdaderos negativos fueron siete (7) y son aquellos que debiendo ser malignos no lo fueron, y debieron ser benignos no salieron las pruebas benignas.

## 7.7 RESULTADO EN EL DIAGNÓSTICO) [1,2]

Es evidente que una buena prueba diagnóstica es la que ofrece resultados positivos en enfermos y negativos en sanos. Por lo tanto, las condiciones que deben ser exigidas a un test son (tabla 12):

**Validez:** Es el grado en que un test mide lo que se supone que debe medir. ¿Con que frecuencia el resultado del test es confirmado por procedimientos diagnósticos más complejos y rigurosos? La sensibilidad y la especificidad de un test son medidas de su validez.

**Reproductividad:** es la capacidad del test para ofrecer los mismos resultados cuando se repite su aplicación en circunstancias similares. La variabilidad biológica del hecho observado, la introducida por el propio observador y la derivada del propio test, determinan su reproductividad.

**Seguridad:** La seguridad viene determinada por el valor predictivo de un resultado positivo o negativo. ¿Con que seguridad un test predecirá la presencia o ausencia de enfermedad? Ante un resultado positivo de un test ¿qué probabilidad existe de que este resultado indique presencia de la enfermedad? Veremos posteriormente que esta probabilidad está muy influenciada por la prevalencia de la patología.

Tabla 1. Relación entre el resultado de una prueba diagnóstica y la presencia o ausencia de una enfermedad.

Resultado de la prueba	Verdadero diagnóstico	
	Enfermo	Sano
Positivo	Verdaderos Positivos (VP)	Falsos Positivos (FP)
Negativo	Falsos Negativos (FN)	Verdaderos Negativos (VN)

Tabla 12 [1,2] . Resultado de una prueba diagnóstica y la ausencia o no de la enfermedad

El caso más sencillo que se nos puede plantear es el de una prueba dicotómica, que clasifica a cada paciente como sano o enfermo en función de que el resultado de la prueba sea positivo o negativo. En casos como éste, generalmente un resultado positivo se asocia con la presencia de enfermedad y un resultado negativo con la ausencia de la misma. Cuando se estudia una muestra de pacientes, los datos obtenidos permiten clasificar a los sujetos en cuatro grupos según una tabla 2x2 como la que se muestra en la Tabla 12, En ella, se enfrenta el resultado de la prueba diagnóstica (en filas) con el estado real de los pacientes (en columnas) o, en su defecto, el resultado de la prueba de referencia o “gold standard” que vayamos a utilizar. El resultado de la prueba puede ser correcto (verdadero positivo y verdadero negativo) o incorrecto (falso positivo y falso negativo). El análisis de su validez puede obtenerse calculando los valores de sensibilidad y especificidad que definimos a continuación:



## SENSIBILIDAD

Es la probabilidad de clasificar correctamente a un individuo enfermo, es decir, la probabilidad de que para un sujeto enfermo se obtenga en la prueba un resultado positivo. La sensibilidad es, por lo tanto, la capacidad del test para detectar la enfermedad. Cuando los datos obtenidos a partir de una muestra de pacientes se clasifican en una tabla como la que se muestra en la Tabla 14, es fácil estimar a partir de ella la sensibilidad como la proporción de pacientes enfermos que obtuvieron un resultado positivo en la prueba diagnóstica. Es decir:

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (8.7.1)$$

## ESPECIFICIDAD

Es la probabilidad de clasificar correctamente a un individuo sano, es decir, la probabilidad de que para un sujeto sano se obtenga un resultado negativo. En otras palabras, se puede definir la especificidad como la capacidad para detectar a los sanos. A partir de una tabla como la Tabla 14, la especificidad se estimaría como:

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (8.7.2)$$

Si partimos de los resultados de nuestro experimento podemos proyectar que tan eficiente es el sistema acorde con los índices de sensibilidad y especificidad. Aplicando los resultados de la tabla 14 a nuestro experimento, tenemos:

Cálculos:

Número de Muestras del experimento:	20
Verdaderos positivos VP o hallazgos:	12
Falsos Negativos FN:	1.5 (promedio de FN (B), FN (M))
Falsos positivos FP:	5.5 (promedio de FP (B), FP (M))
VN:	7

$$\text{Sensibilidad} = VP / (VP + FN) = 12 / (12 + 1.5) = 12 / 13.5 = 0.8888 = \mathbf{88.8\%}$$

$$\text{Especificidad} = VN / (VN + FP) = 7 / (7 + 5.5) = 7 / 12.5 = 0.56 = \mathbf{56\%}$$

### VALOR PREDICTIVO POSITIVO [3,4]:

Es la probabilidad de padecer la enfermedad si se obtiene un resultado positivo en el test. El valor predictivo positivo puede estimarse, por tanto, a partir de la proporción de pacientes con un resultado positivo en la prueba que finalmente resultaron estar enfermos:

$$VPP = \frac{VP}{VP + FP} \quad (8.7.3)$$

### VALOR PREDICTIVO NEGATIVO [3,4]:

Es la probabilidad de que un sujeto con un resultado negativo en la prueba esté realmente sano. Se estima dividiendo el número de verdaderos negativos entre el total de pacientes con un resultado negativo en la prueba:

$$VPN = \frac{VN}{VN + FN} \quad (8.7.4)$$

Cálculos:

Número de Muestras del experimento:	20
Verdaderos positivos VP o hallazgos:	12
Falsos Negativos FN:	1.5 (promedio de FN (B), FN (M))
Falsos positivos FP:	5.5 (promedio de FP (B), FP (M))
VN:	7

$$VPP = 12 / (12 + 5.5) = 12 / 17.5 = 0.68 = \mathbf{68\%}$$

$$VPN = 7 / (7 + 1.5) = 7 / 8.5 = 0.82 = \mathbf{82\%}$$

Además de los conceptos de sensibilidad, especificidad y valores predictivos, se suele hablar del concepto de razón de verosimilitudes, razón de probabilidad, o cociente de probabilidades. Estos miden cuánto más probable es un resultado concreto (positivo o negativo) según la presencia o ausencia de enfermedad:

### RAZÓN DE VEROSIMILITUD POSITIVA:

Se calcula dividiendo la probabilidad de un resultado positivo en los pacientes enfermos entre la probabilidad de un resultado positivo entre los sanos. Es, en definitiva, el cociente entre la fracción de verdaderos positivos (sensibilidad) y la fracción de falsos positivos (1-especificidad):

$$RV+ = \frac{\text{Sensibilidad}}{1 - \text{Especificidad}} \quad (8.7.5)$$

### RAZÓN DE VEROSIMILITUD NEGATIVA:

Se calcula dividiendo la probabilidad de un resultado negativo en presencia de enfermedad entre la probabilidad de un resultado negativo en ausencia de la misma. Se calcula por lo tanto, como el cociente entre la fracción de falsos negativos (1-sensibilidad) y la fracción de verdaderos negativos (especificidad):

$$RV- = \frac{1 - \text{Sensibilidad}}{\text{Especificidad}} \quad (8.7.6)$$

Cálculos:

Sensibilidad = 0.888 = 88.8%

Especificidad = 0.56 = 56%

$$RV+ = 0.888 / (1 - 0.56) = 0.888 / 0.44 = \mathbf{2.018}$$

$$RV- = (1 - 0.888) / 0.56 = 0.112 / 0.56 = \mathbf{0.2}$$

La razón de probabilidades ofrece la ventaja de que relaciona la sensibilidad y la especificidad de una prueba en un solo índice. Además, pueden obtenerse razones de probabilidad según varios niveles de una nueva medida y no es necesario expresar la información de forma dicotómica, como resultado de normal o anormal o bien **positivo** y **negativo**. Por último, al igual que sucede con la sensibilidad y la especificidad, no varía con la prevalencia. Esto permite utilizarlo como índice de comparación entre diferentes pruebas para un mismo diagnóstico.

<b>INDICADORES DE DIAGNÓSTICO</b>	<b>VALORES DEL EXPERIMENTO</b>
<i>SENSIBILIDAD</i>	<i>88.8%</i>
<i>ESPECIFICIDAD</i>	<i>56%</i>
<i>VPP</i>	<i>68%</i>
<i>VPN</i>	<i>82%</i>
<i>RV +</i>	<i>2.018</i>
<i>RV -</i>	<i>0.2</i>

Tabla 13. Resumen de los índices de diagnóstico de las pruebas del experimento

El resultado final del experimento se resume en los índices calculados en la tabla 13. el cual nos muestra la información que los especialistas utilizan clínicamente en la validez de la metodología utilizada en las pruebas de un experimento o análisis clínico de laboratorio.

## REFERENCIAS

[1]

[http://www.fisterra.com/mbe/investiga/pruebas\\_diagnosticas/pruebas\\_diagnosticas.asp#Bibliograf%C3%ADa](http://www.fisterra.com/mbe/investiga/pruebas_diagnosticas/pruebas_diagnosticas.asp#Bibliograf%C3%ADa) Pita Fernández, S., Pértigas Díaz, S., Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario Universitario de la Coruña (España), Cad Aten Primaria 2003; 10: 120-124.

[2] Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Epidemiología clínica. Ciencia básica para la medicina clínica. 2ª ed. Madrid: Editorial médica panamericana; 1994.

[3] Morrison AS. Screening in Chronic disease. Second edition. New York: Oxford University Press; 1992.

[4] Altman D.G., Bland J.M. Statistics Notes: Diagnostic tests 1: sensitivity and specificity. BMJ 1994; 308: 1552. [[Medline](#)]

[5] Altman D.G., Bland J.M. Statistics Notes: Diagnostic tests 2: predictive values. BMJ 1994; 309: 102. [[Medline](#)]

## CONCLUSIONES

La figura 44 es el resultado final de las etapas que conforman la solución implementada. Se resalta en líneas entrecortadas los parámetros de variabilidad en el entrenamiento, los cuales pueden ser sintonizados ante otro tipo de imágenes de entrada diferentes a las probadas MIAS. El algoritmo es parametrizable y adaptable.

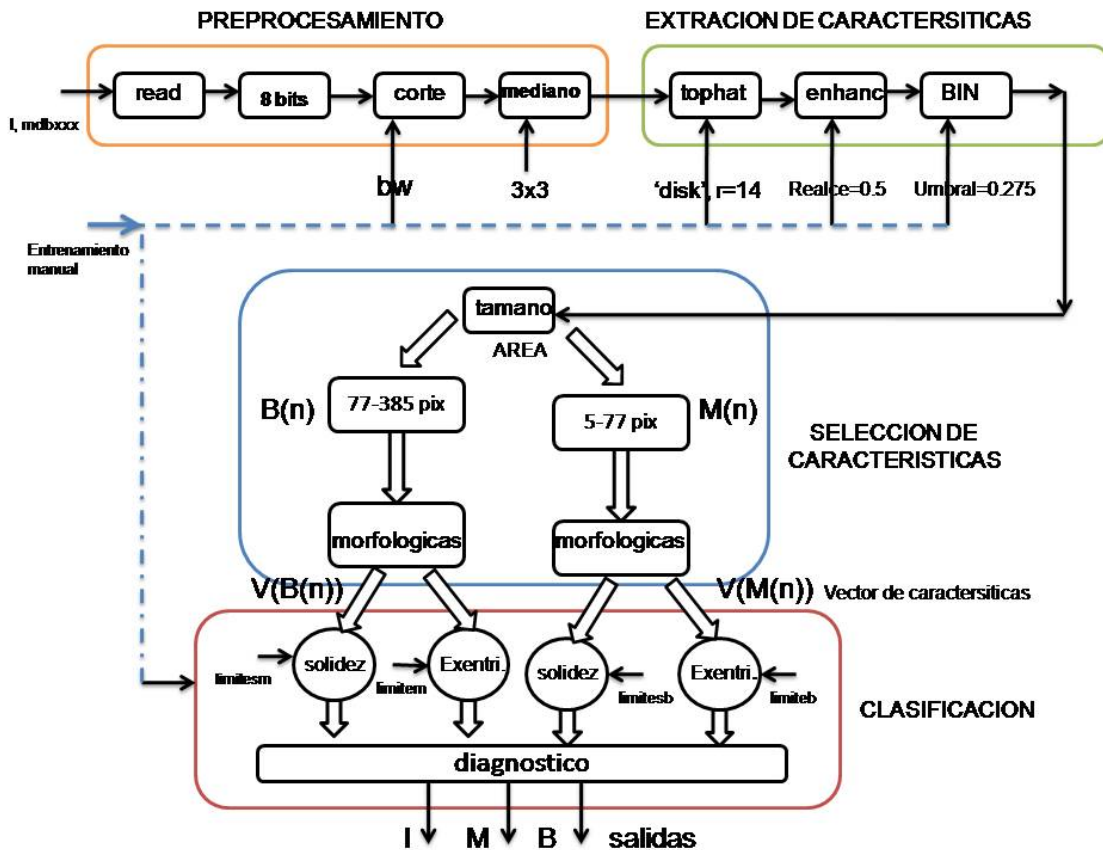


Figura 44. Flujo grama final del algoritmo implementado

De los resultados y experiencia obtenido en este trabajo de tesis de profundización puedo resaltar las siguientes conclusiones y aportes a mi refuerzo y conocimientos adquiridos.

- La detección del cáncer de mama en etapas tempranas es muy importante, técnicas de ayudas inteligentes como lo es un sistema CAD, puede ser una ayuda muy importante a los radiólogos en la lectura de las imágenes digitales.

- Con respecto a la base de datos utilizada MIAS, podemos concluir lo siguiente:
  - De las 322 imágenes en referencia, tan solo el 15% tiene diagnóstico maligno, esto ocasiona muy pocas imágenes para hacer un estudio con mayor número de muestras en el experimento.
  - Tan solo 22 imágenes de las 322 son imágenes con hallazgos de calcificaciones, muy pocas muestras para que el experimento tenga mayores datos de trabajo.
  - En el referenciamiento de la base de datos no se estipula cuantas calcificaciones tienen cada imagen, tan solo se dice el centroide del clúster de ellas. Esto no es una información completa, por ello se tomaron diagnósticos visuales y se contaron los números y centroides de las calcificaciones en cada imagen de las 22 hábiles a trabajar.
  - La validación de las imágenes no se efectuó con un especialista.
- Se escogió el uso de la base de datos MIAS y no DDSM debido al costo en la adquisición y el manejo de la DDSM, a pesar que MIAS es la más utilizada en estudios de este tipo.
- En el pre procesamiento es muy importante la reducción de la región de interés o ROI, resaltando solamente la mama, desafortunadamente en el proceso de digitalización de las imágenes e los sistemas análogos la imagen tiene marcas que ocasionan ruido en el proceso. Las diferentes técnicas de corte de imagen es muy importante, más del 50% de la imagen no es data de trabajo.
- El proceso de segmentación para la extracción de características los métodos morfológicos presentaron un buen comportamiento, en combinación con las técnicas de selección de umbral y realce. La técnica **Top Hat** fue una buena selección en este trabajo.
- En el proceso de selección de características se aplicó la recomendación BIRADS con respecto a tamaño. El descriptor de tamaño utilizado fue área y área convexa, separando acorde con la resolución de las imágenes en dos ventanas, Malignas entre 5 a 77 píxeles y Benignas entre 77 a 385 píxeles.
- Se escogió la técnica de árbol de decisiones como una similitud a la clasificación visual que efectúa un especialista en el descarte de formas y tamaños.
- Para la etapa de clasificación se utilizó la técnica de árbol de decisiones, diseñado en dos ramas cada una con dos nodos de decisión. El descriptor solidez y excentricidad dieron una buena respuesta ante la clasificación de las

formas acordes con el protocolo BIRADS ante calcificaciones malignas y benignas.

- El protocolo BIRADS establece características de forma y tamaño en el diagnóstico, sin embargo es importante sintonizar los descriptores implementados con la ayuda de especialistas, este trabajo solo se baso en la información que la base de datos MIAS en su referenciamiento dice.
- En el proceso de entrenamiento y de afinamiento se logro un buen comportamiento, el índice de sensibilidad fue de 88.8 por ciento, valor clínico confiable. Deseable trabajar este proceso de entrenamiento en conjunto con especialistas.



## TRABAJO FUTURO

- La detección del cáncer de mama en etapas tempranas es muy importante, técnicas de ayudas inteligentes como lo es un sistema CAD, puede ser una ayuda muy importante a los radiólogos en la lectura de las imágenes digitales. Sin embargo los sistemas CAD no se han popularizado en nuestro país, a pesar de existir varias soluciones maduras en el mercado, es una buena oportunidad de profundizar mas sobre esta temática con miras a continuar con este trabajo.
- El proceso de segmentación para la extracción de características los métodos morfológicos presentaron un buen comportamiento, en combinación con las técnicas de selección de umbral y realce. La técnica Top Hat fue una buena selección en este trabajo, sin embargo sería bueno seguir trabajando en metodologías de combinaciones mixtas con filtros gaussianos DOG.
- El ruido en el proceso de toma de las mamografías, las técnicas de corte automático es una etapa que aun puede ser mejorada reduciendo el ROI solo a las regiones en donde existen posibles calcificaciones.
- El protocolo BIRADS incluye otras características a ser incluidas en el análisis: determinación de clúster o proximidad de grupo de calcificaciones como un descriptor adicional a considerar en este trabajo, considerando en un área de 1 cm<sup>2</sup> en el actual si se detectan más de cinco 5 se considera un clúster maligno, proceso que puede ser incluido con técnicas de las **K-medias**.
- A pesar que el procedimiento implementado es parametrizable a otras bases de datos, sería bueno trabajar con la base de datos **DDSM** y otras de localización en hospitales universitarios de nuestro país: HUSH (Hospital Universitario San Ignacio). Esta sería la mejor forma de mejorar la cantidad de imágenes para mejores resultados del experimento.
- Debido a la complejidad en el análisis de las imágenes es importante trabajar con un equipo interdisciplinario con miras a mejorar los resultados, perfiles como: medico radiólogo, Medico patólogo, Ingeniero de sistemas experto en programación de OpenCV con miras a migrar la solución a un sistema GNU basado en plataforma UNIX o LINUX.
- Migrar la plataforma de desarrollo a sistemas abiertos que permitan una interconexión con estándares como: DICOM 3.0, HL7, sistemas PACS y RIS, plataformas sobre LINUX (suse, Centos), sistema de desarrollo OpenCV, SAGE y lenguaje de desarrollo como: java, Python, C++
- Se propone una continuidad basada en un sistema GUI (graphics interface) con el especialista quien pueda reducir el ROI y seleccionar mejor los límites de los umbrales de cada etapa (ver figura 45).

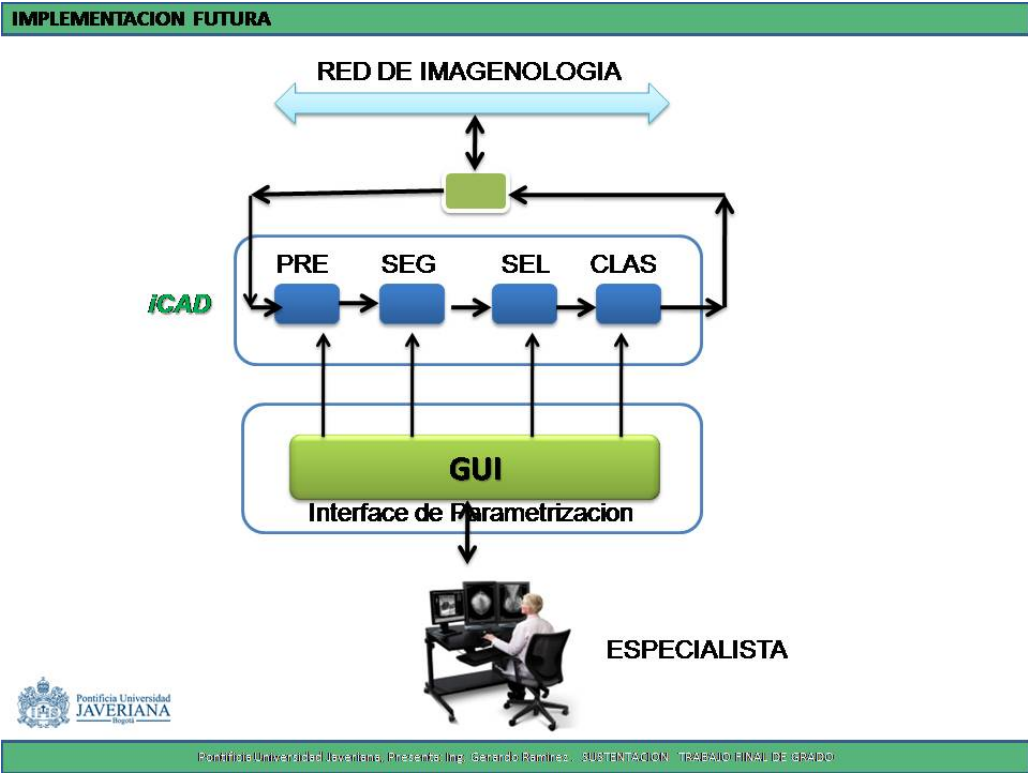


Figura 45. Implementación propuesta como trabajo futuro.

# **GLOSARIO**

## **ANEXOS**

## GLOSARIO

CAD	Computer Added Detections, ayuda asistida por programa computacional.
PACS	Picture Archive Communications Systems
DICOM	Digital Image Communications on Medici
MIAS	Base de datos de la: Mammographic Image Analysis Society
BIRADS	Breast Imaging Report and Data Base Sytem
ARS	American Radiology Society
PATOGENESIS	La Patogénesis describe en general el origen y evolución de una enfermedad (Patología)
NEOPLASIA	Ciclo de crecimiento de las células
CARCINOMA	Crecimiento desenfrenado de células malignas
TUMOR	Grupo grande de células cancerígenas.
CALCIFICACION	Grupo pequeño de células cancerígenas.
METASTASIS	replica de células malignas a otros tejidos y/o órganos del cuerpo
MAMOGRAFO:	Equipo de radiografía que toma imágenes del seno.
MAMOGRAFIA:	Imagen digital que muestra al seno. Es la salida de un equipo de Mamografía

## ANEXO 1

### ENFERMEDAD DEL CANCER

#### A-1.1 INTRODUCCIÓN

La enfermedad del cáncer son la consecuencia del crecimiento desenfrenado de células, las cuales pueden convertirse en malignas o benignas a este fenómeno también se le denomina NEOPLASIA. Las células dentro de su ciclo en el cual crecen y se dividen haciendo un proceso normal a células hijas:

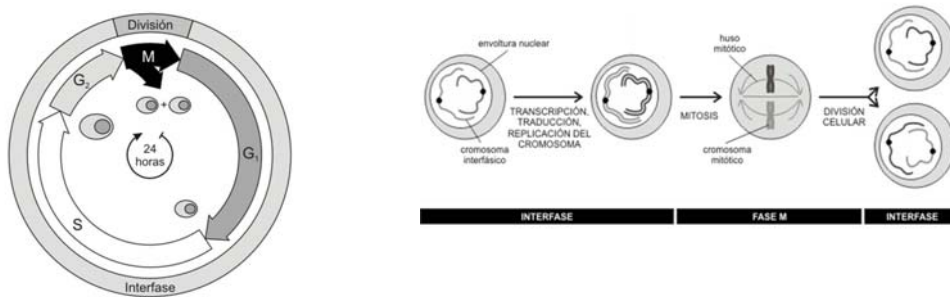


Figura A1-1 [1] – división celular

Dentro del ciclo existen cuatro intervalos en donde la célula opta por dividirse, muchas células pasan su vida entre los ciclos: G<sub>1</sub>, S G<sub>2</sub> antes de optar por dividirse, y es entre la etapa G<sub>2</sub> a G<sub>1</sub> cuando se efectúa la división (conocida como etapa M), no siempre el ciclo se da. El Cáncer comienza cuando dentro del ciclo de vida y de división de una célula se vuelve un proceso desenfrenado ocasionando crecimientos de grupos de células que se convierten en tejidos y posteriormente en agrupaciones que se vuelven tumores. En esta sección veremos como el Cáncer es una enfermedad compleja

#### A-1.2 HISTÓRIA

Posiblemente la primera referencia del cáncer de Seno se remonta a un documento de aproximadamente 3.000 ac. Escrito por Edwin Smith [2] en forma de papiro, cuya autoría se refiere a **Imhotep** arquitecto egipcio el cual manifestaba ocho referencias de diversas enfermedades de la glándula mamaria como abscesos, úlceras, tumores y traumatismos. En 1730 AC. El Código de Hammirabi cita más de 200 referencias de casos de mutilaciones mamarias como reprimenda o castigo por inferir ante el código. Pero en el año 525 ac. **Democedes** curó a **Atosa** esposa de **Dario I** rey de Persia, de un tumor ulcerado del seno.



Figura A1-2 [2] papiro de Edwin Smith , El martirio de Santa Águeda

Virgilio en la Eneida hace referencia a las amazonas que se quitaban sus glándulas mamarias y Tolomeo relata que algunas amazonas amputaban a sus hijas pequeñas del seno derecho para facilitarles acciones militares. **Hipócrates** en el año 400 ac. Hace dos referencias al cáncer mamario; en el **Epidémico** I y el VII adelantándose mucho a la época dicto tres preceptos: los tumores que se curaban con las medicinas eran los de mejores pronósticos, que los que no se curaban con las medicinas necesitaban cuchillo, y que después requerían fuego.

Ya a comienzos de la era cristiana entre los años 30 ac. – 38 dc. El escritor **Celso Cornelio Aulo** hace una revisión enciclopédica sobre la práctica de la medicina y clasifica en tres los tipos de médicos: Dietistas, Farmacéuticos, Cirujanos, según la terapia que empleaban y al referirse al cáncer seno. **Galeno de Pergaño** (200 dc.) Fue seguramente el que realizó el primer tratamiento quirúrgico para el cáncer del seno y en comparación magistral comparo el tumor con un cangrejo (cáncer) que envía sus ramificaciones a diversas partes y recalco la necesidad de que la resección fuera distal a estas ramificaciones, que aumentaban su peligrosidad cuando están cercanas a las venas o arterias (mas adelante esto se conoce como un facilitador de las metástasis). Otra anécdota interesante en la historia del cáncer de mama es la de cristiana **Santa Águeda** en el siglo III dc. Quien **Quintiliaño** gobernador de Sicilia le mando a amputar los dos senos por no ceder a sus pretensiones amorosas. Dice la tradición que como ella poseía extraordinarios poderes de curación a los cuatro días sus senos regresaron a lo normal; por ello es considerada la santa patrona de las nodrizas y de mujeres con patología mamaria, en la figura [2] se aprecia una pintura del acto.

En el siglo XIV Henry de **Mondeville** y **Guy de Chauliac** distinguieron entre canceres tempranos y los tardíos, preconizando en el primero cirugía y cauterización y en los canceres pequeños el uso de pastas con arsénico y zinc.

En el siglo XVI aparece **Ambroise Pare** considerado el padre de la cirugía moderna quien reconoció la relación entre el tumor primario y la axila correspondiente, el

flamenco **Vesalio** practico amplias incisiones quirúrgicas y controló la hemorragia con ligaduras en el mismo siglo.

En el siglo XVIII **Jean Louis Petit** comienza la nueva era de la cirugía mamaria en Francia y hace hincapié en la extirpación de los ganglios axilares y de la fascia pectoral; es decir El comenzó la era de la cirugía radical (hemostasia y mastectomía- quitar áreas del cuerpo).

