



Universidad Autónoma de Querétaro
Facultad de Ingeniería

Clasificación de Microcalcificaciones utilizando Descriptores Morfológicos

Tesis
Que como parte de los requisitos para obtener el
grado de

**Maestro en Ciencias en Instrumentación y
Control Automático**

Presenta

Ing. Rut Audelo Valverde

Dirigido por

Dr. Alberto Pastrana Palma

C.U. Santiago de Querétaro, Querétaro
Junio de 2014



Universidad Autónoma de Querétaro
Facultad de Ingeniería
Maestría en Ciencias Instrumentación y Control Automático

CLASIFICACIÓN DE MICROCALCIFICACIONES UTILIZANDO DESCRIPTORES
MORFOLÓGICOS

TESIS

Que como parte de los requisitos para obtener el grado de
Maestro en Ciencias en Instrumentación y Control Automático

Presenta:

Rut Audelo Valverde

Dirigido por:

Dr. Alberto Pastrana Palma

SINODALES

Dr. Alberto Pastrana Palma
Presidente

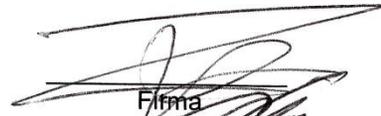
Dr. Damián Vargas Vázquez
Secretario

Dr. Edgar Alejandro Rivas Araiza
Vocal

Dr. Jesús Carlos Pedraza Ortega
Suplente

Dr. Aurora Femat Díaz
Suplente

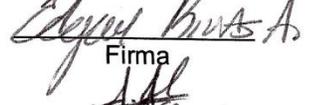
Dr. Aurelio Domínguez González
Director de la Facultad



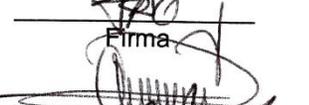
Firma



Firma



Firma



Firma



Firma



Dr. Irineo Torres Pacheco
Director de Investigación y
Posgrado

RESUMEN

El trabajo de tesis presentado a continuación tuvo como objetivo la clasificación de microcalcificaciones encontradas en la base de datos MIAS (*Mammographic Image Analysis Society*) en 2 grupos de microcalcificaciones, los cuales fueron: benignas y malignas, basados en la clasificación de microcalcificaciones de Le Gal (1984). Para trabajar con los descriptores de forma fue necesaria la binarización de las microcalcificaciones de manera individual utilizando el análisis del histograma y el método de Otsu. Las imágenes de la base de datos fueron analizadas a través de los diferentes descriptores morfológicos los cuales fueron: momentos de orden 0, 1, 2 y 3, excentricidad, signatura, compacidad, área y ramificaciones; cuya información pasa al clasificador utilizado en este trabajo red neuronal artificial con el perceptrón multicapa de retropropagación. El número de imágenes (muestras) utilizadas como pruebas experimentales para validar la hipótesis se obtuvieron a través de la determinación del número mínimo de observaciones en investigación, obviando las estimaciones de la varianza de datos. Los resultados obtenidos fueron comparados con diagnósticos realizados por médicos radiólogos de los cuales se obtuvieron los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, para posteriormente analizar el rendimiento del algoritmo de clasificación propuesto mediante la sensibilidad y especificidad.

(Palabras clave: microcalcificación, clasificación, descriptores, red neural)

SUMMARY

The main objective of this thesis is the classification of microcalcifications found in MIAS (Mammographic Image Analysis Society) data base into 2 groups, which were: benignant and malignant, based on Le Gal's classification presented on 1984. First of all, it was necessary to binarize each microcalcification independently using histogram analysis and