En los Estados Unidos **Joseph Pancoast** en el Jefferson Medical College, **Samuel D. Cross**, enfatizaron la necesidad de que el seno y los ganglios axilares comprometidos fueran extirpados en una sola pieza. A finales del siglo pasado sucesivos cirujanos de diversos países sentaron las bases anatómicas definitivas para la estandarización de unos procesos técnico que llegó con Mr. **William S. Halsted** al perfeccionamiento de la mastectomía radical. Procedimiento que imperó más de 60 años en todo el mundo. En las últimas décadas los trabajos de **George Crile** y **Mc. Writer V peter, B. Fisher** y muchos más, han demostrado las ventajas de procedimientos menores y existe el concepto de que la enfermedad del cáncer de seno probablemente desde sus estados iniciales sea una enfermedad sistémica.

En las dos últimas décadas del siglo XX y este nuevo siglo XXI se ha vivido un nuevo enfoque más fisiopatológico y nuevos estudios sobre la cinética celular y la hormona dependencia y los nuevos protocolos de quimioterapia han abierto nuevos horizontes en el manejo del cáncer mamario.

En Colombia los desarrollos de la atención a la enfermedad han sido liderados por el Instituto Nacional de Cancerología [3] y las ligas contra el cáncer las cuales funcionan 45 sedes localizadas en todo el país [4].

### **A-1.3 PATOGENESIS**

La Patogénesis describe en general el origen y evolución de una enfermedad, y la acción de la misma se denomina Patología de la enfermedad. Los procesos de análisis desde la misma sangre y de muestras de tejidos (biopsias) permiten establecer el estado de la enfermedad y una predicción de su evolución. La patología del cáncer es causada como comentamos anteriormente por el crecimiento desenfrenado en el ciclo de vida de las células ocasionando índices del tipo: anormal, con sobrecrecimiento en relación a los tejidos adyacentes [5], descoordinada y sin control, sin finalidad aparente y agresivo con una tendencia muy alta de ser irreversible.

El uso ha establecido como sinónimo la palabra TUMOR (abulta gran cantidad de tejidos enfermos o muertos) como parte de este proceso de crecimiento y duplicación de las células denominadas la Neoplasia o nuevo crecimiento. De una forma simple y elemental acorde con el pronóstico estas pueden ser clasificadas como malignas o benignas y a las malignas en forma general se les denomina CANCER.

Entonces una neoplasia maligna (cáncer) son las células que se multiplican sin control y pueden invadir los tejidos vecinos mediante el transporte por el flujo sanguíneo y el

sistema linfático. **El Carcinoma** es la neoplasia maligna que empieza en el tejido: piel, órganos vitales, glándulas revistiéndolos y contaminándolos. **El sarcoma** es una neoplasia maligna que empieza en el Hueso, cartílagos, grasa, músculos. La leucemia es una neoplasia maligna que empieza en un tejido donde se forman las células sanguíneas como lo es la medula ósea y hace que se produzca un gran número de células sanguíneas anormales y que estas se transmitan por todo el organismo.

Para reconocer si las neoplasias son normales o anormales se aplican criterios morfológicos mediante la diferenciación.

En la figura 1 se observa como es el ciclo de vida de una célula y la podemos ver como una cadena infinita y compleja de nacer-desarrollarse-crear-morir es el comportamiento natural de la vida, así en forma micro la célula dentro de este ciclo crea una nueva célula la cual entrega copia de su información como lo es el DNA. Sin embargo algunas en el proceso de esta división denominada Mitosis es en donde esa información se codifica mal, ocasionando una copia diferente o replicación de la célula, las células se comunican entre sí célula a célula permitiendo que con este contacto respondan a la sensación de ser vecinas. Por eso las células cáncer son las descendientes de una célula normal en la cual algo ha entrado mal. Esta célula dañada con cáncer es diferente a las vecinas y tienden a ser separadas por la comunidad, es un fenómeno muy complejo similar a las comunidades sociales y étnicas, suena curioso pero es así. Pero lo ideal es que la célula mala o enferma se muriera, cosa que no es así, esta sigue su ciclo de vida y a pesar que la comunidad de células buenas empiezan a aislarla esta empieza a dividirse y crear una comunidad de células malas o cancerígenas que se convertirán en un TUMOR (procesos que se denomina Metástasis).





A



B



C



D

D

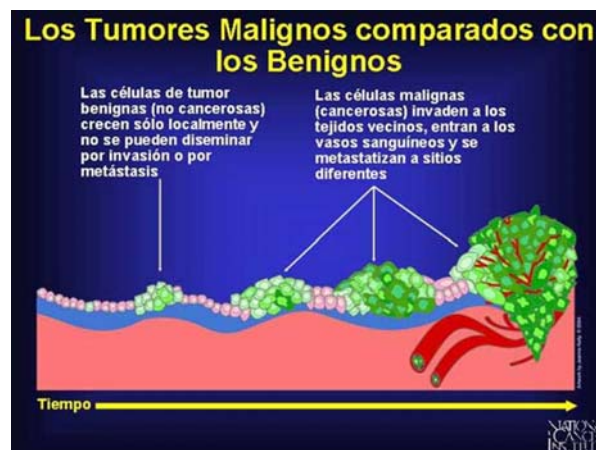


Figura A1-3 [6]: A) proceso del ciclo de crecimiento. B) crecimiento Normal. C) Crecimiento cancerosos. D) proceso invasivo. E) tumores malignos y benignos

Dentro del estudio de la enfermedad del cáncer una de las preguntas y dilemas que aun es un misterio son los agentes que permiten estimular el crecimiento y proliferación de estas células y se está trabajando en la genética de cada persona en la cual la información se trasmite de generación en generación. Sin embargo no solo el mensaje codificado que viene de generación en generación se requiere un ambiente de crecimiento que estimule la proliferación de las neoplasias malignas, y se ha logrado comprobar que estímulos acordes con la alimentación, el stress, el medio ambiente y estilos de vida así como el tabaco, licor uso de sustancias químicas hacen que el proceso de crecimiento continúe y acelere la enfermedad.

#### A-1.4 TIPOS DE CANCER

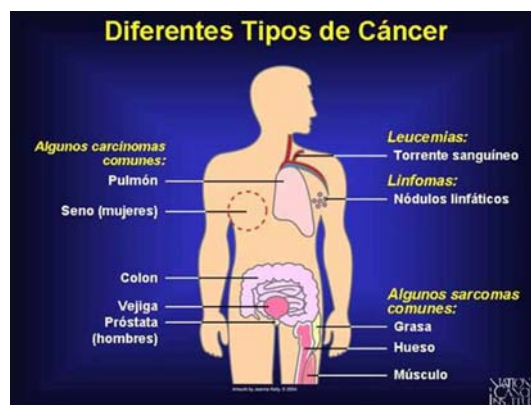


Figura A1-4 [6] Localizaciones de focos de cáncer

Podemos clasificar los diferentes tipos de cáncer acorde con los sistemas a los cuales la neoplasia regenerativa genera el carcinoma respectivo, estas áreas las podemos clasificar como:

- Sistema hematológico: cáncer de la sangre o leucemia
- Aparato respiratorio y tórax: aparato cáncer de pulmón y tórax
- Sistema nervioso: medula espinal, cerebro.
- Sistema digestivo y gastrointestinal: estomago, colon, año, hígado, páncreas, próstata.
- Cutáneo: piel en todas su modalidades
- Seno
- Aparato locomotor: Huesos
- Otros sitios: ojo, endocrino, urinario.

En general se han documentado y diagnosticado más de 100 tipos de cáncer

### A-1.5 PATOLOGÍA DEL CÁNCER DE MAMA

El cáncer de seno o de mama también denominado es la manifestación de neoplasias carcinomas localizadas en las glándulas mamarias de hombre: mujer y hombres. El mayor índice se presenta sobre mujeres y principalmente en las edades de 30-60 años.



Figura A1-5 [7] anatomía del seno

Las glándulas mamarias están compuestas por partes externas e internas. Externamente observamos el pezón y su areola, la piel epitelial circundante del órgano que está localizado en el pecho. Interiormente es un sistema completo de tejido adiposo compuesto por lobulillos, tejido graso, conductos lácteos, conductos linfáticos y músculo interno. La glándula mamaria (seno) histológicamente hablando posee dos elementos fundamentales: aciaños glandulares en donde se produce la leche a través de los ductos mamarios y un conjunto de estructuras tubulares y huecas en forma de árbol.

Con respecto a las posibles causas del cáncer de mama no es tan clara la información que se tiene como causa efecto no uno nominalmente dominante, sin embargo el factor hereditario (genético), efectos hormonales, menarquía temprana (primera menstruación), alimentación, menopausia tardías, edad entre 35 a 60 años, y otro factor muy importante es el estilo de vida: dieta rica en grasa y carne, obesidad, sedentarismo, alcohol y tabaquismo y el factor stress.

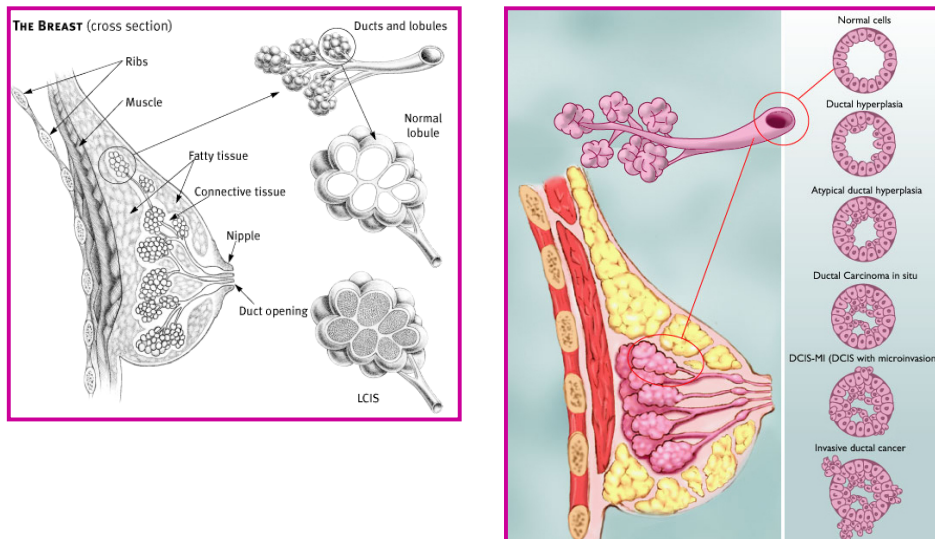


Figura A1-6 [ 7] carcinoma lobulillar en situ

El cáncer de mama es el resultado del crecimiento desenfrenado en las divisiones de las células o neoplasias generando células malignas o denominadas carcinomas las cuales por lo general se generan en los ductos mamarios o glándulas mamarias. Estas células se van acumulando en pequeñas colonias con un largo periodo de vida en algunos casos inmortales caso que las hace muy peligrosas.

Existen dos tipos de clasificaciones de mama: carcinoma **Acinar o Lobular** y carcinoma **Canalicular o Ductal**

La gran mayoría de los canceres de las glándulas mamarias se originan en el epitelio ductal o acinar de la glandula; numerosas clasificaciones se han hecho y desde el punto de vista práctico más del 90% son de origen canicular y menos de un 10% de origen lobulares. Cada variedad puede ser **In Situ o Infiltrante**.

El carcinoma **in situ** se identifica por ser el que no ha roto el ducto basal y por esta razón no se ha replicado y extendido. Estos tienen un proceso de tratamiento menos complejo y se basa más en la extirpación del Tumor. Tenemos los siguientes tipos de sub especies de cáncer de mama Lobular:

- Ductal in situ: este se origina en las células de las paredes de los conductos mamarios. Es un cáncer muy localizado y el cual no se ha extendido a otras zonas ni ha hecho metástasis por este motivo esta enfermedad es Pre maligna y puede extirparse fácilmente, la tasa de curación puede ser del 100% y es muy fácil de detectar mediante la mamografía. Tipos de carcinomas denominados: cruciforme, micro papilar, solido o microacinar, papilar.

- Lobulillar in situ: se origina en las glándulas mamarias o lóbulos, y a aunque no es un verdadero cáncer aumenta el riesgo de que se convierta en el futuro, este suele dar en la edad de menopausia.

El carcinoma ductal infiltrante es el cáncer de mayor representación (70% - 90%) el más común y su carácter es invasivo creciente y bloquea los ductos mamarios creando tumores con límites no claramente delimitados como veremos en el análisis de la imagen de mamografía este aparece con bordes mal conformados, existen las siguientes sub especies:

- Ductal infiltrante: es un cáncer invasivo y se inicia en el conducto mamario que logra atravesarlo y pasa al tejido al tejido adiposo de la mama y luego puede extenderse a otras partes del cuerpo. Es el más frecuente y se da en un 80%. Tipos de carcinomas frecuentes: Medular, coloideo, cruciforme, tubular, adenoideo quístico, meta plástico, apocrino, secretor, enfermedad de paget, carcinoma inflamatorio

Hoy en día se sabe que una célula cancerosa de mama tiene como tiempo de duplicación entre 100 a 300 días, así mismo una neoplasia de mama de 1 Cm cubico de tamaño sufre cerca de 30 duplicaciones antes de alcanzar este tamaño, y este cáncer tiene como mínimo 7 años de evolución [11]. Estas cifras son aterradoras ante el carácter peligros y mórbido de esta enfermedad.

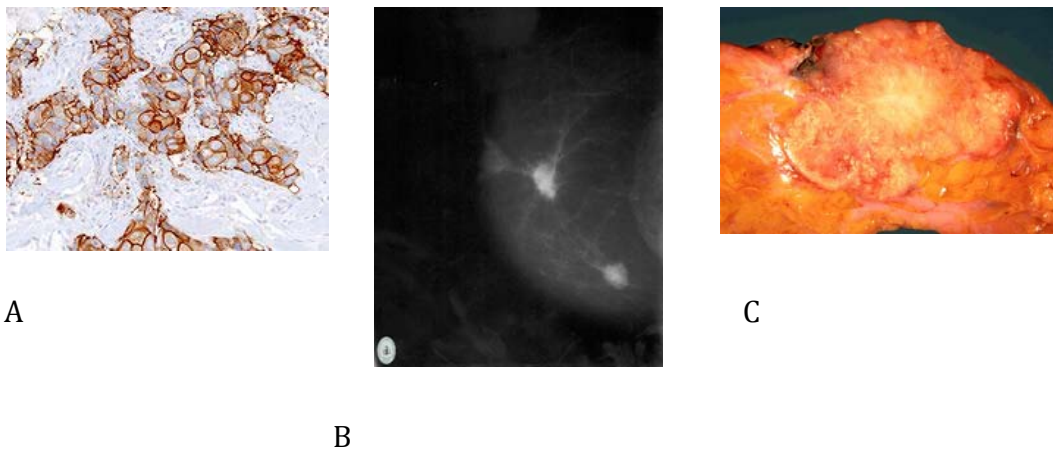


Figura A1-7 [ 7 ] Carcinomas ductal infiltrante: A) Colonia celular necrósica, B) Imagen Rx, C) Tumor extraído.

## A-1.6 EPIDEMIOLOGÍA

La epidemiología es la ciencia que estudia el comportamiento de la enfermedad, su tendencia y proliferación de la misma en la comunidad. El cáncer de mama es una enfermedad con gran impacto mundial, dado que es una de las patologías con más alta prevalencia en mujeres mayores de 50 años de edad y el cáncer con mayor tasa de mortalidad en mujeres en casi todos los países; la OMS (organización mundial de la salud) dictaminó en el 2004 que el cáncer de mama es la mayor causa de muerte por neoplasia en mujeres [8]. Sin embargo el cáncer de mama genera el 10% de todas las neoplasias [9], la incidencia es mayor en los países desarrollados en donde es hasta seis veces más alta que los países en vía de desarrollo; Latinoamérica tiene una incidencia intermedia [10], en EEUU las mujeres de familia latina tiene menores incidencias y mortalidad en comparación con las no latinas, lo cual sugiere que la raza es un factor selectivo. En Colombia es el segundo cáncer más frecuente en mujeres (primero es el de ovario) y en los departamentos: Antioquia, Arauca, Atlántico Bolívar y valle el índice es del 23% de las muertes por cáncer. En Bogotá es la decimosegunda causa de muerte a nivel general.

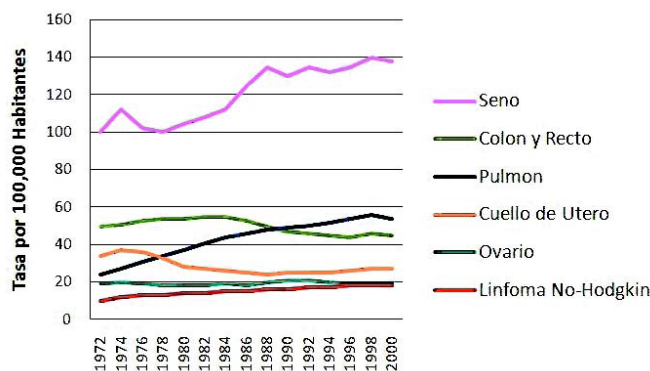


Figura A1-8 [ 8] incidencia de cáncer de mama ajustada a la edad en los EEUU desde 1973 al 2000

Es el Tumor quien tiene la tasa de incidencia más alta en los últimos 30 años y en Colombia en donde paso del quinto al segundo lugar en frecuencia y en EEUU donde actualmente se constituye el 26% de nuevos casos por neoplasias [8]. Se diagnostican más de un millón de nuevos casos en el mundo y más de 4.4 millones de mujeres que padecen esta enfermedad en el mundo. En España de igual forma ocupa el segundo nivel de mortandad representando el 22% de las defunciones. Es una enfermedad exclusiva de las mujeres pues tan solo el 1% ataca a los hombres. La incidencia se mantiene relativamente igual hasta los 30 años y ataca a crecer sustancialmente en cada país hasta las 60-70 años y haciendo que casi el 75% se dé en los 55 años:

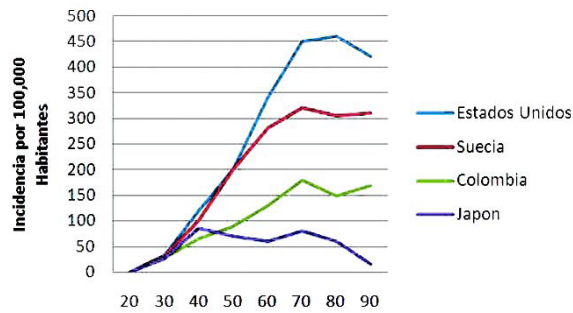


Figura A1-9 [8] incidencia del cáncer de mama por edad en 4 países.

Si se observa que las mayores consultas que hacen las mujeres en orden de frecuencia son: Masas (46%), mastalgia (28%), nódulos (16%) y telorragia (6%). Y solo se comprueba cáncer en un 4.5 % a 8.7% de los pacientes sin importar cual sea el motivo de consulta, es así que la presencia de la masa (TUMOR) es el síntoma que mas relaciona la enfermedad.

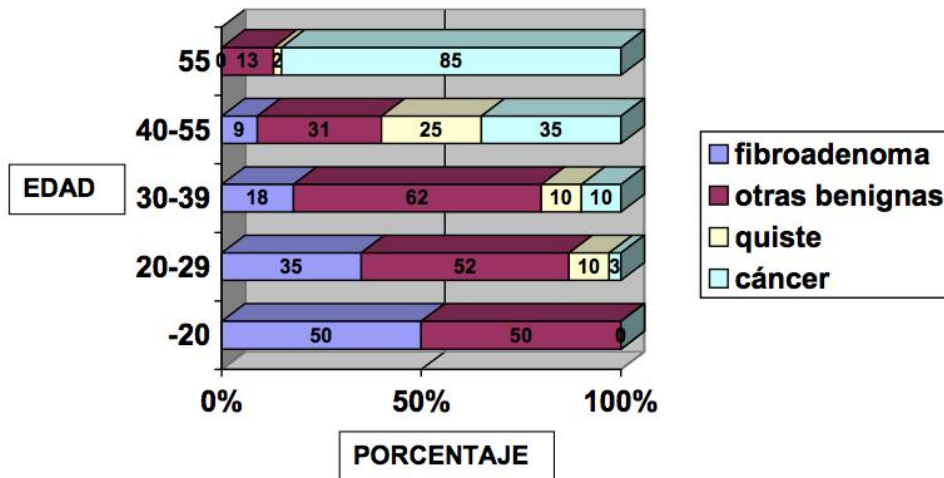


Figura A1-9.1 [21 ]. Etiología de las masas de seno según la edad.

## A-1.7 DIAGNÓSTICO

El diagnóstico temprano es la mejor vía de cura y de poder parar el proceso de la enfermedad, al ser los senos un órgano de fácil acceso es susceptible de un diagnóstico rápido - No Invasivo - para ser sometida a un tratamiento precoz y adecuado, sin embargo en nuestro medio (Colombia) la realidad es otra, no existe una conciencia preventiva hacia los exámenes de rutina si no cuando ya el síntoma ha avanzado, el diagnóstico básico comprende:

- **Auto examen de mama:** es el examen que no causa ninguna incomodidad ni costo, consiste en detectar pequeñas masas y dolores mediante el desplazamiento de masajes entorno a los senos en busca de masas, ganglios o dolor. Se debe práctica cada mes una semana después del inicio de la menstruación.
- **Examen clínico:** examen de mama desarrollado por un médico especialista para verificar cambios en el tamaño del seno y movilidad de los mismos. Se miran aspectos del color de la piel y de los pezones, apariciones de ganglios en las axilas.
- **Mamografía:** es el examen de radiografía de baja radiación considerada no invasiva en la cual un medico radiólogo hace una lectura e interpreta los rasgos en búsqueda de rasgos morfológicos que den indicios de la presencia de carcinomas en sus diferentes modalidades. Se ha calculado que la tamización de la mimeografía reduce en un 20 al 35% la tasa de mortalidad

El método más confiable es la mamografía y de mayor precisión para la detección temprana del cáncer de mama, por ello se recomienda empezar a hacer este procedimiento después de los 38 a 40 años cada dos años y después de los 50 años cada año.

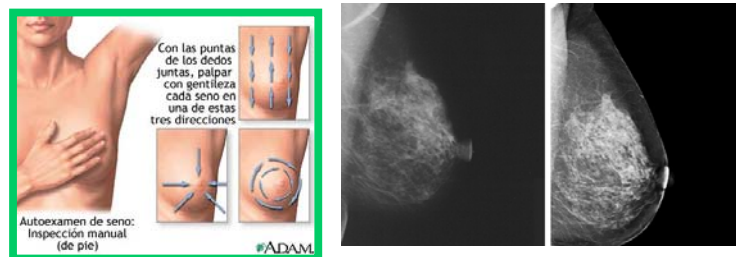


Figura A1-10 [7] Tipos de diagnóstico temprano para detección de carcinomas: auto examen, mamografía

El objetivo primordial del examen es detectar anomalías las cuales son de dos tipos: neoplasias de tamaño grande (masas) denominadas tumores cancerosos y neoplasias de tamaño pequeñas las cuales son pequeñas agrupaciones de calcificaciones denominadas calcificaciones.



### TAMANOS DE CARCINOMAS

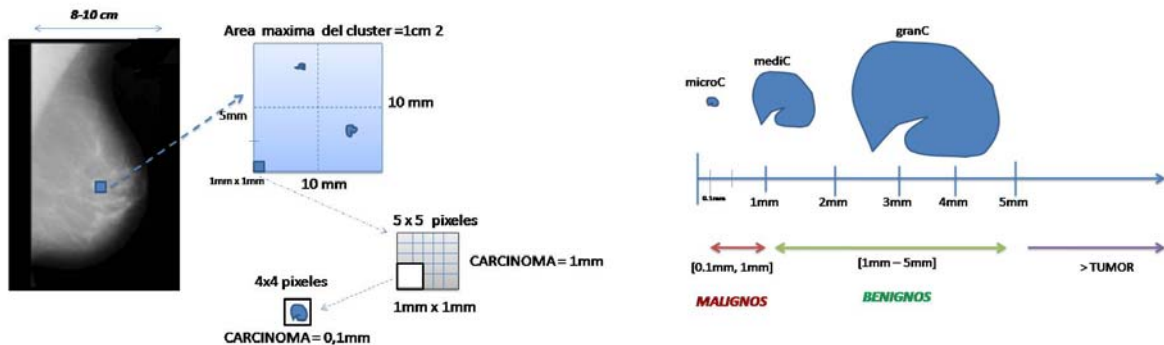


Figura. A1-11 [12] dimensiones de carcinomas, tumor y micro carcinomas

[19] Uno de los procesos en el diagnóstico temprano es la interpretación de las imágenes está basado en un análisis morfológico complejo basado es diferentes vectores: tamaño, textura, formas, con miras a localizar masas, calcificaciones, distorsiones arquitecturales, áreas de densidades asimétricas entre las mamas y ductos lactíferos predominantes [ 14]. La sensibilidad de la mamografía (es la probabilidad de un resultado anormal en presencia de cáncer de mama) es altay está entre el 85% a 95%, en relación al examen clínico que es del 50%. Dependiendo de la edad la sensibilidad se incrementa. Por otro lado la especificidad de la mamografía (probabilidad de un resultado normal es ausencia de la enfermedad) es baja y está entre el 15% y el 30%. Alrededor de las mamografías practicadas en mujeres sin cáncer mamario (resultados de falsos positivos) sugieren su presencia, provocando ansiedad, exámenes complementarios y biopsias mamarias innecesarias que además suben el costo en el tratamiento [15].

Entre el 2% y el 22% de los resultados positivos obtenidos mediante una mamografía a la primera vez, requirieron nuevas evaluaciones para confirmar los resultados y entre el 12% y el 78% requirieron una biopsia [16] . Dada la baja especificidad de la mamografía, el diagnóstico complementario es necesario, la técnica frecuentemente usada es la biopsia, debido a su alta especificidad.

La elevada sensibilidad pero baja especificidad de la mamografía en el diagnóstico del cáncer de mama se puede deber a [17]:

1. La baja diferenciación en la apariencia del tejido canceroso comparado con el Tejido parénquima normal, en especial cuando el tejido predominante en la mama es denso.
2. La variada morfología de los hallazgos, muchos de ellos no relacionados con el cáncer.
3. La semejanza entre las morfologías de los hallazgos.

4. El variado tamaño de los hallazgos.
5. Posibles deficiencias en la técnica para tomar la radiografía, entregando imágenes de baja calidad.
6. La fatiga visual o distracción del radiólogo.

Se ha demostrado que el doble diagnóstico mejora la sensibilidad a lo mas en 15% [18] mientras que cada radiólogo puede perder un pequeño número de casos, el otro los puede detectar. Pero como cada mamografía debe de ser revisada por dos radiólogos, este procedimiento es ineficiente desde el punto de vista práctico debido a la escasez de especialistas y que la productividad individual se ve reducida. Tomar una placa radiográfica puede tomar entre 20 a 30 minutos, pero diagnosticar una mamografía puede tomar al menos 40 minutos. Este tiempo se puede incrementar si es que se encuentran regiones sospechosas que no se pueden declarar fácilmente como procesos benignos.

Una alternativa viable es que el otro radiólogo sea un sistema computacional actuando como una segunda opinión, así los resultados obtenidos por el sistema pueden ser confirmados o rechazados por el especialista. En conclusión, la mamografía es la técnica ideal para el diagnóstico temprano del cáncer de mama, porque no es invasiva y porque puede representar adecuadamente las calcificaciones. Una técnica es ideal porque permite detectar el cáncer antes que infiltre en el tejido circundante.

Las calcificaciones son pequeñas acumulaciones de calcio de 0.1 mm a 2 mm de ancho, y son indicadores favorables de la presencia de cáncer de mama. Frecuentemente son usadas en el diagnóstico del carcinoma intraductal o carcinoma ductal in situ y tienen probada capacidad para detectar estadios tempranos de la enfermedad.

Entre el 30% y el 50% del cáncer de mama en el mundo es diagnosticado debido a la detección de calcificaciones. Los agrupamientos de calcificaciones son grupos de 3 o más calcificaciones individuales que pueden aparecer en áreas menores a 1 cm<sup>2</sup> (**ver figura 11**). Su presencia incrementa la probabilidad de que la lesión sea maligna.

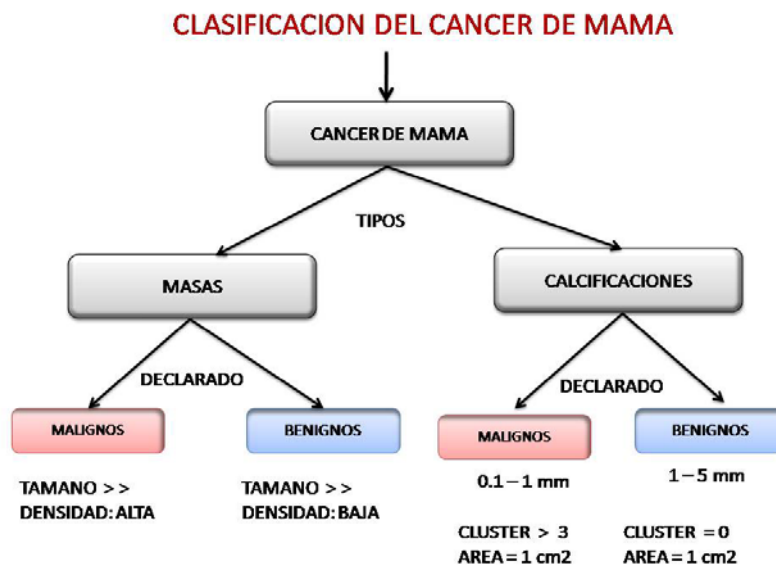


Figura A1-12. Esquema de clasificación del cáncer de mama según: tamaños, áreas, densidad (formulación global)

Finalmente el reto para un radiólogo mediante la lectura de la imagen es localizar y clasificar dentro de los tipos de cáncer dos tipos de vectores: masas y calcificaciones, en la **figura 12** mostramos como los índices: tamaño, área, agrupamiento y densidad son factores importantes en la valoración del diagnóstico para el tratamiento posterior. En el diagnóstico se debe clasificar dentro de los dos tipos si el cáncer es benigno o maligno, entendiendo por benigno todo carcinoma que no ofrece posibilidad de replicación y metástasis, el cual puede ser extraído ofreciendo un índice alto de mejoría en el paciente. Los carcinomas malignos son aquellos que se infiltran en los tejidos vecinos y se replican en áreas circundantes ocasionando crecimientos, contaminaciones y metástasis.

Este trabajo de tesis se profundizara en la búsqueda y detección de los carcinomas pequeños o micro carcinomas (calcificaciones) debido a que los tumores o masas son de fácil vista para el especialista en el proceso de lectura.

Cuando mencionamos el porqué de la baja especificidad en el diagnóstico surge el concepto de Falsos Positivos (FP) y falsos Negativos (FN).

- Un falso negativo **FN** es cuando existiendo la enfermedad pero en el proceso del diagnóstico no se detecta esta, ocasionando que el paciente no sea atendido y permitiendo que la enfermedad continúe su proceso de crecimiento y maduración, en el caso del cáncer de mama en especial una calcificación su índice de crecimiento es muy alto y al tiempo en el cual se vuelva a efectuar de nuevo el examen (siguiente mamografía – 2 años después) sea tarde.

- Un falso Positivo **FP** es cuando no existiendo un carcinoma pero en el proceso de diagnóstico se da como detectada. Esto ocasiona tratamientos innecesarios como lo es la Biopsia, exámenes clínicos y de laboratorio, así como ecografías y demás.

Resolver este problema entre los FN y FP es incrementa el nivel de especificidad del diagnóstico y hacerlo más preciso, como comentamos los costos de una doble lectura, y que en muchos casos no es posible por la ausencia de personal especializado, es en donde nace la buena idea del desarrollo de un sistema experto, que ayude en la segunda lectura incrementando la productividad del radiólogo.

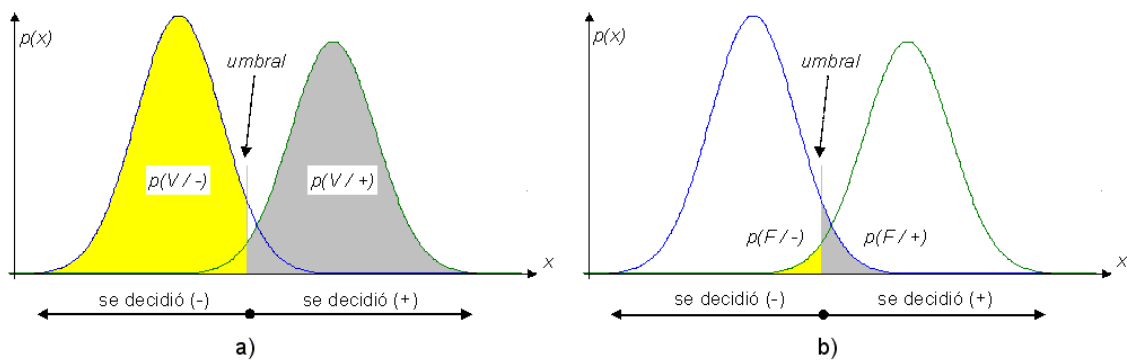


Figura A1-13. [19] Relación entre verdaderos positivos (V+), verdaderos negativos (V-), falsos positivos (F+) y falsos negativos (F-).

El éxito del diseño de un sistema experto que haga una localización y clasificación automática esta en lograr que los umbrales estén balanceados minimizando los FP y FN, esta será una tarea difícil de lograr dentro de la sección de segmentación, localización y clasificación del algoritmo.

## A-1.8 IMAGENOLÓGÍA Y MAMOGRAFÍA

La imagenología es la técnica que se especializa en el manejo de las imágenes médicas con miras a diagnosticar y ayudar en los procedimientos de detección temprana, tratamientos quirúrgicos y post - quirúrgicos. La imagenología está apoyada en la tecnología de sistemas de información, redes de computadores y software especializado integrado con equipos biomédicos de imágenes como: el **Mamografo** (equipo especializado de rayos X), Tomografía axial computarizada **TAC**, resonancia magnética **MRI**, ecografía **ECO**, sistemas de medicina nuclear como lo son: **Gamacamaras**, **PETs** y SPET.



Figura A1-14. [13] interconexión de sistemas de información con equipos de imágenes conformando la tecnología de imagenología ( *Prisma Technologies Inc.* )

Los sistemas de imagenología están conformados por los siguientes sistemas:

- **Sistema de almacenamiento de imágenes:** **PACS** (Picture Archive Communications Systems) el cual lo conforma una plataforma computacional de misión crítica en el cual se almacenan todas las imágenes generadas por los equipos biomédicos invasivos y no invasivos que generan la imagen en los pacientes. El formato que se ha ido estandarizando para la digitalización de la imagen es el **DICOM**
- **Equipos Biomédicos:** son los equipos de rayos X, especializados en la toma de imágenes, de igual forma sistemas basados en resonancia magnética, medicina nuclear, sistemas de ultrasonido y ecografía. La física y matemática han aportado herramientas que los electrónicos han aprovechado para desarrollar estos equipos que hoy por hoy hacen de la medicina una ciencia tecnológica de punta. Compañías líderes en la fabricación de estos equipos son Siemens, General Electric, Phillips, ST, Fujitsu, Kodak, Agfa entre muchos.

- **Redes de telecomunicaciones:** la red de área local LAN es el ente de conectividad que permite la convivencia y la intercomunicación entre los generadores de imágenes el sistema PACS y posteriormente las estaciones de lectura y consulta externa. El protocolo de comunicación es el estándar IP el cual permite conexiones a nivel local, metropolitano y global.
- **Estaciones de lectura:** las estaciones de lectura es la herramienta principal del médico especialista en imágenes (radiólogo), es un computador adaptado con un monitor de alta resolución en la cual se interconecta con la plataforma PACS y extrae las imágenes. En el tema de radiología se ha ido desarrollando de igual forma la plataforma de manejo de seguimiento y control de los radiólogos con el **RIS** (Radiology Information Systems). Las estaciones de lectura en conjunto con software especializado en procesamiento y despliegue de imágenes hacen que los médicos cada día se capaciten y adquieren mayor conocimiento de sistemas, redes y tecnología.
- **Sistemas de apoyo a la lectura:** el radiólogo dentro del trabajo de la lectura interpretando la imagen graba su resultado en un sistema de grabación; posteriormente las transcriptoras transcriben el dictamen a formato digital de texto para poder ser impreso dentro del resultado que debe ser entregado al paciente. Hoy por hoy existen sistemas de reconocimiento de voz a texto permitiendo de una forma rápida la respuesta del resultado.
- **Mamografo:** el mamaografo es un equipo de radios X diseñado especialmente para la toma de imagen del seno, el equipo tiene una incidencia de magnitud de rayos X de baja intensidad. Los mamaografos por muchos años han sido equipos análogos en donde la imagen es impresa en el film (detector) que al atravesar el haz de rayos X entre el tubo generador y la mama queda este grabado en esta placa (detector). La placa debe ser digitalizada mediante un CR (computer radiology) para ser pasada a formato DICOM y ser almacenada automáticamente al PACS. Actualmente existen equipos modernos los cuales incluyen el CR de tal forma que en forma automática digitalizan la imagen a estos sistemas se les denomina mamaografos digitales.

El procedimiento de la mamografía por efecto de achatar el seno para poder adecuarse al área de la placa produce una distorsión visual, de igual forma cuando el proceso se hace mediante mamaografos análogos la digitalización incorpora ruido el cual para la detección y análisis de los carcinomas pequeños denominados calcificaciones es difícil de ser vistos.

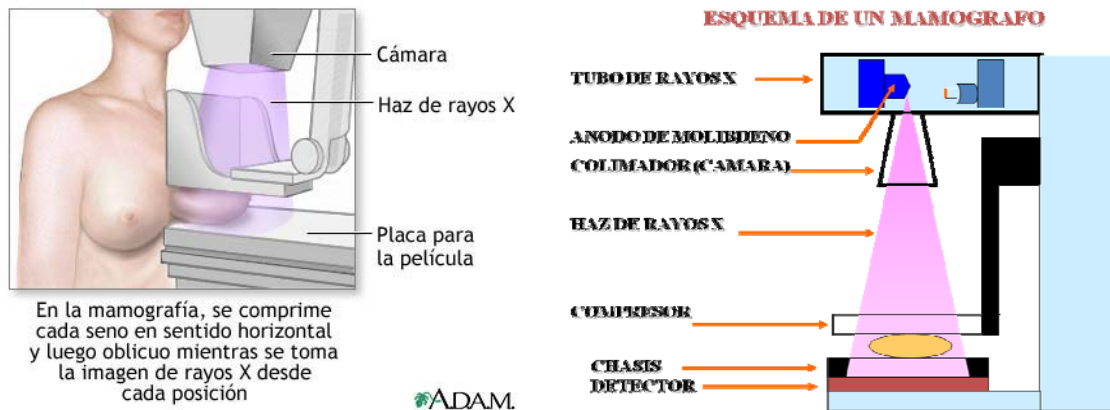


Figura A1-15. Elementos de un equipo mamógrafo

La imagen obtenida en la mamografía, se obtiene usando detectores **pantalla-película**, que graban los fotones de radiación que pasan a través de la mama. Estos detectores son los que determinan la resolución espacial. Algunas pantallas, con más actividad que otras (resolución), tendrán más material fluorescente (intensidad) que al interactuar con fotones de radiación permitirán una exposición más corta, disminuyendo la posibilidad de borrosidad por movimiento y además, reducirá la dosis glandular.

No obstante, la calidad se verá afectada por el aumento de la borrosidad resultante de la producción e interacción de una mayor cantidad de luz. Otro problema en este sistema de detectores, es que la película no tiene la capacidad de respuesta inmediata a los fotones que la inciden. Por eso, vamos a obtener en áreas de mayor densidad, mayor absorción de radiación y por más que en la película se obtenga alta definición, no irá acompañada de un adecuado contraste. Esto determinará que algunas lesiones sean sub diagnósticas.

Por otro lado los nuevos mamografos digitales ofrecen una técnica que mide directamente los fotones que pasan por la mama. Lo importante es que el equipo tiene la capacidad de leer los primeros fotones, lo cual no es obtenido por la técnica anterior. Además es capaz de leer un gran flujo de ellos, lo que nos permite más información sobre una enorme cantidad de puntos de la mama. O sea, obtenemos un mapeo más exacto.

Hay dos tipos de mamógrafos digitales: los **directos e indirectos**. Los primeros, tienen detectores de radiación, que convierten directamente la información, en carga eléctrica. Habitualmente utilizan SELENIO como foto receptores. En los indirectos, la radiación convertida en luz, es recibida por FOTODIODOS, y transformada en carga eléctrica, estos últimos son los que se están difundiendo por su bajo costo y su implementación con pocos elementos de infraestructura mecánica y electrónica. En el sistema pantalla película, la imagen una vez obtenida, no puede modificarse



Figura A1-16. Imágenes de equipos mamografos, estación de lectura. (General Electric)

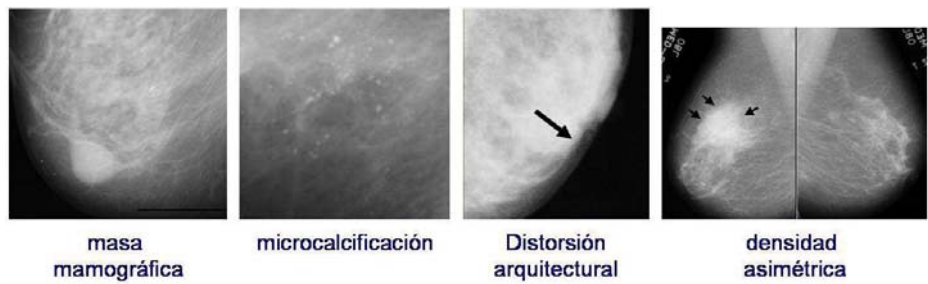


Figura A1-17 [19]: Hallazgos en la mamografía.

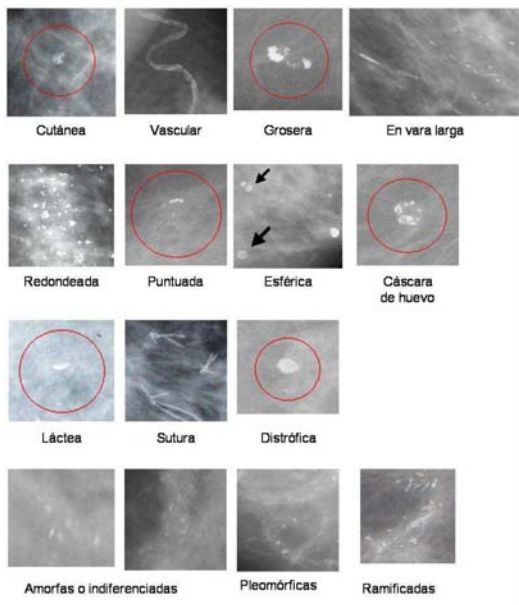


Figura A1-18 [19]:

Diferentes tipos de carcinomas de pequeño tamaño conformando calcificaciones de aspectos morfológicos de diferentes formas: redondos, ovalados, alargados, espiculados, estrellados.

De acuerdo al tamaño y forma es posible identificar si es maligno o benigno y ser clasificados mediante un algoritmo de clasificación automático aplicando una serie de reglas basadas en las especificaciones del protocolo BI-RADS que veremos en la sección 2.10



## A-1.9 TRATAMIENTO

La enfermedad del cáncer es muy compleja y requiere de un proceso largo y costoso, en el caso de cáncer de mama dividimos en tres etapas a la enfermedad: pre-cáncer, cáncer, post-cáncer. El pre cáncer es la etapa en la cual la mujer debe hacerle los exámenes para detección temprana ya vistos en la sección 2.7 y debe empezar desde los 30 años. Una vez se sospecha de alguna anomalía se recurre al especialista para el examen clínico y proceder a una mamografía, si se detecta carcinoma se dispara la etapa de cáncer. La complejidad del desarrollo y tratamiento de la enfermedad se basan en cuatro tipos de tratamiento y combinaciones de estas:

- Cirugía: es el procedimiento de operación quirúrgica y extracción del tumor.
- Radioterapia: es el procedimiento radioactivo para quemar las células carcinomas mediante varias sesiones de radioterapia: 5 días por semana durante 6 a 7 semanas, en dosis de 1.8 a 2.0 Gy/día para un total de 45 a 50 Gy en tratamiento. Uno de los inconvenientes de la radioterapia es la radiación vecina ocasionando muerte de células buenas.
- Quimioterapia: es el procedimiento de introducir un químico (reactivo) en la sangre, se hacen unos 3-5 procesos. Este procedimiento es el más utilizado cuando hay metástasis.
- Terapia Hormonal.
- Combinaciones de las anteriores.

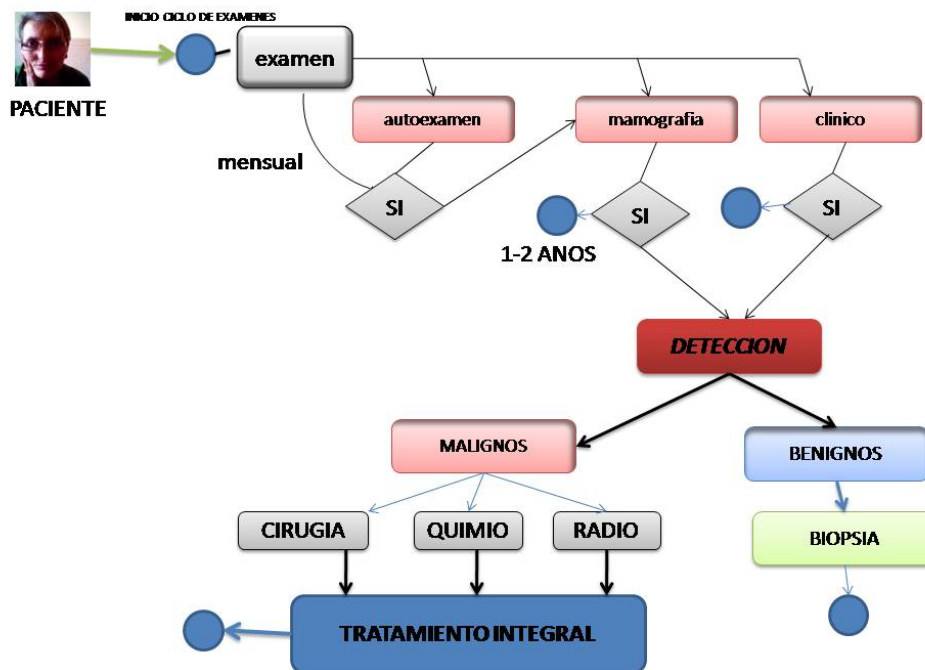


Figura A1-19. Flujo grama del proceso de tratamiento de la enfermedad de cáncer de mama

La mamografía juega un papel muy importante en la detección temprana, pero también en la etapa post cáncer, pues los pacientes tienen dos (2) senos y la enfermedad en uno no necesariamente se replica al mismo tiempo en el otro seno, se puede decir que cada seno izquierdo y derecho tiene probabilidades de afección cercanas pero no iguales. El efecto más traumático del tratamiento es la mastectomía o procedimiento de quitar parte del seno o en su totalidad con miras a minimizar las probabilidades de hacer metástasis en otros órganos.

### **A-1.10 MODELO BI-RADS**

El Colegio Americano de Radiología desarrollo en 1992 el BI-RADS (The Breast Imaging Reporting and Database System [0]), que es una herramienta para clasificar los hallazgos en las mamografías. El BIRADS frecuentemente es utilizado por sistemas automatizados de detección y diagnóstico de cáncer de mama. Una de sus características es que intenta cumplir el rol de "una herramienta de aseguramiento de calidad diseñada para estandarizar el reporte monográfico, reduciendo la confusión en la interpretación de la imagen de la mama y facilitando el control de los resultados" [20]. El BIRADS agrupa las imágenes en **seis** (6) categorías, del 0 al 5 (ver Tabla 1) En cada categoría se sugiere la conducta más adecuada a seguir y por otro lado se indica el valor predictivo positivo para la malignidad (VPP) de las mismas en las diferentes categorías de manejo. [22] El diagnóstico monográfico del cáncer de mama se basa en la detección de alteraciones en la arquitectura glandular que quedan resumidas, en función del grado de sospecha, en un sistema consensuado de nomenclatura denominado BI-RADS (Breast Imaging Reporting Data Systems). Este sistema establece, por un lado, un lenguaje común radiológico con el que se agrupan las distintas alteraciones mamográficas observadas y por otro lado permite correlacionar estas alteraciones con el grado de sospecha en relación con el cáncer de mama. De esta forma, las alteraciones incluidas dentro de los grupos 4 y 5 tienen las mayores posibilidades de correlacionarse, en el estudio histopatológico final, con un proceso maligno. Estas probabilidades oscilan entre un 50-94% en el caso del BIRADS 4 y cercano al 100% en el caso de BIRADS 5 [23]. Siendo que el BIRADS es un sistema de control de calidad, su uso diario implica una evaluación en categorías numéricas de una mamografía, asignado por el médico radiólogo e imagenólogo después de interpretar la mamografía. Ello permite que se presente un reporte radiográfico uniforme y conciso y que pueda ser entendido por múltiples médicos o centros hospitalarios. Consta de 7 clases diferentes de acuerdo a su estatificación, la categoría 6 fue añadida en la 4<sup>ta</sup> por la ACR.

**Clase 0:** Radiografía insuficiente, necesita una evaluación adicional con otro estudio, no es posible determinar alguna patología.

**Clase I:** Mamografía negativa a malignidad, sin ganglios o calcificaciones. 0% de posibilidades de cáncer.

**Clase II:** Mamografía negativa a malignidad, pero con hallazgos benignos (ganglio intramamarios, calcificaciones benignas, etc.). 0% de posibilidades de cáncer.

**Clase III:** Resultado con probable benignidad, pero que requiere control a 6 meses. Puede presentar nódulos circunscritos o algún grupo pequeño de calcificaciones puntiformes y redondeadas. 2.24% de posibilidades de cáncer.

**Clase IV:** Resultado dudoso de malignidad. Requiere una confirmación histopatología. Consta de 3 grados de acuerdo con su porcentaje de malignidad que van del 3 al 94%

1. Baja sospecha de malignidad. 3 a 49%
2. Sospecha media de malignidad. 50 a 89%
3. Sospecha intermedia de malignidad. 90 a 94%

**Clase V:** Alta sospecha de malignidad. Requiere biopsia para confirmar diagnóstico. > De 95% de posibilidades de malignidad.

**Clase VI:** Malignidad comprobada mediante biopsia.

	<b>CLASIFICACIÓN</b>	<b>DESCRIPCIÓN</b>	<b>CONDUCTA</b>	<b>VPP</b>
0		Se necesita más información para entregar un reporte final de la mamografía.	Complementar con mas exámenes	13%
1	NEGATIVO	Mamografía normal: sin carcinomas	Control periódico	
2	BENIGNO	Añormalidades menores. Sin sospecha de cáncer	Control periódico	0%
3	Probablemente BENIGNO	Se detectan masas solidas y regulares	Nuevos controles en 6 meses	0.5 – 2%
4	Probablemente MALIGNO	La apariencia de la lesión no es característica de malignidad, pero la probabilidad de malignidad es suficientemente alta. Las imágenes representativas son: las masas de contornos no definidos, polilobuladas y a veces mal visualizadas, de estructura heterogénea y con calcificaciones heterogéneas (amorfos o granulares)	Se debe considerar biopsia.	15 – 30%
5	MALIGNO	Alta probabilidad de malignidad. Las imágenes representativas son: las masas de contornos irregulares y espiculadas y las calcificaciones irregulares con disposición lineal, ductal o arboriforme.	Se requiere biopsia.	97%

Tabla A1-1. [20] Clasificación de BI-RADS Vs. Valor predictivo positivo VPP

## CLASIFICACION SEGUN OBJETIVO DEL DIAGNOSTICO

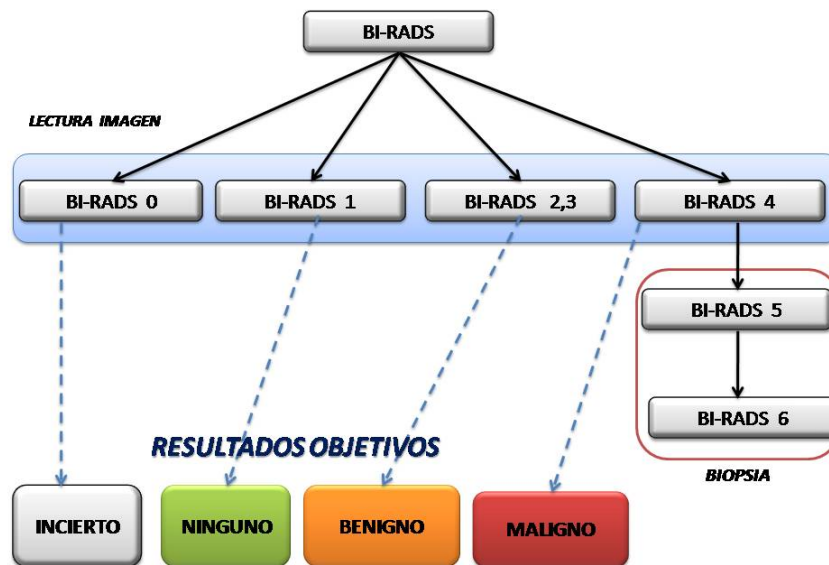


Figura A1-20. Resultados del BIRADS según resultados objetivos: diagnóstico incierto, no existe ningún carcinoma, diagnóstico benigno, diagnóstico maligno.

El objetivo principal de la estandarización BIRADS es identificar cuatro posibles opciones:

- ***Diagnóstico incierto:*** no se logra localizar ninguna marca o carcinoma que permita tener una certeza diagnóstica, esto sucede muchas veces por ruido en la imagen y no haber quedado bien.
- ***Ningún carcinoma:*** imagen limpia, el paciente no tiene ningún carcinoma.
- ***Diagnóstico benigno:*** identificado señales claras que permiten identificar como un carcinoma benigno.
- ***Diagnóstico maligno:*** identificación de carcinoma que requiere de tratamiento mediante una biopsia inicial para identificar la patología y continuidad en el tratamiento visto en la sección 2.9

## **A-1.11 RESUMEN CAPITULO**

En este capítulo se hace una revisión general de la enfermedad del cáncer profundizando en el cáncer de mama: su origen, historia, patología, epidemiología, diagnóstico, imagenología, tratamiento y como se clasifica según la ACR mediante el estándar BIRADS. Como gran importancia queremos resaltar que en el diagnóstico temprano mediante el análisis de la imagen se logra la real prevención y diagnóstico a un buen tiempo, este trabajo debe identificar cuatro (4) marcas que identifican el nivel diagnóstico de la enfermedad: incierta (gris), normal (verde), benigno (naranja) maligno (rojo).

## A-1.12 REFERENCIAS

- [1] <http://www.genomasur.com/lecturas/Guia12a.htm>
- [2] cáncer mamario, Hernando Abaunza Orjuela Md, federación latinoamericana de cirugía FELAC, pag. 1, 1987.
- [3] Instituto nacional de cancerología E.S.E empresa de carácter social inscrita al ministerio de protección social de la república de Colombia: [www.cancer.gov.co](http://www.cancer.gov.co)
- [4] entidad de carácter humanitario no gubernamental adscrita a las federaciones Unión internacional contra el cáncer y sociedad americana contra el cáncer: <http://www.ligacancercolombia.org/>
- [5] Agustín rey y Elena redondo, Biocancer reserach journal, ISBN 1697-6452 Biocancer 1, 2004.
- [6] diagramas de la national cáncer institute en la presentación: <http://www.wikilearning.com/monografia/cancer/10766>
- [7]
- [8] Fernando Andrés Angarita, Sergio Andrés Acuna, Cáncer de seno: de la epidemiología al tratamiento, Univ. Medica, 49 (3):344-372 julio-septiembre del 2008.
- [9] Key Tj., Verkasalo PK, Banks E. Epidemiology of breast cancer. Lancet. 2005; 2:133-40
- [10] Boyle P. breast cancer control: signs of progress, but more work required. Breast 2007;14:429-38
- [11] Yenny Marcela Leuro Umana, Paola Andrea Mayorga Rojas, Tesis: Lineamientos para la promoción y prevención de cáncer de mama a partir de revisión bibliográfica entre 1997-2007 IPS Javesalud-Bogota. Pontificia universidad javeriana, 2008.
- [12] Gerardo Ramirez, informe final trabajo materia reconocimiento de patrones, Pontificia Universidad javeriana, Maestría Electrónica \* autor de la tesis, 2010.
- [13] soluciones de conectividad e imágenes, Prisma Technologies Inc. Ltda., [www.prismtechnologies.net](http://www.prismtechnologies.net), 2010.
- [14] S de Paredes E. radiographic breast anatomy radiology signsof breast cancer, Syllabus: a categorical course in physics – technical aspects of breast imaging. M Yaffe ed. Oak brook IL. RSNA Publications, 1993.

- [15] I. Valdivia B. Resultados de falsos positivos en programas de detección de cáncer mamario: resultados de 10 años. Boletín de la sociedad chilena de climaterio, 3 (1):5-9, 1998.
- [16] M. L. Corner, S. Liu, and E. J. Delp. Statistical segmentation of mammograms. In 3<sup>Rd</sup> International workshop on digital mammography, june 9-12, pages 475-478, Chicago. 1996.
- [17] L. P. Cordella, F. Tortorela, and M. Vento. Combining expert with different features for classifying clustered micro calcifications in mammograms. In International Conference o pattern recognitions (ICPR 2000)-vlolume 4 (septiembre 03-08) page 4324, Barcelona Spain 2000.
- [18] L. de Fontoura and R. marcondes. Shape Analysis and Classification. Theory an Practice. CRC press, 2001.
- [19] Samuel Oporto Díaz, detección automática de agrupaciones de microcalcificaciones en mamaografías digitalizadas, tesis, Instituto tecnológico de monterrey, Diciembre 2004.
- [20] American College of radiology. Breast imaging reposting and data system (BI-RADS). 3<sup>rd</sup>. ed. Reston Va.: American Collage of radiology, 1998.
- [21] Liliana Torregrosa Almonacid, Mauricio Tawil Moreno. BASES DIAGNÓSTICAS Y TERAPÉUTICAS DEL CÁNCER MAMARIO: GUÍA PARA EL MÉDICO GENERAL, Pontificia Universidad Javeriana.
- [22] <http://wikisenology.wikispaces.com/C%C3%A1ncer+de+mama.+Diagn%C3%B3stico+y+estadificaci%C3%B3n>.
- [23] Lehman C, Holt S, Peacock S et al: Use of the American College of Radiology BIRADS Gidelines by Community radiologists: Concordance of assessments and recommendations assigned to screening mammograms. AJR 2002; 179: 15-20

## ANEXO 2

### A-2.1 TECNICAS DE SEGMENTACIÓN

La segmentación es la parte del análisis de imágenes la cual consiste en subdividir una imagen en sus partes constituyentes u objetos también denominadas regiones de interés (ROI). El nivel al que se lleva a cabo esta subdivisión depende del problema a resolver; esto es quiere decir: la segmentación deberá detenerse cuando los objetos de interés de una aplicación hayan sido ***aislados*** o ***delimitados***. En general, la segmentación es una de las tareas más difíciles del procesamiento de imágenes. Esta etapa del proceso determina el eventual éxito o fracaso del análisis [4].

[5] Los métodos para llevar a cabo las segmentaciones varían ampliamente dependiendo de la aplicación específica, tipo de la imagen, y otros factores. Por ejemplo, la segmentación del tejido del cerebro tiene diferentes requerimientos que la segmentación de un hígado. Otros factores como el ruido, volúmenes parciales y movimientos en la imagen, también pueden tener consecuencias significativas en el desempeño de los algoritmos de segmentación. Más aún, cada tipo de imagen tiene un conjunto de características propio con el cual luchar. Actualmente no existe un método de segmentación que alcance resultados aceptables para todo tipo de imagen médica. No existen métodos que sean generales y que puedan ser aplicados a cualquier variedad de datos. De cualquier forma, los métodos que son especializados para aplicaciones particulares pueden obtener mejores resultados tomando en cuenta conocimiento a priori. Por lo tanto, la selección de un método apropiado para un problema de segmentación puede ser muy difícil.

[5] Clásicamente, la segmentación de imágenes se define como la partición de una imagen en regiones constituyentes no solapadas, las cuales son homogéneas con respecto a alguna característica como una intensidad o una textura. Si el dominio de la imagen esta dado por  $I$ , entonces el problema de segmentación consiste en determinar el conjunto:  $S_k \subset I$  cuya unión es la imagen  $I$  completa. Por lo tanto, el conjunto que conforma la segmentación debe satisfacer:

$$S = \bigcup_{k=1}^K S_k \quad (5.1)$$

Donde  $S_k \cap S_j = \emptyset$  para  $k \neq j$ , y cada  $S_k$  está conectado. Idealmente, un método de segmentación encuentra aquellos conjuntos que corresponden a distintas estructuras o regiones anatómicas de interés en la imagen. Al eliminar la restricción



de que las regiones estén conectadas, determinar los conjuntos  $S_k$  es llamado **clasificación de píxel** y a los conjuntos se les llama **clases**. La clasificación de píxeles frecuentemente es un objetivo deseable en el tratamiento de imágenes médicas, particularmente cuando se necesita clasificar regiones desconectadas que pertenecen al mismo tejido (calcificaciones). La del número de clases  $K$  en la clasificación de píxeles puede ser un problema complejo, por lo que generalmente se asume conocida, basado en conocimientos previos de la anatomía que se está tomando en consideración.

El **etiquetado** es el proceso de asignar una designación significativa a cada región clase y puede ser llevada a cabo separadamente de la segmentación. Este proceso mapea el índice numérico del conjunto  $S_k$ , a una designación anatómica. En imágenes médicas, frecuentemente las etiquetas son obvias y pueden ser determinadas mediante inspección del Técnico o Radiólogo. El etiquetado automático por computador es deseable cuando las Etiquetas no son obvias o el sistema realiza procesamiento automático de imágenes. Una situación típica en la que se utiliza el etiquetado ocurre en la mamografía digital cuando se segmenta la imagen en distintas regiones y las regiones se etiquetan como sanas o tumorosas.

La segmentación se desarrolla por dos grandes tópicos: detección de bordes (objetos) y realces (fondos) en una imagen [1,2,3].

### A-2.1.1 UMBRALIZACIÓN

La umbralización (*thresholding*) es un método que busca segmentar imágenes escalares creando una partición binaria de las intensidades de las imágenes. Una umbralización trata de determinar un valor de intensidad, llamado umbral (*threshold*), que separa las clases deseadas. La segmentación se logra agrupando todos los píxeles con mayor intensidad al umbral en una clase, y todos los otros píxeles en otra clase. La Figura A2-1 muestra un histograma de una imagen que aparentemente posee tres clases, lo cual nos deja dos posibles valores de umbral, en los valles del histograma. La determinación de más de un valor umbral es un proceso llamado multiumbralización (*multithresholding*).

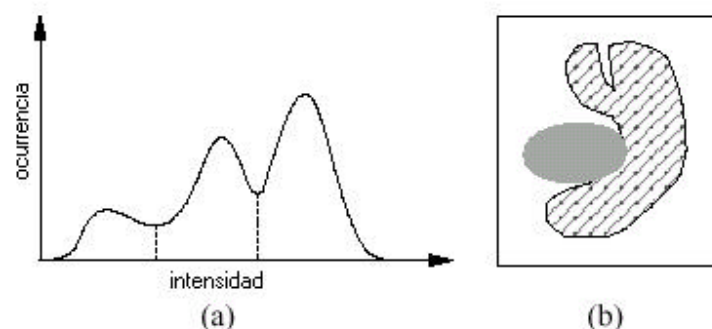


Figura A2-1 [5]. Umbralización: (a) histograma de intensidades de grises en la imagen mostrando los posibles umbrales (b) imagen en escala de grises

La umbralización es una técnica efectiva para obtener la segmentación de imágenes donde estructuras diferentes tienen intensidades contrastantes u otras características diferenciables. La partición usualmente es generada interactivamente, pero también existen métodos automáticos. Para imágenes escalares, los métodos interactivos pueden estar basados en la apreciación visual del usuario ya que la operación de umbralización es implementable en tiempo real. Generalmente, la umbralización es el paso inicial de una secuencia de operaciones de procesamiento de imágenes. Su principal limitación es que en su forma más simple solo se generan dos clases y por lo tanto no se puede aplicar a imágenes multicanal.

[6] Al aplicar un umbral,  $T$ , la imagen en escala de grises,  $f(x,y)$ , quedará binarizada; etiquetando con '1' los píxeles correspondientes al objeto y con '0' aquellos que son del fondo. Por ejemplo, si los objetos son claros respecto del fondo, se aplicará:

$$g(x,y) = \begin{cases} 1 & f(x,y) > T \\ 0 & f(x,y) \leq T \end{cases} \quad (5.1)$$

En el caso de que los objetos sean oscuros respecto del fondo, la asignación sería a la Inversa:

$$g(x,y) = \begin{cases} 1 & f(x,y) < T \\ 0 & f(x,y) \geq T \end{cases} \quad (5.2)$$

El umbral puede depender de  $f(x,y)$ , de alguna propiedad local del píxel,  $p(x,y)$ , y hasta de su propia posición:

$$T = T(f(x,y), p(x,y), x, y) \quad (5.3)$$

Si el umbral sólo depende de  $f(x,y)$  se dice que es un umbral global; en el caso de que además dependa de  $p(x,y)$ , por ejemplo, el valor medio de los píxeles vecinos, el umbral es denominado local; y si depende también de la posición  $(x,y)$  del píxel, se denominará dinámico. La mayoría de las técnicas de umbralización se basan en estadísticas sobre el histograma unidimensional de una imagen.

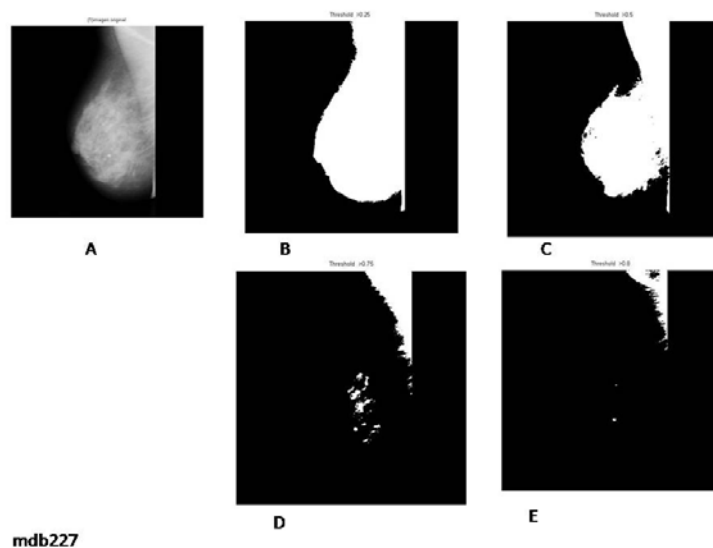


Figura A2-2. Aplicación de la técnica de umbral sencilla: A) imagen original, B) umbral >0.25. C) umbral >0.5 D) umbral >0.75 E) umbral >0.8

En la figura A2-2, observamos el procesamiento de la técnica de umbralización con diferentes umbrales o valores de  $T$ : 0.25, 0.5, 0.75, 0.8 en estas vemos como la segmentación va acotando los brillantes o posibles señales de interés como ROI, en especial con umbral 0.8 se resalta el punto de calcificación de la imagen.

### A-2.1.2 DETECCIÓN DE BORDES

[7] Las técnicas de detección de bordes (contornos) son útiles en diferentes contextos, en particular la detección de bordes es una de las etapas del proceso de segmentación cuyo objeto es particionar la imagen en regiones asociadas a los diferentes elementos que componen la escena, y que puede ser utilizada posteriormente para el análisis automático de los mismos mediante algoritmos de reconocimiento de formas [3].

Un borde en una imagen, es un *límite* o contorno en el cual ocurren cambios significativos en algún parámetro físico de la imagen, tal como: la reflectancia superficial, la iluminación o la distancia de la superficie visible al observador. Los cambios en los parámetros físicos de la imagen se manifiestan de diversas formas, incluyendo cambios en intensidad, color y textura; En la figura A2-3, se muestran los diagramas esquemáticos de bordes unidimensionales y bidimensionales, modelados usualmente como un incremento en rampa, en los niveles de gris de la imagen, en este caso de un nivel de gris bajo a un nivel de gris alto, siendo válida la definición para el

caso contrario. En el caso bidimensional, cada punto  $(x,y)$  define la posición del pixel y la coordenada z define la amplitud del nivel de gris.

Para el caso unidimensional, el borde se caracteriza por la altura, la pendiente y la coordenada  $x_0$  del punto medio de la pendiente, entonces se puede decir que existe un borde si tanto el ángulo de la pendiente como la altura superan un cierto valor umbral. Para el caso bidimensional, también es importante la orientación con respecto al eje x, en cualquier caso lo deseable es que el detector de contorno produzca un pixel indicador del contorno, ubicado en el punto medio de la pendiente.

En la figura A2-3, se muestra un enfoque clásico para la detección de bordes, según el cual la imagen, se somete a una acentuación de contornos, seguida por un detector de borde por umbral. Si denotamos como  $x(i,j)$  la imagen de entrada,  $G(i,j)$  la imagen luego de la acentuación de bordes,  $Ub$  el umbral para bordes de bajo a alto y  $Ua$  el umbral para bordes de alto hacia bajo, entonces podemos decir que tenemos:

Un borde en sentido positivo si:

$$G(i, j) \geq Ub \quad (5.4)$$

Un borde en sentido negativo si:

$$G(i, j) < Ua \quad (5.5)$$

La selección del valor umbral, es uno de los aspectos importantes en detección de bordes. Un nivel de umbral muy elevado, no permitirá la detección de elementos estructurales de la imagen si estos no tienen suficiente amplitud, del mismo modo, un umbral de muy poca amplitud causará que el ruido se detecte falsamente como bordes en la imagen.

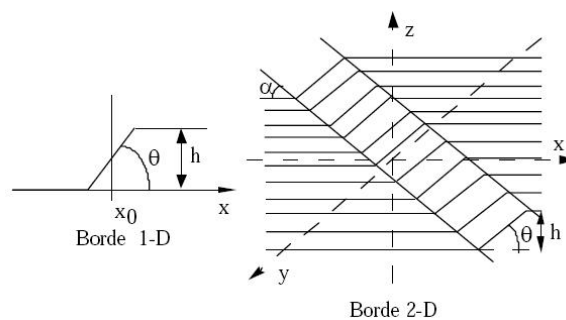


Figura A2-3. [7] modelo para detección de borde unidimensional y Bidimensional.

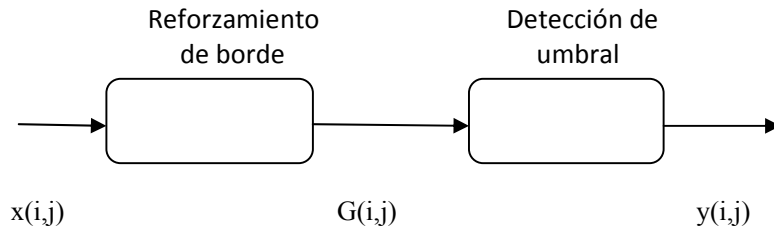


figura A2-4. [7] ejemplo de una implementación de detección de bordes.

### Detección de bordes basados en técnicas de gradientes

Esta técnica se basa en el gradiente o la derivada de la señal. En la figura A2-5 se analiza la aplicación de esta técnica para detectar un borde unidimensional, en una función  $f(x)$  (figura A2-5a). En tal caso el punto  $x_0$  corresponde a la ubicación del borde y una manera de determinar tal valor, es mediante el uso de la primera derivada  $f'(x)$  tal como se muestra en la figura A2-5b o de la segunda derivada tal como se muestra en la figura A2-5c, en tales figuras, el valor de  $x_0$  puede determinarse buscando los extremos (máximo o mínimo) en la primera derivada ó la posición del cruce por cero cuando la segunda derivada cambia de signo.

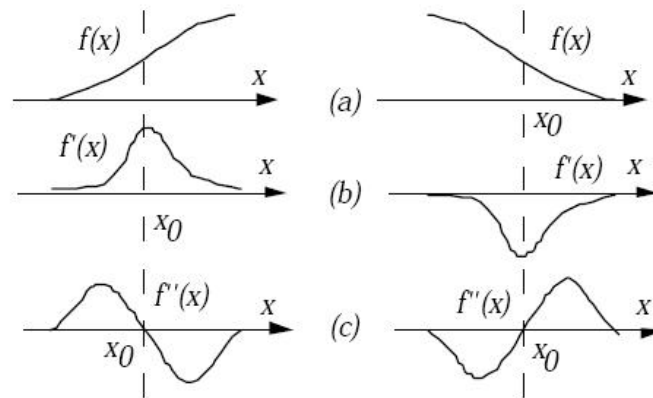


Figura A2-5. Diagrama del gradiente de una señal  $f(x)$ .

Para una función el gradiente de  $f$  de coordenadas  $(x,y)$  se define como el vector:

$$\nabla f' = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} \quad (5.6)$$

El modulo del vector esta dado por:

$$\nabla f = \text{mag} (\nabla f') = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} \quad (5.7)$$

Esta ecuación se puede aproximar a:

(5.8):

$$\nabla f \approx |(z7 + z8 + z9) - (z1 + z2 + z3)| + |(z3 + z6 + z9) - (z1 + z4 + z7)|$$

La diferencia entre la tercera y la primera fila de la región 3X3 da una aproximación de la derivada en la dirección  $x$ , y la diferencia entre la tercera y la primera columna da una aproximación de la derivada en la dirección  $y$ . la matriz como la de la figura A2-6, se conocen con el nombre de mascarar en la jerga de procesamiento de imágenes:

Z1	Z2	Z3
Z4	Z5	Z6
Z7	Z8	Z9

Figura A2-6. Matriz z, en la que cada elemento tiene un nivel diferente de gris.

De acuerdo al los valores que se dan en cada índice de  $Z_i$  se crean unos operadores que son muy utilizados en procesamiento de imágenes medicas y en general. En las siguientes figuras A2-7 veremos algunos de más uso y que en este trabajo se han probado buscando el mejor comportamiento dentro del proceso de análisis y segmentación de la imagen de mamografía.

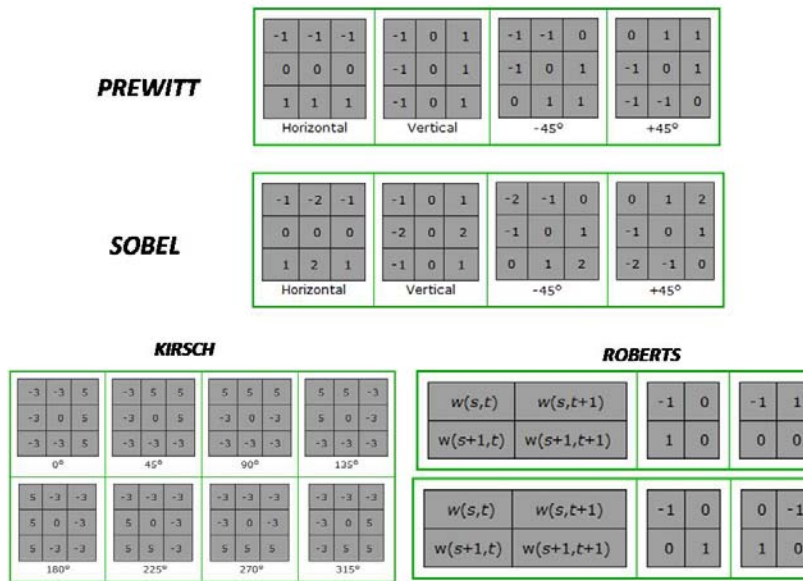


Figura A2-7. Tabla de operadores según valores índices Zi bajo las técnicas de gradientes. Prewitt, Sobel, Kirsch, Roberts.

Cuando aplicamos la segunda derivada a la formula (5.6) se define el operador Laplaciano. Este operador funciona mejor cuando hay cambios abruptos en los niveles de grises, como por ejemplo una función paso, o una esquina.

$$\nabla^2 f(x,y) = \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2} \quad (5.8)$$

Y la función que define la máscara es:

$$\nabla^2 f(x,y) = f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1) - 4f(x,y) \quad (5.9)$$

Dándonos las siguientes mascarar:

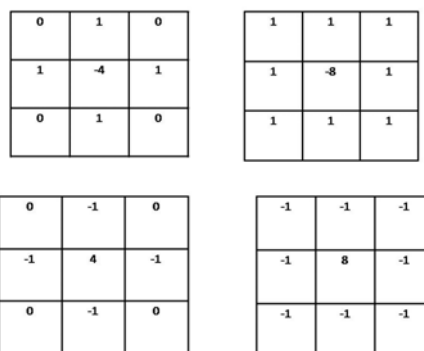


Figura A2-8. Tablas operador laplaciano

Existen otros tipos de operadores del gradiente como lo son: detector de cruce por cero, detector de borde Canny (Jhon canny quien trabajo este método en 1986) , tal vez el más poderoso para la detección de bordes, en resumen podemos resaltar:

- La primera derivada devuelve bordes más gruesos
- La segunda derivada discrimina más adecuadamente a líneas finas, puntos aislados y ruido
- La segunda derivada devuelve doble respuesta a un cambio de intensidad (valor absoluto: doble borde)
- El cambio de signo de la segunda derivada indica si el cambio de intensidad es de negro a blanco o viceversa

En la figura A2-9 se muestra diferentes detectores efectuados a la imagen mdb227.



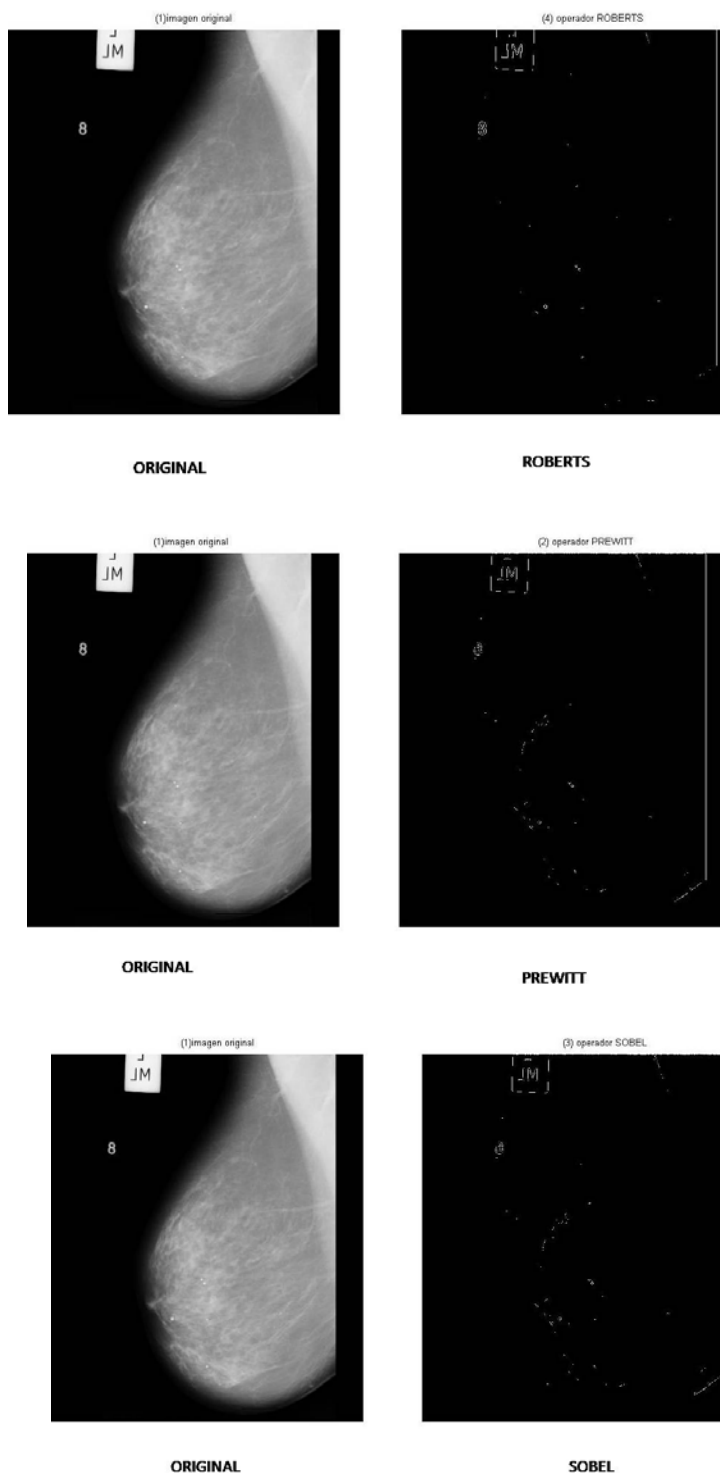


Figura A2-9. Ejemplos de aplicar operadores a una imagen. A) Roberts, B) Prewitt, C) Sobel

La mayoría de las veces para obtener el resultado deseado es preciso combinar diferentes tipos de filtros junto con otros tipos de operaciones de imágenes como operaciones aritméticas, manipulación de histogramas, etc. Algunas de estas combinaciones da lugar a nuevos tipos de filtros como los denominados **top hat**.

Dependiendo de lo que uno quiera detectar es importante definir en el procedimiento si el objeto tiene predominancia blanca u oscura, esto es muy importante al considerar la aplicación de la tabla de los diferentes operadores. En la siguiente sección vamos a explicar el filtro **top hat** el cual se escogió después de trabajar a las múltiples opciones de los operadores.

## A-2.2 TECNICA TOP HAT - OPERADORES MORFOLÓGICOS

La transformación de Top hat es el resultado de la aplicación de las técnicas morfológicas utilizadas en los procesos de segmentación. Tenemos las siguientes tipos de técnicas más utilizadas y de gran importancia en el procesamiento de imágenes:

- Erosión
- Dilatación
- Apertura
- Cierre
- Morfología en escala de grises
- Transformada top hat.

[8] La teoría básica de los operadores morfológicos se basan en la teoría de conjuntos (aportada por *Minkowsky* 1901 y *hadwiger* 1956) y que posteriormente se desarrollo en la escuela de parís: **Centre of morphology matematique de l'cole des mines of paris**. El objetivo de las operaciones morfológicas es la extracción de estructuras geométricas en los conjuntos en los que se opere (imágenes filtradas) mediante la utilización de otra forma conocida denominado elemento estructurante (EE) , el tamaño y la forma de este elemento se escoge a priori de acuerdo a la morfología del conjunto sobre el que va a interactuar y de acuerdo a la extracción de formas que se desean obtener.

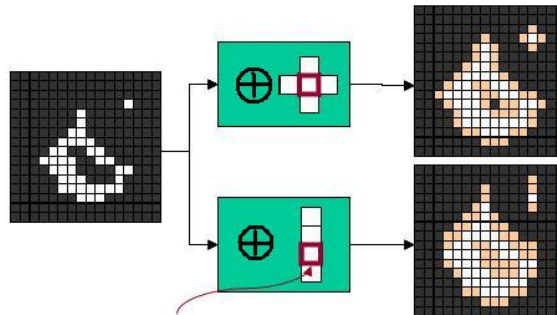
## Dilatación.

[9] Dilatación de la imagen  $A$  por el EE (elemento estructurante)  $C$ : En rigor,  $A$  no es una imagen, sino un conjunto, y el resultado de su dilatación,  $\delta_C(A)$ , tampoco es una imagen, sino otro conjunto. Sin embargo, por brevedad, en lo sucesivo utilizaremos expresiones como "la imagen  $A$ " para referirnos a "la imagen  $a(x,y)$  definida por el conjunto  $A$ ".

$$\delta_C(A) = A \oplus C = \{\mathbf{x} \mid (C)_{\mathbf{x}} \cap A \neq \emptyset\} \quad (5.10)$$

Se obtiene en base a la reflexión de  $C$  con respecto a su origen y un desplazamiento  $\mathbf{x}$ .

La salida de la dilatación (figura A2-10) es el conjunto de puntos barridos por el centro del EE mientras algún punto de  $C$  coincide con alguno de  $A$ . Alternativamente (véase figura), puede interpretarse la dilatación como el resultado de reemplazar cada píxel blanco de la imagen original por una réplica del elemento estructurante EE.



(En esta figura y las siguientes, el origen de coordenadas del EE aparece destacado)

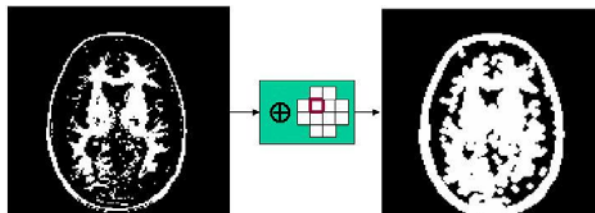


Figura A2-10. [9] Ejemplo del operador morfológico de dilatación.

### Propiedades:

- Añade todos los puntos del fondo que tocan el borde de un objeto  $\rightarrow$  la dilatación es extensiva.
- Rellena entrantes en los que no quepa el EE (pequeños agujeros y bahías)

## Erosión

[9] La salida de la erosión es el conjunto de puntos barridos por el centro del EE mientras se cumpla que todos los puntos de  $C$  estaban contenidos en  $A$ . Elimina grupos de píxeles donde el EE (elemento estructurante) no cabe (figura A2-11).

$$\varepsilon_C(A) = A \ominus C = \{x \mid C_x \subset A\} \quad (5.11)$$

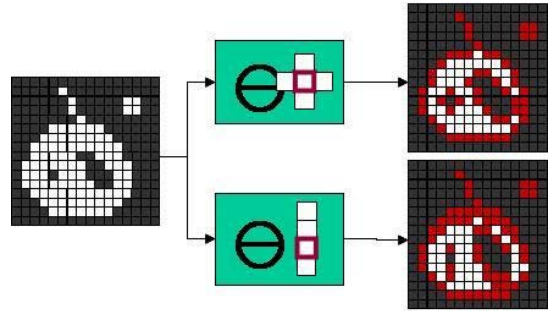


Figura A2-11 [9]. Ejemplo del operador morfológico de erosión

(Los píxeles que cambian de blanco al negro por efecto de la erosión aparecen en rojo).

### Propiedades:

- La erosión es anti extensiva  $\rightarrow$  reduce el tamaño del objeto
- Elimina elementos (figura A2-12) en los que no quepa el EE (pequeñas islas y protuberancias). Una de las aplicaciones más típicas de la erosión es la eliminación de detalles irrelevantes (desde el punto de vista del tamaño)

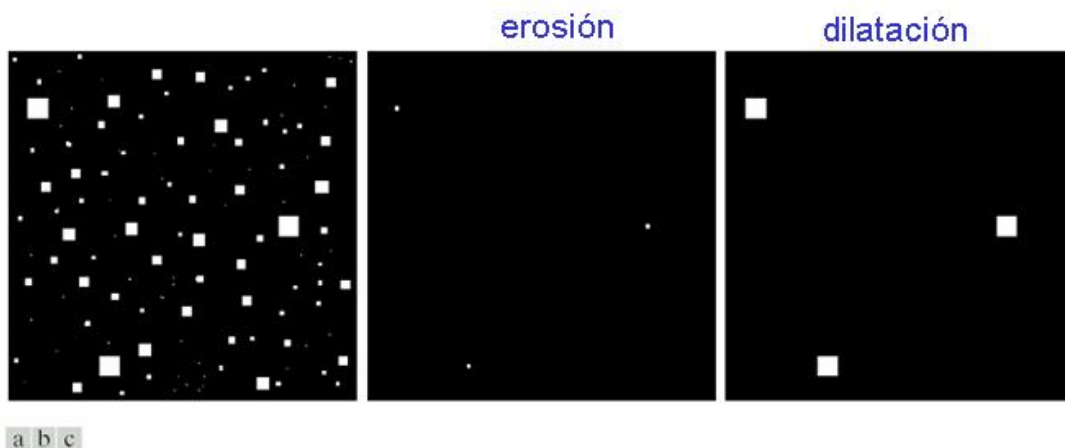
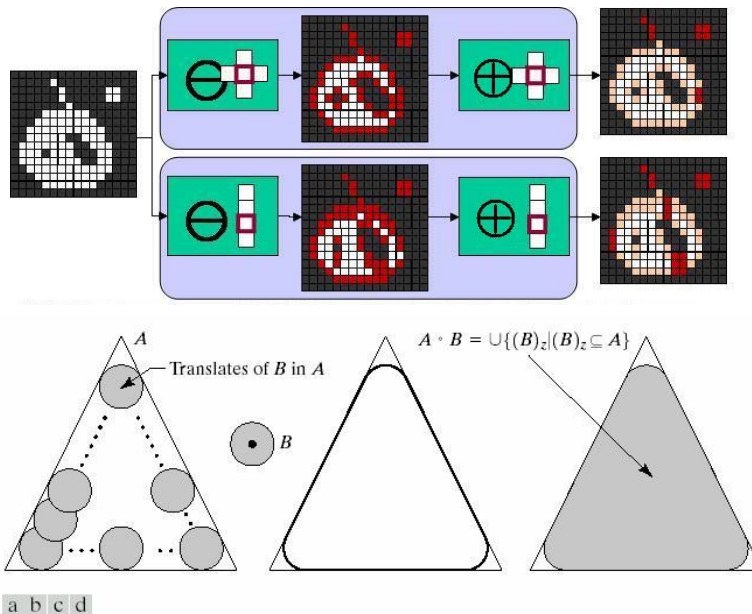


FIGURE 9.7 (a) Image of squares of size 1, 3, 5, 7, 9, and 15 pixels on the side. (b) Erosion of (a) with a square structuring element of 1's, 13 pixels on the side. (c) Dilatation of (b) with the same structuring element.

Figura A2-12. Otro ejemplo de cómo la erosión y dilatación separa elementos de una imagen.

## Apertura

[9] Es la composición de un operador de erosión y otro de dilatación con el mismo elemento estructurante (figura A2-13).



**FIGURE 9.8** (a) Structuring element  $B$  "rolling" along the inner boundary of  $A$  (the dot indicates the origin of  $B$ ). (c) The heavy line is the outer boundary of the opening. (d) Complete opening (shaded).

Tomada de Gonzalez, Woods, *Digital Image Processing* (2nd Ed.), Prentice Hall, 2002

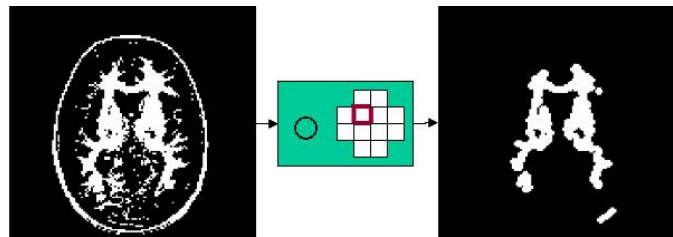


Figura A2-13 [9] Ejemplo del operador morfológico de apertura

### Propiedades:

- Es un operador anti extensivo:  $A \circ B$  es un subconjunto de  $A$
- Es un operador idempotente:  $(A \circ B) \circ B = A \circ B$

### Algunos efectos:

- Alisa contornos (redondear las esquinas donde no quepa el EE).
- Eliminar las protuberancias donde no quepa el EE
- Separa objetos en puntos estrechos

## Cierre

Es la composición de un operador de dilatación seguido de otro de erosión con el mismo elemento estructurante (figura A2-14):

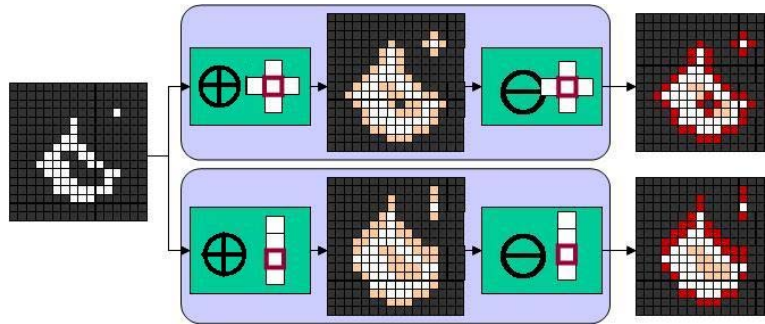


Figura A2-14. [9] Ejemplo del operador morfológico de cierre

### Propiedades:

1. Es un operador extensivo:  $A \subset A \bullet B$
2. Es un operador idempotente:  $(A \bullet B) \bullet B = A \bullet B$

### Algunos efectos:

- Tiende a alisar porciones del contorno, fusiona estrechas grietas, y rellena vacíos en el contorno.
- Rellena agujeros pequeños
- Elimina los entrantes pequeños
- Conecta objetos vecinos

**Apertura y Cierre son Duales** con respecto al complementario y a la transposición:

$$\rightarrow (A \bullet B)^c = (A^c \circ B^c)$$



## ANEXO 3

### TECNICAS DE CLASIFICACIÓN USUALES

#### A3-1 REDES NEURONALES

El cerebro es un procesador de información muy notable, es capaz de procesar a gran velocidad grandes cantidades de información procedentes de los sentidos, combinarla o compararla con la información almacenada y dar respuestas adecuadas incluso en situaciones nuevas. Aunque todavía no se sabe mucho sobre la forma o manera en que el cerebro aprende a procesar la información, se han desarrollado imitar tales habilidades; denominados Redes Neuronales Artificiales (RNA). La elaboración de estos modelos supone en primer lugar la deducción de los rasgos o características esenciales de las neuronas biológicas y segundo lugar, la implementación se pueda simular [9].

El cerebro humano se compone de decenas de billones de neuronas interconectadas entre sí formando circuitos o redes que desarrollan funciones específicas. Una neurona típica recoge señales procedentes de otras neuronas a través de una inmensa cantidad de delicadas estructuras llamadas dendritas. La neurona emite impulsos de actividad eléctrica a lo largo de una fibra larga y delgada, denominada axón, que se divide en millares de ramificaciones (ver Figura A3-1) [10,11]. Las extremidades de estas ramificaciones llegan hasta las dendritas de otras neuronas y establecen conexiones llamadas sinapsis, en las cuales se produce una transformación de un impulso eléctrico en un mensaje neuroquímico mediante la liberación de sustancias llamadas neurotransmisor.

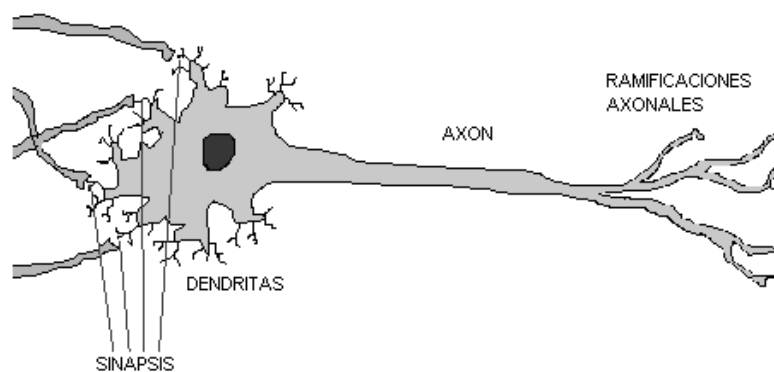


Figura A3-1.[9] Neurona y sus conexiones sinápticas



El efecto de los neurotransmisores sobre la neurona receptora puede ser excitatorio o inhibitorio, y es variable, de manera que se puede hablar de la fuerza o efectividad de una sinapsis [2]. Las señales excitatorias e inhibitorias recibidas por una neurona se combinan, y en función de la estimulación total recibida la neurona toma un cierto nivel de activación, que se traduce en la generación de breves impulsos nerviosos con una determinada frecuencia o tasa de disparo, y su propagación a lo largo del axón hacia las neuronas con las cuales se establece comunicación. Este efecto se muestra en la Figura A3-1 como la sinapsis.

De esta manera, la información se transmite de unas neuronas a otras y va siendo procesada a través de las conexiones sinápticas y las propias neuronas. El aprendizaje de las redes neuronales se produce mediante la variación de la efectividad de las sinapsis.

### ***Redes neuronales artificiales***

Conforme la ciencia y la tecnología han ido avanzando uno de los grandes retos es construir sistemas inteligentes, entendiéndose como sistema como cualquier dispositivo físico, lógico capaz de realizar una tarea requerida, esta es una de los objetivos de la disciplina científica conocida como inteligencia artificial.

Dentro de la Inteligencia Artificial se pueden distinguir dos grandes áreas. Una se ocupa de la construcción de sistemas con características que se puedan definir como inteligentes. A este campo se lo denomina Inteligencia Artificial Simbólica [12]. En este caso, se define el problema a resolver y se diseña el sistema capaz de resolverlo siguiendo esquemas prefijados por la disciplina. Los Sistemas Expertos siguen este esquema: se introducen una serie de reglas lógicas que recogen el conocimiento de un experto sobre una materia, y mediante mecanismos de inferencia parecidos a los que empleamos al razonar, se sacan conclusiones.

Frente a esta perspectiva se encuentra la otra gran área de la Inteligencia Artificial: la Sub-simbólica [12]. En este caso no se realizan diseños a alto nivel de sistemas capaces de resolver los problemas utilizando las técnicas de la disciplina, sino que se parte de sistemas genéricos que van adaptándose y construyéndose hasta formar por sí mismos un sistema capaz de resolver el problema.

Una perspectiva simbólica consiste en el estudio de los mecanismos de razonamiento humano a alto nivel: como nos enfrentamos a un problema, cómo lo abordamos y resolvemos; y se elaboran programas que realizan las mismas tareas. Cuanto mejor haya podido entenderse la forma de razonar humana, más eficiente será el sistema producido a la hora de resolver los problemas planteados. Es en este campo donde se encuentran las redes neuronales artificiales.

Existen diferentes formas de definir lo que son las redes neuronales: “una *red neuronal artificial RNA* es un sistema masivo de procesamiento paralelo distribuido que está compuesto de unidades de proceso simples, y tiene una predisposición natural para almacenar conocimiento experimental y transformarlo para uso predeterminado”.

Las RNA se parecen al cerebro en dos aspectos:

- La RNA adquiere conocimiento de su entorno a través de un proceso de aprendizaje.
- La fuerza o importancia de las interconexiones entre las neuronas, conocidas como pesos sinápticos, se usan para almacenar el conocimiento adquirido.

### **Estructura básica**

Una **RNA** se compone de varias unidades de proceso, cada una de las cuales se compone de un núcleo encargado de los cálculos, una red o vector de conexiones de entrada, y una salida (ver Figura A3-2) [10,12]:

**Conexiones ponderadas:** hacen el papel de las conexiones sinápticas, el peso de la conexión equivale a la fuerza o efectividad de la sinapsis. La existencia de conexiones determina si es posible que una unidad influya sobre otra, el valor de los pesos y el signo de los mismos definen el tipo (excitatorio/inhibitorio) y la intensidad de la influencia.

**Función de red:** Calcula el valor de base o entrada total a la unidad, generalmente como una simple suma ponderada de todas las entradas, es decir, de las entradas multiplicadas por el peso o valor de las conexiones. Equivale a la combinación de las señales excitadoras e inhibitorias de las neuronas biológicas.

**Función de activación:** Se encarga de calcular el nivel o estado de activación de la neurona en función de la entrada total.

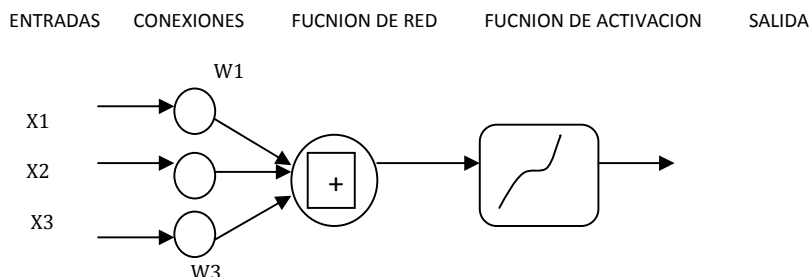


Figura A3-2. Diagrama básico de una red RNA

**Salida:** calcula la salida de la neurona en función de la activación de la misma neurona, normalmente no se aplica más que la función identidad y se toma como salida el valor de activación. El valor de salida cumplirá la función de la tasa de disparo en las neuronas biológicas.

### Implementación de una RNA

Existen dos fases típicas en la implementación y puesta en producción de una RNA: fase de aprendizaje o de entrenamiento, y la fase de pruebas.

- en la fase de aprendizaje (figura A3-3), los parámetros de la red se obtiene a partir de datos representativos llamados patrones de entrenamiento o base de datos de referencia. El objetivo de esta fase es lograr que la RNA entregue la salida deseada para los datos de entrada con la mayor eficiencia posible. Existen tres (3) tipos de procedimientos utilizados para esta fase:
  - **Supervisado:** los datos para el entrenamiento están constituidos por varios pares de patrones de entrenamiento entrada / salida, el hecho de conocer la salida implica que el entrenamiento se beneficia con la supervisión del maestro o especialista, muchas veces este proceso es en forma manual prueba a prueba.
  - **No supervisado:** los datos de entrenamiento consisten solo en patrones de entrada, por lo tanto la red es entrenada sin el beneficio de un especialista o maestro. La red aprende a adaptarse acorde con la experiencia aprendida prueba a prueba.
  - **Híbrida:** es una mezcla de las dos anteriores y cuyo propósito es optimizar el aprendizaje.
- En la fase de prueba, se utilizan datos diferentes a la de aprendizaje. En esta fase por lo general se verifica la eficiencia de la RNA diseñada.

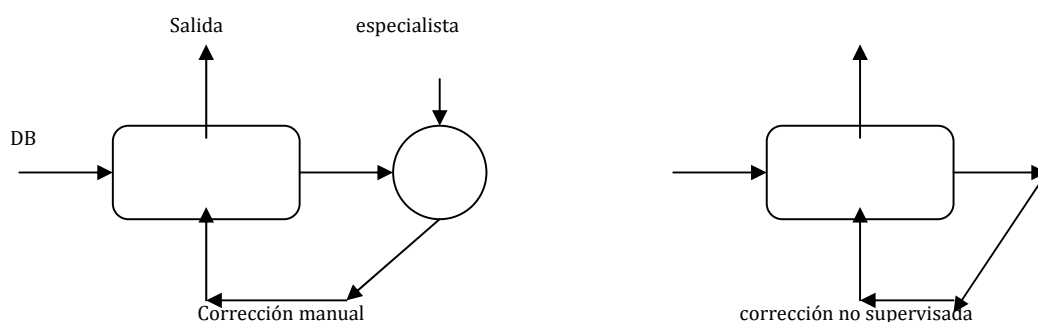


Figura A3-3. Diagrama de entrenamiento de una RNA

### **Clasificación de las RNA**

Existen varios tipos de clasificación de las RNA; por tipo de aprendizaje (mencionados anteriormente), por mecanismos de asociación, por arquitectura, etc. En la clasificación basada por la arquitectura se tienen dos grupos principales:

Unidireccional: También es conocida como FeedForward. La estructura de esta red consiste en capas de neuronas donde la salida de una neurona de una capa, alimenta todas las neuronas de la capa siguiente tal como se observa en la Figura A3-4.

Recurrente: e aquella red que posee las conexiones de realimentación entre la salida y capas intermedias (ver figura A3-4).

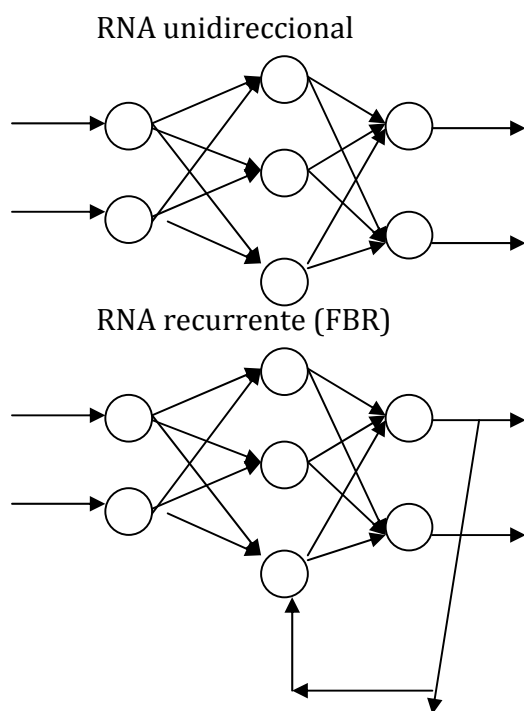


Figura A3-4. Esquemas de redes neuronales.

En general los siguientes parámetros son importantes para definir la arquitectura de una red neuronal: número de capas, número de neuronas en cada capa y tipo de conexión entre dos neuronas, que definen en si la red es unidireccional o recurrente.

### Activación de una RNA

La función de activación de una RNA depende del valor que se va obteniendo en cada una de las unidades de proceso que constituyen a la RNA, se filtra a través de una función conocida como función de activación, y es la que nos da la salida de cada neurona o de la red completa. De acuerdo al objetivo del entrenamiento de la RNA se suele escoger una función de activación como las que se muestran en la siguiente tabla A3-1:

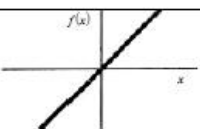
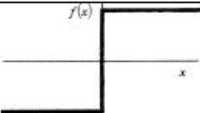
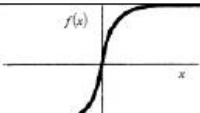
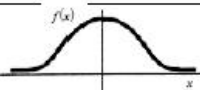
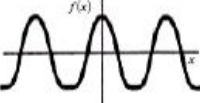
Nombre	Función	Rango	Gráfica
Identidad	$y = x$	$[-\infty, +\infty]$	
Escalón o limitador duro	$y = \text{sign}(x)$ $y = \text{hardli m}(x)$	$\{-1, +1\}$ $\{0, +1\}$	
Sigmoidal	$y = \frac{1}{1 + e^{-x}}$ $y = \text{tgh}(x)$	$[0, +1]$ $[-1, +1]$	
Gaussiana	$y = Ae^{-Bx^2}$	$[0, +1]$	
Sinusoidal	$y = A \text{sen}(\omega x + \varphi)$	$[-1, +1]$	

Tabla A3-1 [9] Funciones de activación más utilizadas en RNA.

### Redes RNA más importantes.

Aunque son muchas los tipos de redes neuronales, profundizar sobre estas nos requeriría de una extensa explicación de sus características, activaciones, implementaciones, solo mencionaremos las más utilizadas y que en trabajos similares a la detección automática de calcificaciones han sido utilizadas (ver tabla A3-1).

**Perceptron:** [12] este fue el primer modelo de RNA desarrollado por Rosenblatt (1958) y despertó un gran interés en los años 60, debido a su capacidad para prender a reconocer patrones sencillos. Un perceptron está formado por varias neuronas lineales como entradas a la red y una sola salida (figura A3-5). La única neurona de salida realiza la suma ponderada de las

entradas, resta el umbral y pasa el resultado a una función de transferencia tipo escalón.

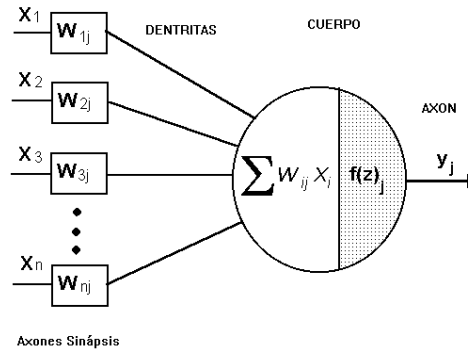


Figura A3-5. [11] Modelo del perceptron

Aquí el cuerpo de la neurona se representa como un sumador lineal de los estímulos externos  $Z_j$ , seguida de una función no lineal  $y_j = f(Z_j)$ . La función  $f(Z_j)$  es llamada la función de activación (función escalón, ver tabla A2-1), y es la función que utiliza la suma de estímulos para determinar la actividad de salida de la neurona.

Los pesos pueden ser positivos (excitación) o negativos (inhibición). Los pesos junto con las funciones  $f(Z)$  dictan la operación de la red neuronal. Normalmente las funciones no se modifican de tal forma que el estado de la red neuronal depende del valor de los factores de peso (sinapsis) que se aplica a los estímulos de la neurona.

En un perceptrón, cada entrada es multiplicada por el peso  $W$  correspondiente, y los resultados son sumados, siendo evaluados contra el valor de umbral, si el resultado es mayor al mismo, el perceptrón se activa.

**Entrenamiento:** El entrenamiento de un perceptrón es por medio de la regla de aprendizaje delta: Para cada peso  $W$  se realiza un ajuste  $dW$  según la regla:

$$dW = LR ( T - Y ) X$$

Donde  $LR$  es la razón de aprendizaje,  $T$  el valor deseado,  $Y$  el valor obtenido, y  $X$  la entrada aplicada al perceptrón.

**Limitaciones:** El perceptrón es capaz tan sólo de resolver funciones definidas por un hiperplano (objeto de dimensión N-1 contenida en un espacio de dimensión N). que corte un espacio de dimensión N. Un ejemplo de una función que no puede ser resuelta es el operador lógico XOR. Una explicación más sencilla de un hiperplano sería, hablando en un plano de dos dimensiones, una línea que separa a los elementos existentes en dos grupos. El perceptrón sólo puede resolver una función, si todos los posibles resultados del problema pueden separarse de ésta forma (en dos secciones) es decir, que no se combinen entre sí

**Perceptron MLP (multicapa):** un perceptron MLP es una red del tipo *feedforward* de varias capas de neuronas entre la entrada y la salida, esta forma de RNA permite establecer regiones de decisión más complejas que las de dos planos (figura A3-6). El problema está en el entrenamiento de las capas ocultas.

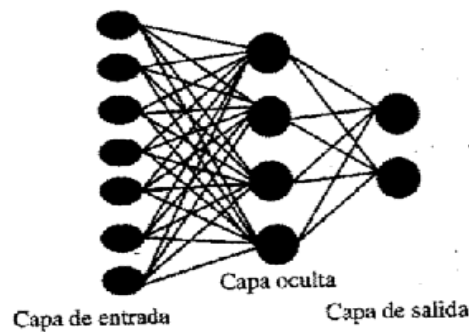


Figura A3-6. [12] Modelo de una RNA - MLP

**Adaline - madeline:** [11] las redes ADELIN (Adaptative linear element) y Madeline (multiple adeline) fueron desarrolladas por Bernie Widrow (1960) en la universidad de stanford (USA) poco después del aporte del perceptron. La arquitectura son esencialmente la misma a la del perceptron, ambas usan función de transferencia del tipo escalón, con una única neurona de salida, mientras que madeline puede tener varias salidas. La gran diferencia con el perceptron es la técnica de aprendizaje en la cual se utiliza la regla Delta de hidrow-Hoff o regla del mínimo error cuadrado medio (LMS), basado en la búsqueda del mínimo de una expresión de error entre la salida deseada y la salida lineal obtenida antes de aplicarle la función de activación escalón. Esto hace que pueda procesar información analógica.

**Modelo Hopfield:** [10] Las redes de Hopfield son redes de adaptación probabilística, recurrentes, funcionalmente entrarían en la categoría de las memorias auto asociativas, es decir, que aprenden a reconstruir los patrones de entrada que memorizaron durante el entrenamiento. Son arquitecturas de

una capa con interconexión total, funciones de activación booleana de umbral (cada unidad puede tomar dos estados, 0 o 1, dependiendo de si la estimulación total recibida supera determinado umbral), adaptación probabilística de la activación de las unidades, conexiones recurrentes y simétricas, y regla de aprendizaje no supervisado. Mientras que las redes en cascada (no recurrentes) dan soluciones estables, los modelos recurrentes dan soluciones inestables (dinámicas), lo que no siempre es aconsejable.

La principal aportación de Hopfield consistió precisamente en conseguir que tales redes recurrentes fueran así mismo estables. Imaginó un sistema físico capaz de operar como una memoria auto asociativa, que almacenara información y fuera capaz de recuperarla aunque la misma se hubiera deteriorado.

La Red de Hopfield es recurrente y completamente conectada. Funciona como una memoria asociativa no lineal que puede almacenar internamente patrones presentados de forma incompleta o con ruido. De esta forma puede ser usada como una herramienta de optimización (figura A3-7). El estado de cada neurona puede ser actualizado un número indefinido de veces, independientemente del resto de las neuronas de la red pero en paralelo.

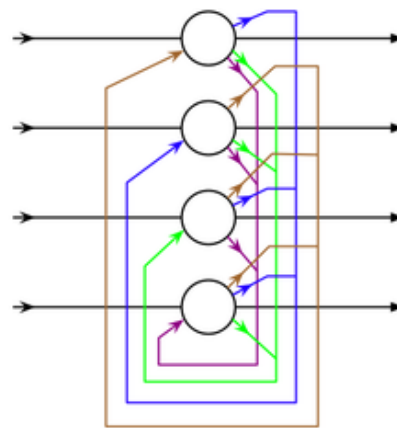


Figura A3-7 [13] Modelo de una RNA Hopfield

Las redes de Hopfield se han aplicado a campos como la percepción el reconocimiento de imágenes y optimización de problemas, mostrando gran inmunidad al ruido y robustez. Incluso se han llegado a desarrollar chips específicos para este tipo de redes. El estudio de las representaciones de secuencias temporales es un área de gran interés, con aplicaciones en reconocimiento automático de voces y movimientos.

***Función de base radial (RBF):*** [9] Las Redes de Neuronas de Base Radial son redes multicapa con conexiones hacia delante, y se caracterizan por estar formadas por una única capa oculta, y cada neurona de esta etapa posee un



carácter local, en sentido de que cada neurona de la red se activa en una región diferente del espacio de patrones de entrada. Este carácter local viene dado por el uso de las llamadas funciones de base radial como funciones de activación. En 1988, Broomhead y Lowe introducen la Función de Base Radial (RBF) como un método para hacer ajuste a funciones no lineales (clasificación no lineal), aunque la mayor contribución a la teoría, diseño y aplicaciones de este tipo de Uno de los objetivos iniciales de los desarrolladores de esta red era construir una red de neuronas que requiriese menos tiempo que el perceptron multicapa y de ese modo disponer de una red de neuronas que pudiera ser apropiada para aplicaciones en tiempo real. Esto se consiguió incorporando funciones de activaciones locales en las neuronas ocultas de la red, lo cual permitía que sólo unas pocas neuronas ocultas tuvieran que ser procesadas para nuevos patrones de entrada.

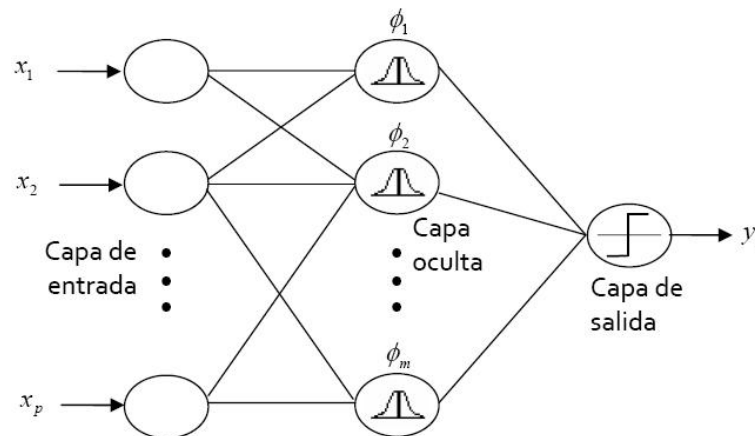


Figura A3-8. [9] Modelo de una RNA función de base radial RBF.

La RBF en su segunda capa (capa oculta) es del tipo gaussiana (ver tabla 5). Y lineal en la capa de salida. La primer capa (capa de entrada) se encarga de leer los datos que se van a introducir a la RBF. La segunda capa (capa oculta) está completamente interconectada entre todos sus nodos con las capas de entrada y de salida y se activa a través de una función radial, y, La tercer y última capa (capa de salida), se interconecta a la capa oculta y es activada a través de una función lineal continua.

**SVM (support vector machine):** son otra variedad de RNA la cual se basa en algoritmos supervisados de aprendizaje, Desarrollada por **Vladimir Vapnik** en los Lab Bell de AT&T. Dado un conjunto de ejemplos de entrenamiento (de muestras) podemos etiquetar las clases y entrenar una SVM para construir un modelo que prediga la clase de una nueva muestra. Intuitivamente, una SVM es un modelo que representa a los puntos de muestra en el espacio, separando las clases por un espacio lo más amplio posible. Cuando las nuevas muestras se ponen en correspondencia con dicho modelo, en función de su proximidad pueden ser clasificadas a una u otra clase. Más formalmente, una SVM construye un hiperplano o conjunto de

hiperplaños en un espacio de dimensionalidad muy alta (o incluso infinita) que puede ser utilizado en problemas de clasificación o regresión. Una buena separación entre las clases permitirá una clasificación correcta.

### A3-2 LÓGICA DIFUSA

Otra técnica muy utilizada para la clasificación de imágenes médicas es la lógica difusa, la cual nace a partir de un artículo del ingeniero Lofti A. Zadeh titulado "Fuzzy sets" ("Conjuntos difusos"), que se publicó en 1965. La lógica difusa difiere de la lógica clásica: "en la lógica clásica, según Bertrand Russell, de un conjunto de antecedentes lógicos se infieren unos consecuentes". Se trata de reglas de la forma "si" (antecedente), "entonces" (consecuente). "En la lógica difusa, tanto los antecedentes como los consecuentes son conjuntos difusos, que intentan cuantificar la vaguedad de los calificadores". O blanco o negro. O verdadero o falso. Si algo es verdadero, no puede ser falso. Y a la inversa. La eficiencia de este principio de la lógica clásica que fundamenta el método científico se ve limitada cuando se penetra en el mundo de los *grises*. La lógica clásica tiene dificultades para definir cuantitativamente elementos que son blancos pero que también son negros, y que se diferencian unos de otros por un leve matiz que permite a los humanos calificarlos de gris claro, de oscuro, de muy oscuro o, por fin, de negro. ¿Cuántos graños de arena debe perder un desierto para dejar de ser considerado como tal? ¿Cuántos grados debe bajar la temperatura de una sala para que consideremos que hace frío? La lógica difusa (*fuzzy logic*) abre una vía para tratar científicamente estos conceptos cualitativos. "Es una extensión de la lógica clásica que sirve para cuantificar la vaguedad".

Cómo se aplican los conjuntos difusos, por ejemplo, en un sistema de control que regula la luz de una sala. Los parámetros del sistema son adjetivos, habitualmente entre tres y seis, que definen la luz, como:

"poquísima, poca, suficiente, mucha y muchísima". Entonces, se realizan varias medidas de la luz y se asigna el grado en que cada una pertenece a un adjetivo. Así:

- Una luz muy tenue a poquísima = 0,6
- Poquísima a Poca = 0,3
- Poca a suficiente = 0,1

Cada magnitud medida queda repartida entre diferentes conjuntos y su grado de pertenencia a un conjunto se representa con una función, es un rango no son extremos como por lo regular hace la lógica clásica, esta es la gran diferencia y la hace muy útil en muchas aplicaciones en el área industrial, doméstica y vehicular (figura A3-9). Se han hecho también grandes aportes mediante el uso de esta técnica en la detección de calcificaciones.

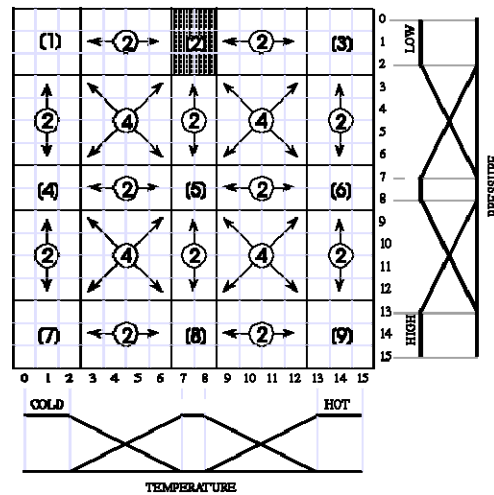


Figura A3-9. Diagrama de función de decisión de lógica difusa.

En 1991 Lofti A. Zadeh, profesor de la Universidad de Berkeley y padre de la lógica difusa, acuñó el término *softcomputing* (en castellano, informática blanda). Esta rama de la inteligencia artificial se ocupa del diseño de sistemas expertos capaces de gestionar de una manera adecuada información imprecisa, incierta y/o incompleta. La lógica difusa es una de las principales técnicas de esta rama, junto con los algoritmos evolutivos y las redes neuronales. Más de 2.000 expertos y expertas en informática blanda de todo el mundo se reunieron en Barcelona en el *IEEE World Congress on Computational Intelligence*, en julio de 2010.

### A3-3 ALGORÍTMOS GENÉTICOS

[15] El algoritmo genético ( **AG** ) es una técnica de búsqueda basada en la teoría de la evolución de Darwin, que ha cobrado tremenda popularidad en todo el mundo durante los últimos años. En los últimos años, la comunidad científica internacional ha mostrado un creciente interés en una nueva técnica de búsqueda basada en la teoría de la evolución y que se conoce como el **algoritmo genético**. Esta técnica se basa en los mecanismos de *selección* que utiliza la naturaleza, de acuerdo a los cuales los individuos *más aptos* de una población son los que sobreviven, al adaptarse más fácilmente a los cambios que se producen en su entorno. Hoy en día se sabe que estos cambios se efectúan en los genes de un individuo, y que sus atributos más deseables

(i.e., los que le permiten adaptarse mejor a su entorno) se transmiten a sus descendientes cuando éste se reproduce en cada nueva generación.

Un investigador de la Universidad de Michigan llamado **John Holland** era consciente de la importancia de la selección natural, y a fines de los 60s desarrolló una técnica que permitió incorporarla a un programa. Su objetivo era lograr que las computadoras aprendieran por sí mismas. A la técnica que inventó Holland se le llamó originalmente "planes reproductivos", pero se hizo popular bajo el nombre "algoritmo genético" tras la publicación de su libro en 1975.

Una definición bastante completa de un algoritmo genético es la propuesta por John Koza [16] (stanford university):

*"Es un algoritmo matemático altamente paralelo que transforma un conjunto de objetos matemáticos individuales con respecto al tiempo usando operaciones modeladas de acuerdo al principio Darwiniano de reproducción y supervivencia del más apto, y tras haberse presentado de forma natural una serie de operaciones genéticas de entre las que destaca la recombinación sexual. Cada uno de estos objetos matemáticos suele ser una cadena de caracteres (letras o números) de longitud fija que se ajusta al modelo de las cadenas de cromosomas, y se les asocia con una cierta función matemática que refleja su aptitud. "*

Los Algoritmos Genéticos (AGs) son métodos adaptativos que pueden usarse para resolver problemas de búsqueda y optimización. Están basados en el proceso genético de los organismos vivos. A lo largo de las generaciones, las poblaciones evolucionan en la naturaleza de acorde con los principios de la selección natural y la supervivencia de los más fuertes, postulados por **Darwin**. Por imitación de este proceso, los Algoritmos Genéticos son capaces de ir creando soluciones para problemas del mundo real. La evolución de dichas soluciones hacia valores óptimos del problema depende en buena medida de una adecuada codificación de las mismas.

Un **algoritmo genético** consiste en una función matemática o una rutina de software que toma como entradas a los ejemplares y retorna como salidas cuáles de ellos deben generar descendencia para la nueva generación.

Versiones más complejas de algoritmos genéticos generan un ciclo iterativo que directamente toma a la especie (el total de los ejemplares) y crea una nueva generación que reemplaza a la antigua una cantidad de veces determinada por su propio diseño. Una de sus características principales es la de ir perfeccionando su propia heurística en el proceso de ejecución, por lo que no requiere largos períodos de entrenamiento especializado por parte del ser humano, principal defecto de otros métodos para solucionar problemas, como los Sistemas Expertos.

Los principios básicos de los Algoritmos Genéticos fueron establecidos por Holland, y se encuentran bien descritos en varios textos. Goldberg, Davis, Michalewicz, Reeves.

En la naturaleza los individuos de una población compiten entre sí en la búsqueda de recursos tales como comida, agua y refugio. Incluso los miembros de una misma especie compiten a menudo en la búsqueda de un compañero. Aquellos individuos que tienen más éxito en sobrevivir y en atraer compañeros tienen mayor probabilidad de generar un gran número de descendientes. Por el contrario individuos poco dotados producirán un menor número de descendientes. Esto significa que los genes de los individuos mejor adaptados se propagarán en sucesivas generaciones hacia un número de individuos creciente. La combinación de buenas características provenientes de diferentes ancestros, puede a veces producir descendientes "superindividuos", cuya adaptación es mucho mayor que la de cualquiera de sus ancestros. De esta manera, las especies evolucionan logrando unas características cada vez mejor adaptadas al entorno en el que viven.

Los Algoritmos Genéticos usan una analogía directa con el comportamiento natural. Trabajan con una población de individuos, cada uno de los cuales representa una solución factible a un problema dado. A cada individuo se le asigna un valor ó puntuación, relacionado con la bondad de dicha solución. En la naturaleza esto equivaldría al grado de efectividad de un organismo para competir por unos determinados recursos. Cuanto mayor sea la adaptación de un individuo al problema, mayor será la probabilidad de que el mismo sea seleccionado para reproducirse, cruzando su material genético con otro individuo seleccionado de igual forma. Este cruce producirá nuevos individuos. Descendientes de los anteriores. Los cuales comparten algunas de las características de sus padres. Cuanto menor sea la adaptación de un individuo, menor será la probabilidad de que dicho individuo sea seleccionado para la reproducción, y por tanto de que su material genético se propague en sucesivas generaciones.

De esta manera se produce una nueva población de posibles soluciones, la cual reemplaza a la anterior y verifica la interesante propiedad de que contiene una mayor proporción de buenas características en comparación con la población anterior. Así a lo largo de las generaciones las buenas características se propagan a través de la población. Favoreciendo el cruce de los individuos mejor adaptados, van siendo exploradas las áreas más prometedoras del espacio de búsqueda. Si el Algoritmo Genético ha sido bien diseñado, la población convergerá hacia una solución óptima del problema.

[17] Lo que aventaja a los AG frente a otros algoritmos tradicionales de búsqueda es que se diferencian de estos en los siguientes aspectos:

- Trabajan con una codificación de un conjunto de parámetros, no con los parámetros mismos.
- Trabajan con un conjunto de puntos, no con un único punto y su entorno (su técnica de búsqueda es global.) Utilizan un subconjunto del espacio total, para obtener información sobre el universo de búsqueda, a través de las evaluaciones de la función a optimizar. Esas evaluaciones se emplean de forma eficiente para clasificar los subconjuntos de acuerdo con su idoneidad.
- No necesitan conocimientos específicos sobre el problema a resolver; es decir, no están sujetos a restricciones. Por ejemplo, se pueden aplicar a funciones no continuas, lo cual les abre un amplio campo de aplicaciones que no podrían ser tratadas por los métodos tradicionales.
- Utilizan operadores probabilísticos, en vez de los típicos operadores determinísticos de las técnicas tradicionales.
- Resulta sumamente fácil ejecutarlos en las modernas arquitecturas masivas en paralelo.

*Procedimiento del proceso de un AG (ver figura A3-10):*

- - Generar una población inicial.
- - Iterar hasta un criterio de parada.
- - Evaluar cada individuo de la población.
- - Seleccionar los progenitores.
- - Aplicar el operador de cruce y mutación a estos progenitores.
- - Incluir la nueva descendencia para formar una nueva generación.

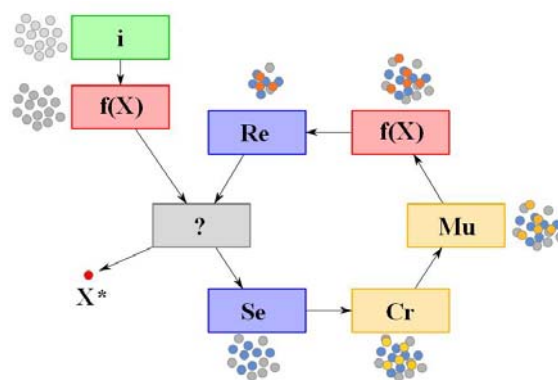


Figura A3-10. [18] diagrama de mutación según algoritmos genéticos.

*Funcionamiento de un Algoritmo Genético Simple:*

El siguiente algoritmo genérico nos muestra una simple implementación de un AG.

```
BEGIN /* Algoritmo Genético Simple */
  Generar una población inicial.
  Computar la función de evaluación de cada individuo.
  WHILE NOT Terminado DO
    BEGIN /* Producir nueva generación */
      FOR Tamaño población/2 DO
        BEGIN /*Ciclo Reproductivo */
          Seleccionar dos individuos de la anterior generación,
            Para el cruce (probabilidad de selección proporcional a la
            función de evaluación del individuo).
          Cruzar con cierta probabilidad los dos
            Individuos obteniendo dos descendientes.
          Mutar los dos descendientes con cierta probabilidad.
          Computar la función de evaluación de los dos
            Descendientes mutados.
          Insertar los dos descendientes mutados en la nueva generación.
        END
      IF la población ha convergido THEN
        Terminado:= TRUE
    END
  END
END
```

#### **A3-4 REDES BAYESIANAS**

[19] El Teorema de Bayes es ampliamente utilizado en diversas áreas del conocimiento debido a su fortaleza garantizado por el alto nivel de flexibilidad con el que cuenta, flexibilidad que permite realizar inferencias a todo nivel, los mismos que, son altamente fidedignos y cercanos al más exhaustivo método riguroso (cómo por ejemplo, comparando sus resultados con los obtenidos con los de una Tabla de Distribución Conjunta Completa). En lo que respecta a la Inteligencia Artificial, su uso es diverso, siendo la más común, en la aplicación a sistemas expertos

El teorema es ampliamente usado en la teoría de la probabilidad, según subyace a todos los sistemas modernos de inferencia probabilística. El teorema es derivado de la fórmula de probabilidad condicional y permite establecer la probabilidad a *posteriori* de una variable Y, dado un conjunto de eventos X:

$$P^*(Y) = P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (\text{A3.1})$$

Una Red Bayesiana es un par, donde es un grafo acíclico dirigido (GAD) tal que los nodos representan las variables del problema  $X=\{X1, X2, X3, \dots, Xn\}$  y los arcos representan las dependencias probabilísticas, y  $P=\{p(X1|pa(X1)), \dots, p(Xn|pa(Xn))\}$  es un conjunto de n distribuciones de probabilidad condicionada, una para cada variable, siendo  $pa(Xi)$  el conjunto de padres del nodo en el grafo  $D$ . Donde además el cálculo de la probabilidad conjunta del problema se obtiene mediante el producto de los elementos de tal forma que:

$$P(x) = \prod_{i=1}^n p(x_i | pa(X_i)) \quad (\text{A3.2})$$

La red bayesiana de tipo causal o red de diagnóstico es la que realiza su inferencia basándose en evidencias (consecuencias) representadas por síntomas frecuentes y determinantes de un hecho (causa). Su representación gráfica está dada por un grafo en el que, los nodos padre representan una causa y los nodos hijos, vienen a ser las consecuencias; por tanto, la inferencia en una Red Bayesiana de diagnóstico se realiza de abajo hacia arriba (de los nodos hijo hacia los nodos padre).

Para un sistema experto de diagnóstico, por ejemplo, el modelado se una red bayesiana que la represente, podría ser, por ejemplo, una red binaria que encapsule la representatividad de la presencia o ausencia de un síntoma o enfermedad (figura A3-11). En la siguiente imagen se observa el modelado de una red de diagnóstico (Tomado de [1]), los nodos hijo (síntomas) de color naranja están presentes, lo que indica que los cambios en los nodos padre (enfermedades) respectivos se calcularán.

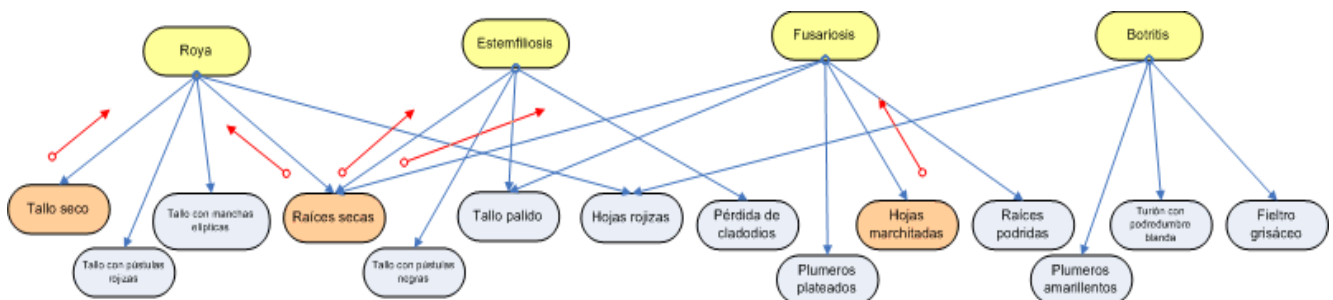


Figura A3-11. [19] Ejemplo de una red usando el algoritmo de Bayes

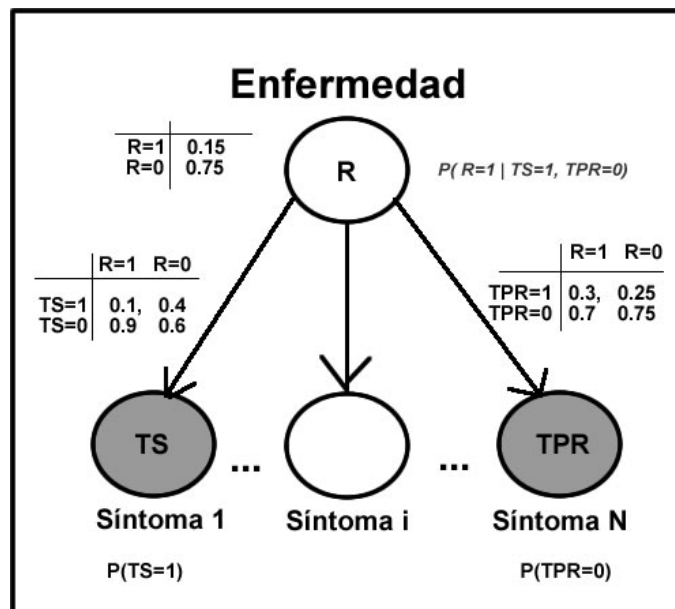


La construcción de una red de éste tipo requiere de un estudio estadístico previo, para determinar las enfermedades y síntomas relacionados, que están presentes en el ambiente a implantarse, esto implica, lograr estimaciones de qué enfermedades están presentes en dicho entorno, y qué síntomas están asociados a cada una de ellas y en qué porcentaje. Obviamente los síntomas con alto grado de *sensibilidad* y especificidad (se verán en la sección 7.2) con respecto a una enfermedad, serán considerados como síntomas verdaderos de dicha enfermedad.

Una Red de Bayes consiste de un conjunto de nodos y de un conjunto de arcos dirigidos entre estos nodos.

- Los arcos reflejan relaciones causa-efecto dentro del dominio. Estos efectos normalmente no son determinísticos (ej.: enfermedad - síntoma).
- La “fuerza” de un efecto es modelada como una probabilidad.
- La base matemática está dada por el Teorema de Bayes.

Una Red Bayesiana es un grafo dirigido acíclico en donde cada nodo representa una variable aleatoria y los arcos entre los nodos representan dependencias probabilísticas entre las variables. Cada nodo contiene los estados de la variable que representa y una tabla de probabilidad condicional (CPT). La CPT (figura A3-12) de un nodo contiene las probabilidades de que el nodo esté en un estado específico dados los estados de sus padres.



**TS**

Síntoma 1

$P(TS=1)$

...

**Síntoma i**

...

**TPR**

Síntoma N

$P(TPR=0)$

Figura A3-12. Tabla de probabilidad condicional en una red de bayes.

Un nodo representa una variable aleatoria discreta con un número finito de estados o una variable aleatoria continua (distribución Gaussiana).

- Si un nodo no tiene padres, el nodo contendrá una tabla de probabilidad marginal. Si el nodo es una variable dependiente, el nodo contendrá una tabla de probabilidad condicional (CPT).
- Si el nodo tiene padres, contiene una tabla de probabilidad condicional (CPT).
- Si el nodo es discreto, cada celda en la CPT de un nodo contiene una probabilidad condicional para cada estado en que pueda estar el nodo dada una configuración específica de los estados de sus padres.
- Si el nodo es continuo, la CPT contiene una media y una varianza para cada configuración de los estados de sus padres discretos (una si no hay padres discretos) y un coeficiente de regresión por cada padre continuo.

Cómo obtener los valores de las probabilidades?

- Conocimiento de expertos en el dominio
- Estudios estadísticos
- Derivados analíticamente
- Aprenderlos a partir de los datos

Las redes de Bayes son tal vez uno de los sistemas de clasificación más utilizados en el área médica y de investigación científica.

### A3-5 ÁRBOL DE DECISIONES

[20] La técnica de árbol de decisiones es también utilizada como un sistema de clasificación lineal, El árbol es una excelente ayuda para la elección entre varios cursos de acción. Proveen una estructura sumamente efectiva dentro de la cual estimar cuales son las opciones e investigar las posibles consecuencias de seleccionar cada una de ellas. También ayudan a construir una imagen balanceada de los riesgos y recompensas asociados con cada posible curso de acción.

Los árboles de decisión son uno de los métodos de *aprendizaje inductivo* más usados, en los arboles el aprendizaje se sigue por estas opciones:

**Hipótesis de aprendizaje inductivo:** cualquier hipótesis encontrada que clasifique un número suficientemente grande de ejemplos de entrenamiento clasificará otros ejemplos no observados.

- **Razonamiento deductivo:** partiendo de unas premisas se llega necesariamente a una conclusión. No aporta información nueva.
- **Razonamiento abductivo:** partiendo del conocimiento de unos efectos (síntomas) se llega a la causa (enfermedad).

Se trata de aproximar una función desconocida a partir de ejemplos positivos y negativos de esa función. Esos ejemplos serán en realidad pares  $\langle x, f(x) \rangle$ , donde  $x$  es el valor de entrada y  $f(x)$  el valor de la función aplicada a  $x$ . Dado un conjunto de ejemplos de  $f$ , la inducción consiste en obtener una función  $h$  que aproxime  $f$ . A esta función  $h$  se la denomina hipótesis.

Para comenzar a dibujar un árbol de decisión debemos escribir cuál es la decisión que necesitamos tomar, y seguimos las siguientes normas:

**Nodo de decisión:** Indica que una decisión necesita tomarse en ese punto del proceso. Está representado por un cuadrado



**Nodo de probabilidad:** Indica que en ese punto del proceso ocurre un evento aleatorio (estado de la naturaleza) Está representado por un círculo.



**Rama:** Nos muestra los distintos posibles caminos que se pueden emprender dado que tomamos una decisión u ocurre algún evento aleatorio



### **Como se construye un árbol de decisiones.**

Dibujaremos un recuadro para representar esto en la parte izquierda de una página grande de papel. Desde este recuadro se deben dibujar líneas hacia la derecha para cada posible solución, y escribir cuál es la solución sobre cada línea. Se debe mantener las líneas lo más apartadas posibles para poder expandir tanto como se pueda el esquema.

Al final de cada línea se debe estimar cuál puede ser el resultado. Si este resultado es incierto, se puede dibujar un pequeño círculo. Si el resultado es otra decisión que necesita ser tomada, se debe dibujar otro recuadro. Los recuadros representan decisiones, y los círculos representan resultados inciertos. Se debe escribir la decisión o el causante arriba de los cuadros o círculos. Si se completa la solución al final de la línea, se puede dejar en blanco.

Comenzando por los recuadros de una nueva decisión en el diagrama, dibujar líneas que salgan representando las opciones que podemos seleccionar. Desde los círculos se deben dibujar líneas que representen las posibles consecuencias. Nuevamente se debe hacer una pequeña inscripción sobre las líneas que digan que significan. Seguir realizando esto hasta que tengamos dibujado tantas consecuencias y decisiones como sea posible (ver figura A3-13).

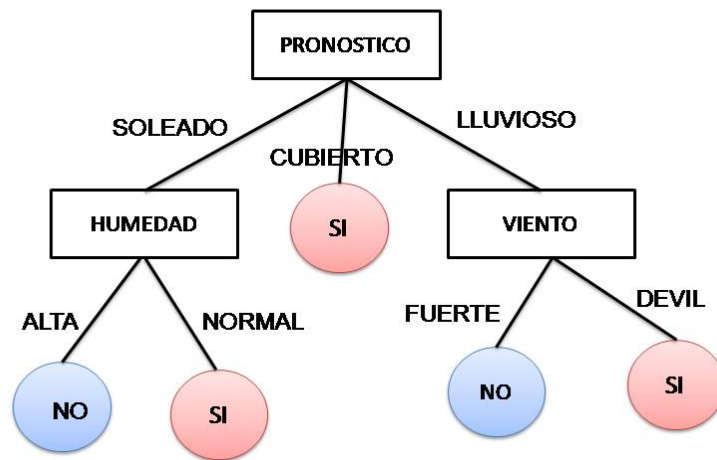


Figura A3-13. Ejemplo de un árbol de decisiones.

Significado del árbol, por las ramas del SI:

Pronostico = soleado /\ humedad = normal

Pronostico = cubierto

Pronostico = lluvioso /\ viento = débil

***Entonces SI puede Jugar.***

Los atributos están formados por parejas del tipo <atributo, valor> (ver figura A3-13), como por ejemplo <humedad, alta> o <humedad, normal> y a estos valores se les puede dar un peso o valor numérico si se desea, cuando se utiliza la técnica de árboles de decisión para clasificación se requiere un sistema de aprendizaje supervisado, la función de clasificación tiene valores de salida discretos, y a ser posible booleanos. Los patrones de ejemplo pueden contener errores, tanto por ruido presente en los ejemplos (errores en los valores) como errores en la clasificación de los ejemplos del conjunto de entrenamiento (recordemos que es un aprendizaje supervisado) Los ejemplos pueden contener atributos sin valor.

## A3-6 RESUMEN

En el anexo 4 se resaltan algunas contribuciones de desarrollos de la comunidad en materia de clasificación y detección automática de calcificaciones, en este capítulo vimos las diferentes técnicas que más se utilizan en la clasificación de patrones en imágenes medicas digitales, observamos las dos modalidades en el diseño de un clasificador: las técnicas supervisadas o no supervisadas, y como en el proceso de entrenamiento se requiere de una base de conocimiento del especialista para que el clasificador aprenda a reconocer y tener la mayor aproximación. Hemos hecho un resumen muy rápido sin profundizar sobre todo en la teoría de redes neuronales RNA en la cual hay muchos métodos y técnicas y queremos comentar que esta tal vez es en la que más se ha trabajado en sistemas expertos en reconocimiento de calcificaciones y tumores de cáncer de mama. En el siguiente capítulo nos dedicaremos a la parte experimental y diseño del proyecto concluyendo con los resultados obtenidos en este trabajo.

### A3-7 REFERENCIAS

- [1] Calot, tesis: "reconocimiento de patrones en imágenes medicas basado en sistemas inteligentes", Universidad de buenos aires, 2008.
- [2] R.C. González y R.E. Woods, *Digital Imagen Processing*, Addison Wesley, 3ª Edición, 2008. (cap. 12)
- [3] Vélez et al., *Visión por Computador*, Ed. Dyckinson – Serv. Publ. URJC, 2003. (cap. 5)
- [4] A. de la Escalera, *Visión por computador: Fundamentos y métodos*, Pearson-Prentice Hall, 2001. (caps. 4 y 5)
- [5] M. Sonka et al., *Image Processing, Analysis, and Machine Vision*, PWS Publishing, 1999. (cap. 7)
- [6] L.G. Shapiro y G.C. Stockman, *Computer Vision*, Prentice Hall, 2001. (cap. 4)
- [7] D.A. Forsyth y J. Ponce. *Computer Vision. A Modern Approach*, Pearson, 2003. (cap. 22)
- [8] Duda, Hart, Stork. *Pattern Classifications*, 2d. Wiley. 2000
- [9] José Augusto Moreno, clasificación de mamografías usando la función de base radial de rango tipo M, Instituto técnico nacional, Tesis para optar MSc. México, 2007.
- [10] Nikola K. Kazabov, "foundations of neural networks, fuzzy systems and knowledge engineering". The MIT press. 1998 paginas: 251-259
- [11] Mario Gómez, "teoría y simulación de redes neuronales artificiales". Universidad politécnica de valencia. 1999.
- [12] Pedro I. Vinuela, Ines M. Galvan, "redes neuronales artificiales, un enfoque practico", Pearson-Prentice hall, 2004. Pag:1-14
- [13] [http://en.wikipedia.org/wiki/Hopfield\\_net](http://en.wikipedia.org/wiki/Hopfield_net)
- [14] Universidad politécnica de cataluna,
- [15] <http://eddyalfaro.galeon.com/geneticos.html>
- [16] John Koza: <http://www.genetic-programming.com/johnkoza.html>
- [17] N. G. Londoño, Universidad Nacional de Colombia, escuela de estadística, Medellín. 2006.
- [18] [http://upload.wikimedia.org/wikipedia/commons/5/58/Evolutionary\\_algorithm.svg](http://upload.wikimedia.org/wikipedia/commons/5/58/Evolutionary_algorithm.svg)

[19] Jorge Valverde, Universidad nacional d Trujillo, Peru <http://jc-info.blogspot.com/2009/07/teorema-de-bayes-y-redes-bayesianas-de.html>

[20] Juan A. Botia Blaya, “aprendizaje computacional”, Departamento de ingeniería de la información y las telecomunicaciones, Universidad de Murcia, [juanbot@um.es](mailto:juanbot@um.es)

## ANEXO 4

### A4-1 TRABAJOS DESARROLLADOS PREVIOS – ESTADO DEL ARTE

La comunidad científica ha desarrollado desde 1985 a la fecha muchos trabajos orientados al procesamiento de imágenes de mamografía en diferentes aspectos de identificación de carcinomas, el gran interés del desarrollo de sistemas de ayuda automatizados CAD (Computer Aided Detections) y CADx (Computer Aided Diagnosis) [7].

Podemos clasificar los trabajos de investigación desarrollados por la comunidad de acuerdo por las áreas de procesamiento (ver tabla A4-1): segmentación, extracción de características, selección de características, clasificación. También acorde con el tipo de fenómeno de búsqueda: masas y calcificaciones.

<i>AUTOR</i>	<i>AÑO</i>	<i>METODO</i>	<i>uC</i>	<i>CI</i>	<i>TITULO</i>
D. Betal. Roberts	1997	MM		X	Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology
Zhao, M. Shridhar	1992	MM	X		Morphology on detection of calcifications in mammograms
Fuhr. O. treiber	2003	LoG		X	Cluster oriented detections of microcalcifications in simulated low-dose mammography
Papadopoulos, Fotiadis	2002	TX	X		An automated detection of calcifications system based on a hybrid neural networks classifier
El-naqa, Y. Yang.	2002	SVM		X	Support vector machine approach for detection of microcalcifications.
Bazzani. Bevilacqua.	2000	SVM	X	X	Automatic detection of cluster microcalcifications in digital mammograms using SVM classifier
Bocchi. Coppini.	2004	FR, RN	X	X	Detection of single and clustered microcalcifications in mammograms using fractal models and neural networks.
Nishikawa.	1993	DI		X	CAD of clustered microcalcifications
Ricketts.	1992	DI		X	CAD using clustered.

Tabla A4-1. [1, pag 14]. Algunas publicaciones de trabajos sobre el tema de procesamiento de imágenes de mamografía en diferentes técnicas de procesamiento.



## **METODOS:**

FR: Fractales            MM: matemática morfológica            LoG: laplaciano gaussiano  
AD: adaptativo        DI: diferencia de imágenes            DoG: diferencia de gaussianos  
TX: textural            RN: redes neuronales            SVM: support vector machine  
uC: técnica orientada a identificar microcalcificaciones  
Cl: técnica orientada a identificar agrupamiento de calcificaciones

Acorde con procedimientos encontramos las siguientes publicaciones en cada área del Procesamiento de las imágenes de mamografía [7]:

## **SEGMENTACIÓN.**

- F. Lefebvre, H. Benali, R. Gilles, E. Kahn, R. Di Paola, A Fractal Approach to the Segmentation of Microcalcifications in Digital Mammograms, Medical Physics, Vol. 22, No. 4, 1995, pp. 381-390
- W. Qian, M. Kallergi, L.P. Clarke, H.-D. Li, P. Venugopal, D. Song, R.A. Clark, Tree Structured Wavelet Transform Segmentation of Microcalcifications in Digital Mammography, Medical Physics, Vol. 22, No. 8, 1995, pp. 1247-1254
- H. Li, K.J.R. Liu, S.-C.B. Lo, Fractal Modeling and Segmentation for the Enhancement of Microcalcifications in Digital Mammograms, IEEE Transactions on Medical Imaging, Vol. 16, No. 6, 1997, pp. 785-798
- M.A. Kupinski, M.L. Giger, Automated Seeded Lesion Segmentation on Digital Mammograms, IEEE Transactions on Medical Imaging, Vol. 17, No. 4, 1998, pp. 510-517
- M.A. Gavrielides, J.Y. Lo, R. Vargas-Voracek, C.E. Floyd Jr., Segmentation of Suspicious Clustered Microcalcifications in Mammograms, Medical Physics, Vol. 27, No. 1, 2000, pp. 13-22
- G.M. te Brake, N. Karssemeijer, Segmentation of Suspicious Densities in Digital Mammograms, Medical Physics, Vol. 28, No. 2, 2001, pp. 259-266
- H. Li, Y. Wang, K.J.R. Liu, S.-C.B. Lo, M.T. Freedman, Computerized Radiographic Mass Detection - Part I: Lesion Site Selection by Morphological Enhancement and Contextual Segmentation, IEEE Transactions on Medical Imaging, Vol. 20, No. 4, 2001, pp. 289-301

### **EXTRACCIÓN DE CARACTERÍSTICAS**

- C.J. D'Orsi, D.B. Kopans, Mammographic Feature Analysis, *Seminars in Roentgenology*, Vol. 28, No. 3, 1993, pp. 204-230
- A.P. Dhawan, Y. Chitre, C. Kaiser-Bonasso, M. Moskowitz, Analysis of Mammographic Microcalcifications Using Gray-Level Image Structure Features, *IEEE Transactions on Medical Imaging*, Vol. 15, No. 3, 1996, pp. 246-259
- H.-P. Chan, B. Sahiner, K.L. Lam, N. Petrick, M.A. Helvie, M.M. Goodsitt, D.D. Adler, Computerized Analysis of Mammographic Microcalcifications in Morphological and Texture Feature Spaces, *Medical Physics*, Vol. 25, No. 10, 1998, pp. 2007-2019
- B. Verma, J. Zakos, A Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy-Neural and Feature Extraction Techniques, *IEEE Transactions on Information Technology in Biomedicine*, Vol. 5, No. 1, 2001, pp. 46-54
- S. Timp, N. Karssemeijer, Interval Change Analysis to Improve Computer Aided Detection in Mammography, *Medical Image Analysis*, Vol. 10, No. 1, 2006, pp. 82-95

### **SELECCIÓN DE CARACTERÍSTICAS**

- B. Zheng, Y.-H. Chang, X.-H. Wang, W.F. Good, D. Gur, Feature Selection for Computerized Mass Detection in Digitized Mammograms by Using a Genetic Algorithm, *Academic Radiology*, Vol. 6, No. 6, 1999, pp. 327-33
- Z. Huo, M.L. Giger, D.E. Wolverton, W. Zhong, S. Cumming, O.I. Olopade, Computerized Analysis of Mammographic Parenchymal Patterns for Breast Cancer Risk Assessment: Feature Selection, *Medical Physics*, Vol. 27, No. 1, 2000, pp. 4-12
- B. Verma, P. Zhang, A Novel Neural-Genetic Algorithm to Find the Most Significant Combination of Features in Digital Mammograms, *Applied Soft Computing Journal*, Vol. 7, No. 2, 2007, pp. 612-625

### **CLASIFICACIÓN**

- Y. Wu, K. Doi, M.L. Giger, R.M. Nishikawa, Computerized Detection of Clustered Microcalcifications in Digital Mammograms: Applications of Artificial Neural Networks, *Medical Physics*, Vol. 19, No. 3, 1992, pp. 555-560
- H.-P. Chan, S.-C.B. Lo, B. Sahiner, Kwok Leung Lam, M.A. Helvie, Computer-Aided Detection of Mammographic Microcalcifications: Pattern Recognition with an Artificial Neural Network, *Medical Physics*, Vol. 22, No. 10, 1995, pp. 1555-1567

- Y. Jiang, R.M. Nishikawa, D.E. Wolverton, C.E. Metz, M.L. Giger, R.A. Schmidt, C.J. Vyborny, K. Doi, Malignant and Benign Clustered Microcalcifications: Automated Feature Analysis and Classification, *Radiology*, Vol. 198, No. 3, 1996, pp. 671-678
- H.-P. Chan, B. Sahiner, N. Patrick, M.A. Helvie, K.L. Lam, D.D. Adler, M.M. Goodsitt, Computerized Classification of Malignant and Benign Microcalcifications on Mammograms: Texture Analysis Using an Artificial Neural Network, *Physics in Medicine and Biology*, Vol. 42, No. 3, 1997, pp. 549-567
- W.J.H. Veldkamp, N. Karssemeijer, J.D.M. Otten, J.H.C.L. Hendriks, Automated Classification of Clustered Microcalcifications into Malignant and Benign Types, *Medical Physics*, Vol. 27, No. 11, 2000, pp. 2600-2608
- Papadopoulos, D.I. Fotiadis, A. Likas, An Automatic Microcalcification Detection System Based on a Hybrid Neural Network Classifier, *Artificial Intelligence in Medicine*, Vol. 25, No. 2, 2002, pp. 149-167
- L. Wei, Y. Yang, R.M. Nishikawa, Y. Jiang, A Study on Several Machine-Learning Methods for Classification of Malignant and Benign Clustered Microcalcifications, *IEEE Transactions on Medical Imaging*, Vol. 24, No. 2, 2005, pp. 371-380
- P. Zhang, B. Verma, K. Kumar, Neural vs. Statistical Classifier in Conjunction with Genetic Algorithm Based Feature Selection, *Pattern Recognition Letters*, Vol. 26, No. 7, 2005, pp. 909-919

## A4-2 GRUPOS DE INVESTIGACIÓN EN LA TECNICA

Son muchos los grupos de investigación que están trabajando en el desarrollo y profundización de las diferentes técnicas dentro el ciclo del procesamiento, clasificación y diagnóstico de carcinomas mediante sistemas expertos CAD, no solo las Universidades si no también la industria, para no ir muy lejos el grupo **BioIngenium** de la Universidad Nacional de Colombia – sede Bogotá, tiene un equipo de investigadores desarrollando y publicando [8] nuevos trabajos en el tema, como ayuda al entrenamiento de los estudiantes de Radiología, la figura A4-1 muestra el proceso que ha trabajado:

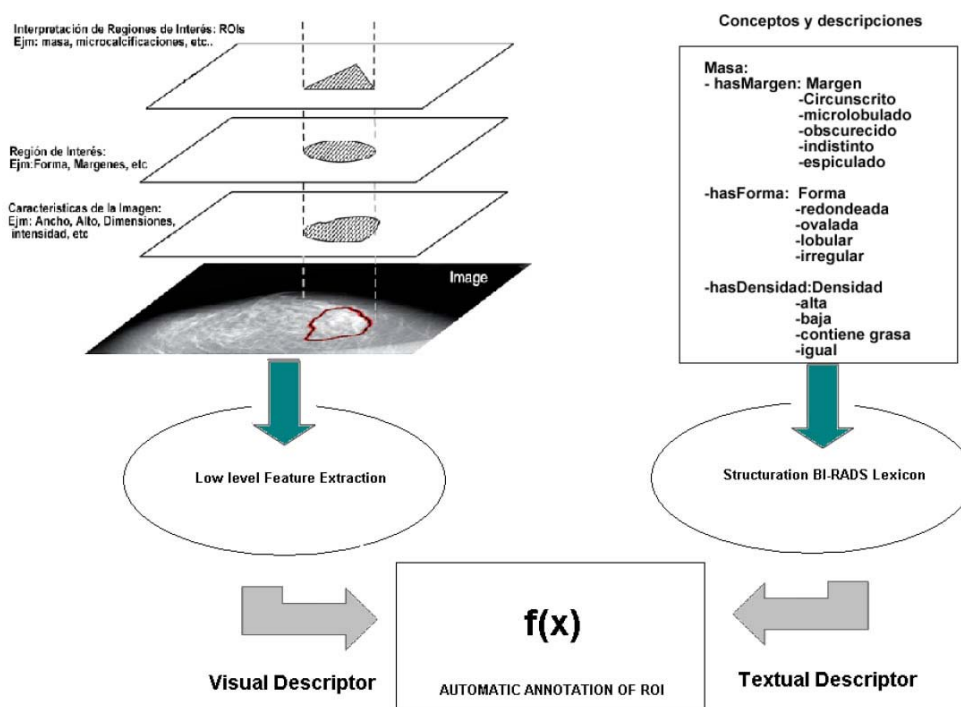


Figura A4-1 [8] diagrama del sistema de investigación del grupo de BioIngenium

Otras universidades de reconocimiento mundial por sus publicaciones y grupos de investigación son:

- Australia: Universidad Finders; school of science: computers engineering and and mathematics.
- Australia: Universidad de Westem; school os electrical, electronic and computer engineering.

- Brasil: Universidad de Sao Paulo; laboratorio de procesamiento de imágenes médicas y odontológicas.
- Canada: Universidad de Alberta; Biomedical engineering group. Department of electrical and computer engineering.
- Canada: Universidad de Calgary; Biomedical engineer group, Schulich School of engineering.
- Canada: Universidad de Toronto; Department of medical biophysics, faculty of medicine.
- Croatia: Universidad de Zagreb; communications lab, faculty of electrical engineering.
- Alemania: Instituto Fraunhofer Mevis
- Alemania: Universidad de Bremen; center of complex systems and visualizations
- Alemania: clarite comprehensive center
- Israel: Jerusalem school of engineering.
- Italia: Universidad de Palermo; departamento de física relativa.
- Japón: Universidad de Gifu; Fujita Lab
- Nueva Zelanda: Universidad de Cantebury; computational imaging group
- España: Universidad de Girona; Image analysis lab, computer visions group, department of electronic.
- España: Universidad de Extremadura; clasificación de patrones y análisis de imágenes.

Además de las muchas universidades las cuales tiene grupos de investigación existen muchas asociaciones científicas que promueven la investigación y el trabajo en tecnologías de sistemas expertos, áreas de **Computer Vision** y diseños de CAD en diferentes áreas de la medicina: MRI, CT, PET, Ultrasound. La asociación **SPIE** [9] “society’s interdisciplinary approach to supporting the science for optical light”. En febrero del 2011 se realizara el evento SPIE – Medical imaging -. El cual ofrece más de 1000 presentaciones sobre avances de la ciencia en temas como CAD, de igual forma son muchos los fabricantes que han desarrollado sistemas que están en producción y han recibido aprobación por la **FDA** (Federal Drougs and food administrations) organismo del gobierno americano quien vela por estandarizar y dar un alto grado de certificación en los sistemas médicos.

# ANEXO 5

## 1 – PROGRAMA MATLAB

```
%*****
%      Ing. Gerardo Ramírez de la E.
%      Pontificia Universidad Javeriana
%      Facultad de Ingeniería, departamento de Electrónica
%      Trabajo de tesis:
%      "DETECCIÓN DE CARCINOMAS EN LAS IMÁGENES DE MAMOGRAFIA
%      MEDIANTE TECNICAS DE PROCESAMIENTO DE IMÁGENES Y
%      RECONOCIMIENTO DE PATRONES"
%      trabajo para optar el título de MSc.
%      Bogotá, Enero 2011.
%      mail: gramirez3333@gmail.com, móvil: (57) 320.4996050
%=====
%      VARIABLES DEL PROGRAMA
%
% realce          nivel del umbral de la imagen de salida tophat
% umbral_         umbral de binarización a imagen salida de realce
% r_tophat        radio a utilizar en detector de borde top hat
% im1             imagen de entrada original
% imag            imagen recortada, manualmente (proceso previo)
% im001           imagen escalarizada a 8 bits
% imediana        imagen de salida al filtro mediana 3x3
% itoh            imagen de salida segmentada por Top hat
% oehc            imagen de salida después del realce
% obin_tophat     imagen binarizada
% roi             imagen para localizar objetos y seleccionarlos
% [Lb Ne]         matriz en donde guarda elementos localizados
%                Lb : Matriz
%                Ne : número de elementos encontrados
% propied         vector de propiedades de morfológicas
% roi_b           región de interés para ventana benignas
%===== variables benignas =====
% [Lbb Nebb] = bwlabel(roi_b);      matriz en donde guarda elementos probablemente benignos
% propied_b= regionprops(Lbb,'all'); vector de características en región roi_b
% count_roi1b                       contador de objetos encontrados en imagen binarizada
% sroib=find([propied_b.Area]<77)    define áreas <77
% count_roi2b                       contador de objetos menores a <77 píxeles (malignos)
% total_roib                         total objetos posiblemente benignos;
% roi_b                             región de interés para ventana benignas
% [Lbb Nebb] = bwlabel(roi_b);      matriz en donde guarda elementos probablemente benignos
% propied_b= regionprops(Lbb,'all'); vector de características en región roi_b
% count_roi1b                       contador de objetos encontrados en imagen binarizada
% sroib=find([propied_b.Area]<77)    define áreas <77
% count_roi2b                       contador de objetos menores a <77 píxeles (malignos)
% total_roib                         total objetos posiblemente benignos;
```

```

%===== variables malignas =====
% [Lbm Nebm] = bwlabel(roi_m);      matriz en donde guarda elementos probablemente malignos
% propied_m= regionprops(Lbm,'all'); vector de características en region roi_m
% count_roi1m                       contador de objetos encontrados en imagen binarizada
% sroib=find([propied_m.Area]>77)   define areas >77
% count_roi2m                       contador de objetos mayores a >77 pixeles (malignos)
% total_roiM                         total objetos posiblemente total_roiB total_roiMte malignos;
% roi_m                             region de interes para ventana malignos
% [Lbm Nebm] = bwlabel(roi_b);      matriz en donde guarda elementos probablemente malignos
% propied_m= regionprops(Lbm,'all'); vector de características en region roi_m
% count_roi1m                       contador de objetos encontrados en imagen binarizada
% sroim=find([propied_m.Area]>77)   define areas > 77
% count_roi2m                       contador de objetos mayores a >77 pixeles (malignos)
% total_roiM                         total objetos posiblemente malignos;
% roimin                             imagen nueva con objetos entre 0-77 pixeles.
% [Lbmin Nebmin] = bwlabel(roi_min); matriz en donde guarda elementos
%                                   probablemente malignos muy pequenos
% propied_min= regionprops(Lbmin,'all'); vector de características
% count_roi1min                     contador de objetos menores a encontrar
% sroimin                           vector de objetos con area <5 sroimin
%*****VECTOR DE CARACTERISTICAS*****
%                                   SENALES BENINAS:
% Area (n) =                         arealb(n)
% Perimetro (n) =                    perimlb(n)
% ecenrticidad (n) =                 eccentriclb(n)
% Areaconvexa (n) =                  areconvexlb(n)
% Solidez (n) =                      solidlb(n)
%                                   SENALES MALIGNAS:
% Area (n) =                         arealm(n)
% Perimetro (n) =                    perimlm(n)
% ecenrticidad (n) =                 eccentriclm(n)
% Areaconvexa (n) =                  areaconvexlm(n)
% solidez (n) =                      solidlm(n)
%===== Limites de decision en el clasificador =====
% limite_solb                         limite de decision de solidez en el clasificador B
% limite_execb                         limite de decision de excentricidad en el clasificador B
% limite_                               limite de decision de solidez en el clasificador M
% limite_execm                         limite de decision de excentricidad en el clasificador
% total_benignos                       total numero de benignos clasificados
% roi_bp                               region de interes para calculos de benignos a clasificar
% benigno_exec(n)                      vector de excentricidad en benignos a clasificar
% benigno_sol(n)                       vector de solidez en benignos a clasificar
% benigno_insertidumbre(n)            vector de conteo de incertidumbre
% total_insertidumbreB                total de objetos inciertos en ventana benignos
% roi_bm                               region de interes para calculos de malignos a clasificar
% maligno_exec(n)                     vector de excentricidad en malignos a clasificar
% maligno_sol(n)                       vector de solidez en malignos a clasificar
% maligno_insertidumbre(n)            vector de conteo de incertidumbre
% total_insertidumbreM                total de objetos inciertos en ventana malignos

```

```

%*****
close all;
clear all;
clc;
%=====parametros=====
realce = 0.5 ;
umbralbin_tophat = 0.25 ;
r_tophat = 14 ;

limite_execb = 0.95 ;
limite_solb = 0.745 ;

limite_execm = 0.25;
limite_solm = 0.935 ;

%=====LECTURA DE LA IMAGEN=====
im1 = imread('mdb024.pgm');
imag = imread('mdb024-r.pgm');
%*****
%
%                PREPROCESAMIENTO
%*****
%===Convertir a escala de 8 bits - 256 niveles de gris
im001 = uint8 (imag);
%=====
%preprocesamiento de la imagen: filtro mediaño
%entrada fml
%Salida f1
%=====
fml = im001;
f1 = medfilt2 (fml , [3 3] , 'symmetric');
imediana=f1;
%=====Figure preprocesamiento 2=====
figure;
subplot (1,3,1), imshow(im1); title(' imagen original');
subplot (1,3,2), imshow(imag); title(' imagen recortada');
subplot (1,3,3), imshow(imediana); title ('imagen filtrada');
%*****
%                EXTRACCION DE CARACTERISTICAS
%*****
%Opcion 1. - tophat
% entrada itohl
% Salida otohl
%=====
itoh = imediana ;
se= strel('disk', r_tophat) ;
otoh = imtophat(itoh , se);
% procesamiento de la imagen: enhanced
iehc= otoh ;
oehc= imadjust(iehc,[0 realce],[0 1]);
% Binarizacion
inbin_tophat = oehc ;
obin_tophat=im2bw(inbin_tophat,umbralbin_tophat);

```



```

%=====Figure 2 =====
figure;
subplot (1,3,1), imshow(otoh); title('SEGMENTACIÓN Top hat');
subplot (1,3,2), imshow(oehc); title('realce iamgen');
subplot (1,3,3), imshow(obin_tophat); title ('Imagen Binarizada');

%***** SELECCION DE CARACTERISTICAS *****
% rango imagenes Binignas: 77 a 385 pixeles
% rango imagenes Malignas: 5 a 77 pixeles
% imágenes mayores a 386 pixeles se consideran posibles tumores
%=====
roi_b = obin_tophat ;
roi_m = obin_tophat ;

%=====
% paso 1)
% procedimiento para quitar objetos menores a <77 pixeles
% para dejara solo obejtos probables de ser Benignos
%=====
figure; imshow(roi_b); title('Imagen a discriminar ');
%*****
% Etiquetar elementos conectados
% Neb: numero de elementos
% Lb: Matrix en donde guarda objetos encontrados
%=====
[Lbb Nebb] = bwlabel(roi_b);
%=====
% Calcular propiedades de los objetos de la imagen
%=====
propied_b= regionprops(Lbb , 'all');
%=====

hold on

count_roi1b=0; % contador de objetos encontrados
%===== Graficar las 'cajas' de frontera de los objetos =====
% encuentra objetos y los enciera en una ventana de color verde (green)
for n=1:size(propied_b,1)
    rectangle('Position',propied_b(n).BoundingBox,'EdgeColor','g','LineWidth',2);
    count_roi1b=count_roi1b+1;
end
%=====Crear ventana Benignas=====
pause (3)
%===== Buscar áreas < 77 =====
sroi_b=find([propied_b.Area]<77);
%===== Marcar áreas < 77 =====
count_roi2b=0;
%==== encuentra objetos y los enciera en una ventana de color rojo (red)
for n=1:size(sroi_b,2)
    rectangle('Position',propied_b(sroi_b(n)).BoundingBox,'EdgeColor','r','LineWidth',2);
    count_roi2b=count_roi2b+1;
end

```

```

pause (3)
%=====Eliminar áreas menores a < 77=====
% elimina los objetos en color rojo.
for n=1:size(sroib,2)
    d=round(propied_b(sroib(n)).BoundingBox);
    roi_b(d(2):d(2)+d(4),d(1):d(1)+d(3))=0; % elimina los objetos
end
%===== total objetos posiblemente benignos; total_roid=====
total_roid=count_roidb-count_roid2b;
figure, imshow(roi_b);
title(['Objetos encontrados probablemente benignos: ',num2str(total_roid)])
%*****

%*****
% paso 2)
%     procedimiento para quitar objetos mayores a >77 pixeles
%     para dejra solo obejtos probables de ser Malignos
%     en el rango de 0 a 77 pixeles
%=====
figure; imshow(roi_m); title('(3)Imagen a discriminar ');
%*****
% Etiquetar elementos conectados
% Neb: numero de elementos
% Lb: Matrix en donde guarda objetos encontrados
%=====
[Lbm Nebm] = bwlabel(roi_m);
%=====
% Calcular propiedades de los objetos de la imagen
%=====
propied_m= regionprops(Lbm , 'all');
%=====

hold on

count_roidm=0;
%=====Graficar las 'cajas' de frontera de los objetos =====
% los encierra en color verde
for n=1:size(propied_m,1)
    rectangle('Position',propied_m(n).BoundingBox,'EdgeColor','g','LineWidth',2)
    count_roidm=count_roidm+1;
end

pause (3)
% ===== Buscar áreas > a 77 =====
sroim=find([propied_m.Area]>77);
%===== Marcar áreas > a 77 =====
count_roid2m=0;
for n=1:size(sroim,2)
    rectangle('Position',propied_m(sroim(n)).BoundingBox,'EdgeColor','r','LineWidth',2)
    count_roid2m=count_roid2m+1;
end

```

```

pause (3)
%=====Eliminar áreas mayores a 77 =====
for n=1:size(sroim,2)
    d=round(propied_m(sroim(n)).BoundingBox);
    roi_m(d(2):d(2)+d(4),d(1):d(1)+d(3))=0;
end
total_roim=count_roi1m-count_roi2m;
figure, imshow(roi_m);
title(['(8)Total Objetos encontrados: ',num2str(total_roim)])

%*****
% paso 3)
% procedimiento para quitar objetos menores a 5 pixeles
% para quitar objetos que son ruido y fuera de escala
%=====
roi_min = roi_m ;
figure; imshow(roi_min); title('(3)Imagen a discriminar ');
%*****
% Etiquetar elementos conectados
% Neb: numero de elementos
% Lb: Matrix en donde guarda objetos encontrados
%=====
[Lbmin Nebmin] = bwlabel(roi_min);
%=====
% Calcular propiedades de los objetos de la imagen
%=====
propied_min= regionprops(Lbmin , 'all');
%=====

hold on

count_roi1min=0; % contador de objetos encontrados
%===== Graficar las 'cajas' de frontera de los objetos =====
for n=1:size(propied_min,1)
    rectangle('Position',propied_min(n).BoundingBox,'EdgeColor','g','LineWidth',2)
    count_roi1min=count_roi1min+1;
end

pause (3)
% ===== Buscar áreas < a 5 =====
sroimin=find([propied_min.Area]<5);
%===== Marcar áreas < 5 =====
count_roi2min=0;
for n=1:size(sroimin,2)
    rectangle('Position',propied_min(sroimin(n)).BoundingBox,'EdgeColor','r','LineWidth',2)
    count_roi2min=count_roi2min+1;
end

```

```

pause (3)
%===== Eliminar áreas menores a 5 =====
for n=1:size(sroi_min,2)
    d=round(propied_min(sroi_min(n)).BoundingBox);
    roi_min(d(2):d(2)+d(4),d(1):d(1)+d(3))=0;
end
total_roimin=count_roi1min-count_roi2min;
figure, imshow(roi_min);
title(['Objetos encontrados probablemente
malignos:',num2str(total_roim)]);

%=====
%                               imagenes de salida seleccionadas
% Benignas:                       roi_b
% total objetos probables : total_roib
% Malignas:                       roi_min
% total objetos probables : total_roimin
%
%*****VECTOR DE CARACTERISTICAS*****
%                               SENALES BENINAS:
% Area (n) =                       areaIb(n)
% Perimetro (n) =                   perimIb(n)
% ecentricidad (n) =                eccentrIb(n)
% Areaconvexa (n) =                 areconvexIb(n)
% Solidez (n) =                     solidIb(n)
%
%                               SENALES MALIGNAS:
% Area (n) =                       areaIm(n)
% Perimetro (n) =                   perimIm(n)
% ecentricidad (n) =                eccentrIm(n)
% Areaconvexa (n) =                 areaconvexIm(n)
% solidez (n) =                     solidIm(n)
%=====Resultado=====
figure;
subplot (1,2,1), imshow(roi_b); title(['benignos probables =', num2str(total_roib)]);
subplot (1,2,2), imshow(roi_min); title(['malignos probables =', num2str(total_roimin)]);

%===== VECTOR DE CARACTERISTICAS =====
%
% 1) Objetos probablemente benignos
%=====
roi_bp = roi_b ;
[Lbbp Nebbp] = bwlabel(roi_bp);
propied_bp= regionprops(Lbbp, 'all');
for n=1:size(propied_bp,1)
    areaIb(n)    = propied_bp(n).Area ;
    perimIb(n)   = propied_bp(n).Perimeter ;
    eccentrIb(n) = propied_bp(n).Eccentricity ;
    areaconvexIb(n) = propied_bp(n).ConvexArea ;
    solidIb(n)  = propied_bp(n).Solidity ;
end
end

```

```

%=====
% 2) Objetos probablemente Malignos
%=====
roi_mp = roi_min ;
[Lbmp Nebmp] = bwlabel(roi_mp);
propied_mp= regionprops(Lbmp,'all');
for n=1:size(propied_mp,1)
    areaIm(n)      = propied_mp(n).Area ;
    perimIm(n)     = propied_mp(n).Perimeter ;
    eccentrIm(n)   = propied_mp(n).Eccentricity ;
    areaconvexIm(n) = propied_mp(n).ConvexArea ;
    solidIm(n)     = propied_mp(n).Solidity ;
end

%===== CLASIFICADOR =====
% clasificador implementado por arbol de decisiones
% dos ramas: solido, circular
% cada una con dos nodos de decision:
% solidez y excentricidad
% limite_solb      limite de decision de solidez en el clasificador B
% limite_exeCb     limite de decision de excentricidad en el clasificador
B
% limite_solm      limite de decision de solidez en el clasificador M
% limite_exeCm     limite de decision de excentricidad en el clasificador
%=====
% calculos del lado de etiquetas benignos
%=====
total_benignos = 0;
total_insertidumbreB = 0;
roi_bp = roi_b ;
[Lbbp Nebbp] = bwlabel(roi_bp);
propied_bp= regionprops(Lbbp,'all');
for n=1:size(propied_bp,1)
    eccentrIb(n) = propied_bp(n).Eccentricity ;
    if eccentrIb(n) <= limite_exeCb
        benigno_exec(n)=1 ;
    else
        benigno_exec(n)=0 ;
    end
    solidIb(n) = propied_bp(n).Solidity ;
    if solidIb(n) > limite_solb
        benigno_sol(n) =1 ;
    else
        benigno_sol(n)= 0 ;
    end
    if ((benigno_exec(n)>=1) && (benigno_sol(n)>=1)) ;
        benigno_true(n) =1 ;
        total_benignos = total_benignos +1 ;
    else
        benigno_true(n)=0 ;
    end
    if ((benigno_exec(n)<=0) && (benigno_sol(n)<=0));
        benigno_insertidumbre(n) = 1 ;
        total_insertidumbreB = total_insertidumbreB + 1 ;
    end
end

```

```

else
    benigno_inserdidumbre(n) = 0 ;
end
end

%=====
% calculos del lado de etiquetas malignos
%=====
total_malignos = 0;
total_inserdidumbreM = 0;
roi_mp = roi_min ;
[Lbmp Nebmp] = bwlabel(roi_mp);
propied_mp= regionprops(Lbmp,'all');
for n=1:size(propied_mp,1)
    eccentrIm(n) = propied_mp(n).Eccentricity ;
    if eccentrIm(n) >= limite_execm
        maligno_exec(n)=1 ;
    else
        maligno_exec(n)=0 ;
    end
    solidIm(n) = propied_mp(n).Solidity ;
    if solidIm(n) < limite_solm
        maligno_sol(n) =1 ;
    else
        maligno_sol(n)=0 ;
    end
    if ((maligno_exec(n)>=1) && (maligno_sol(n)>=1)) ;
        maligno_true(n)=1 ;
        total_malignos = total_malignos + 1 ;
    else
        maligno_true(n)=0 ;
    end
    if ((maligno_exec(n)<=0) && (maligno_sol(n)<=0)) ;
        maligno_inserdidumbre(n)=1 ;
        total_inserdidumbreM = total_inserdidumbreM + 1;
    else
        maligno_inserdidumbre(n)=0 ;
    end
end
end

figure;
subplot (1,3,1), imshow(imediana);
title ('Imagen de entrada');

subplot (1,3,2), imshow(roi_b);
title(['B(+)=', num2str(total_benignos), ' B(l)=' ,num2str(total_inserdidumbreB)]);

subplot (1,3,3), imshow(roi_min);
title(['M(+)=', num2str(total_malignos), ' M(l)=' ,num2str(total_inserdidumbreB)]);

```

## ANEXO 4

### *Ref. mias MIAS (organizarlo en columnas)*

1	2	3	4	5	6	7
=====						
mdb001	G	CIRC	B	535	425	197
mdb002	G	CIRC	B	522	280	69
mdb003	D	NORM				
mdb004	D	NORM				
mdb005	F	CIRC	B	477	133	30
mdb005	F	CIRC	B	500	168	26
mdb006	F	NORM				
mdb007	G	NORM				
mdb008	G	NORM				
mdb009	F	NORM				
mdb010	F	CIRC	B	525	425	33
mdb011	F	NORM				
mdb012	F	CIRC	B	471	458	40
mdb013	G	MISC	B	667	365	31
mdb014	G	NORM				
mdb015	G	CIRC	B	595	864	68
mdb016	G	NORM				
mdb017	G	CIRC	B	547	573	48
mdb018	G	NORM				
mdb019	G	CIRC	B	653	477	49
mdb020	G	NORM				
mdb021	G	CIRC	B	493	125	49
mdb022	G	NORM				
mdb023	G	CIRC	M	538	681	29
mdb024	G	NORM				
mdb025	F	CIRC	B	674	443	79
mdb026	F	NORM				
mdb027	F	NORM				
mdb028	F	CIRC	M	338	314	56
mdb029	G	NORM				
mdb030	G	MISC	B	322	676	43
mdb031	G	NORM				
mdb032	G	MISC	B	388	742	66
mdb033	D	NORM				
mdb034	D	NORM				
mdb035	D	NORM				
mdb036	D	NORM				
mdb037	D	NORM				
mdb038	D	NORM				
mdb039	D	NORM				
mdb040	D	NORM				
mdb041	G	NORM				
mdb042	G	NORM				
mdb043	G	NORM				
mdb044	G	NORM				
mdb045	G	NORM				
mdb046	G	NORM				
mdb047	G	NORM				
mdb048	G	NORM				
mdb049	G	NORM				
mdb050	G	NORM				
mdb051	G	NORM				
mdb052	G	NORM				
mdb053	D	NORM				
mdb054	D	NORM				
mdb055	G	NORM				
mdb056	G	NORM				
mdb057	D	NORM				
mdb058	D	MISC	M	318	359	27

mdb059 F CIRC B  
 mdb060 F NORM  
 mdb061 D NORM  
 mdb062 D NORM  
 mdb063 D MISC B 546 463 33  
 mdb064 D NORM  
 mdb065 D NORM  
 mdb066 D NORM  
 mdb067 D NORM  
 mdb068 D NORM  
 mdb069 F CIRC B 462 406 44  
 mdb070 F NORM  
 mdb071 G NORM  
 mdb072 G ASYM M 266 517 28  
 mdb073 G NORM  
 mdb074 G NORM  
 mdb075 F ASYM M 468 717 23  
 mdb076 F NORM  
 mdb077 F NORM  
 mdb078 F NORM  
 mdb079 F NORM  
 mdb080 F CIRC B 432 149 20  
 mdb081 G ASYM B 492 473 131  
 mdb082 G NORM  
 mdb083 G ASYM B 544 194 38  
 mdb084 G NORM  
 mdb085 G NORM  
 mdb086 G NORM  
 mdb087 F NORM  
 mdb088 F NORM  
 mdb089 G NORM  
 mdb090 G ASYM M 510 547 49  
 mdb091 F CIRC B 680 494 20  
 mdb092 F ASYM M 423 662 43  
 mdb093 G NORM  
 mdb094 G NORM  
 mdb095 F ASYM M 466 517 29  
 mdb096 F NORM  
 mdb097 F ASYM B 612 297 34  
 mdb098 F NORM  
 mdb099 D ASYM B 714 340 23  
 mdb100 D NORM  
 mdb101 D NORM  
 mdb102 D ASYM M 415 460 38  
 mdb103 D NORM  
 mdb104 D ASYM B 357 365 50  
 mdb105 D ASYM M 516 279 98  
 mdb106 D NORM  
 mdb107 D ASYM B 600 621 111  
 mdb108 D NORM  
 mdb109 D NORM  
 mdb110 D ASYM M 190 427 51  
 mdb111 D ASYM M 505 575 107  
 mdb112 D NORM  
 mdb113 G NORM  
 mdb114 G NORM  
 mdb115 G ARCH M 461 532 117  
 mdb116 G NORM  
 mdb117 G ARCH M 480 576 84  
 mdb118 G NORM  
 mdb119 G NORM  
 mdb120 G ARCH M 423 262 79  
 mdb121 G ARCH B 492 434 87  
 mdb122 G NORM  
 mdb123 G NORM  
 mdb124 G ARCH M 366 620 33  
 mdb125 D ARCH M 700 552 60  
 mdb126 D ARCH B 191 549 23  
 mdb127 G ARCH B 523 551 48



mdb128 G NORM  
 mdb129 D NORM  
 mdb130 D ARCH M 220 552 28  
 mdb131 F NORM  
 mdb132 F CIRC B 252 788 52  
 mdb132 F CIRC B 335 766 18  
 mdb133 F NORM  
 mdb134 F MISC M 469 728 49  
 mdb135 F NORM  
 mdb136 F NORM  
 mdb137 D NORM  
 mdb138 D NORM  
 mdb139 F NORM  
 mdb140 F NORM  
 mdb141 F CIRC M 470 759 29  
 mdb142 F CIRC B 347 636 26  
 mdb143 F NORM  
 mdb144 F MISC B 233 994 29  
 mdb144 F MISC M 313 540 27  
 mdb145 D SPIC B 669 543 49  
 mdb146 D NORM  
 mdb147 F NORM  
 mdb148 F SPIC M 326 607 174  
 mdb149 F NORM  
 mdb150 F ARCH B 351 661 62  
 mdb151 F NORM  
 mdb152 F ARCH B 675 486 48  
 mdb153 F NORM  
 mdb154 F NORM  
 mdb155 F ARCH M 448 480 95  
 mdb156 F NORM  
 mdb157 F NORM  
 mdb158 F ARCH M 540 565 88  
 mdb159 F NORM  
 mdb160 F ARCH B 536 519 61  
 mdb161 D NORM  
 mdb162 D NORM  
 mdb163 D ARCH B 391 365 50  
 mdb164 D NORM  
 mdb165 D ARCH B 537 490 42  
 mdb166 D NORM  
 mdb167 F ARCH B 574 657 35  
 mdb168 F NORM  
 mdb169 D NORM  
 mdb170 D ARCH M 489 480 82  
 mdb171 D ARCH M 462 627 62  
 mdb172 D NORM  
 mdb173 F NORM  
 mdb174 F NORM  
 mdb175 G SPIC B 592 670 33  
 mdb176 G NORM  
 mdb177 G NORM  
 mdb178 G SPIC M 492 600 70  
 mdb179 D SPIC M 600 514 67  
 mdb180 D NORM  
 mdb181 G SPIC M 519 362 54  
 mdb182 G NORM  
 mdb183 F NORM  
 mdb184 F SPIC M 352 624 114  
 mdb185 G NORM  
 mdb186 G SPIC M 403 524 47  
 mdb187 G NORM  
 mdb188 G SPIC B 406 617 61  
 mdb189 G NORM  
 mdb190 G SPIC B 512 621 31  
 mdb191 G SPIC B 594 516 41  
 mdb192 G NORM  
 mdb193 D SPIC B 399 563 132  
 mdb194 D NORM

mdb195 F SPIC B 725 129 26  
 mdb196 F NORM  
 mdb197 D NORM  
 mdb198 D SPIC B 568 612 93  
 mdb199 D SPIC B 641 177 31  
 mdb200 D NORM  
 mdb201 D NORM  
 mdb202 D SPIC M 557 772 37  
 mdb203 F NORM  
 mdb204 F SPIC B 336 399 21  
 mdb205 F NORM  
 mdb206 F SPIC M 368 200 17  
 mdb207 D SPIC B 571 564 19  
 mdb208 D NORM  
 mdb209 G CALC M 647 503 87  
 mdb210 G NORM  
 mdb211 G CALC M 680 327 13  
 mdb212 G NORM  
 mdb213 G CALC M 547 520 45  
 mdb214 G NORM  
 mdb215 D NORM  
 mdb216 D CALC M \*NOTE 3\*  
 mdb217 G NORM  
 mdb218 G CALC B 519 629 8  
 mdb219 G CALC B 546 756 29  
 mdb220 G NORM  
 mdb221 D NORM  
 mdb222 D CALC B 398 427 17  
 mdb223 D CALC B 523 482 29  
 mdb223 D CALC B 591 529 6  
 mdb224 D NORM  
 mdb225 D NORM  
 mdb226 D CALC B 287 610 7  
 mdb226 D CALC B 329 550 25  
 mdb226 D CALC B 531 721 8  
 mdb227 G CALC B 504 467 9  
 mdb228 G NORM  
 mdb229 F NORM  
 mdb230 F NORM  
 mdb231 F CALC M 603 538 44  
 mdb232 F NORM  
 mdb233 G CALC M \*NOTE 3\*  
 mdb234 G NORM  
 mdb235 D NORM  
 mdb236 D CALC B 276 824 14  
 mdb237 F NORM  
 mdb238 F CALC M 522 553 17  
 mdb239 D CALC M 645 755 40  
 mdb239 D CALC M 567 808 25  
 mdb240 D CALC B 643 614 23  
 mdb241 D CALC M 453 678 38  
 mdb242 D NORM  
 mdb243 D NORM  
 mdb244 D CIRC B 466 567 52  
 mdb245 F CALC M \*NOTE 3\*  
 mdb246 F NORM  
 mdb247 F NORM  
 mdb248 F CALC B 378 601 10  
 mdb249 D CALC M 544 508 48  
 mdb249 D CALC M 575 639 64  
 mdb250 D NORM  
 mdb251 F NORM  
 mdb252 F CALC B 439 367 23  
 mdb253 D CALC M 733 564 28  
 mdb254 D NORM  
 mdb255 F NORM  
 mdb256 F CALC M 400 484 37  
 mdb257 D NORM  
 mdb258 D NORM

mdb259 D NORM  
mdb260 D NORM  
mdb261 D NORM  
mdb262 D NORM  
mdb263 G NORM  
mdb264 G MISC M 596 431 36  
mdb265 G MISC M 593 498 60  
mdb266 G NORM  
mdb267 F MISC M 793 481 56  
mdb268 F NORM  
mdb269 G NORM  
mdb270 G CIRC M 356 945 72  
mdb271 F MISC M 784 270 68  
mdb272 F NORM  
mdb273 F NORM  
mdb274 F MISC M 127 505 123  
mdb275 G NORM  
mdb276 G NORM  
mdb277 G NORM  
mdb278 G NORM  
mdb279 G NORM  
mdb280 G NORM  
mdb281 D NORM  
mdb282 D NORM  
mdb283 D NORM  
mdb284 D NORM  
mdb285 D NORM  
mdb286 D NORM  
mdb287 D NORM  
mdb288 D NORM  
mdb289 D NORM  
mdb290 D CIRC B 337 353 45  
mdb291 G NORM  
mdb292 G NORM  
mdb293 F NORM  
mdb294 F NORM  
mdb295 D NORM  
mdb296 D NORM  
mdb297 F NORM  
mdb298 F NORM  
mdb299 F NORM  
mdb300 F NORM  
mdb301 F NORM  
mdb302 F NORM  
mdb303 F NORM  
mdb304 F NORM  
mdb305 F NORM  
mdb306 F NORM  
mdb307 F NORM  
mdb308 F NORM  
mdb309 F NORM  
mdb310 F NORM  
mdb311 F NORM  
mdb312 F MISC B 240 263 20  
mdb313 F NORM  
mdb314 F MISC B 518 191 39  
mdb315 D CIRC B 516 447 93  
mdb316 D NORM  
mdb317 D NORM  
mdb318 D NORM  
mdb319 D NORM  
mdb320 D NORM  
mdb321 D NORM  
mdb322 D NORM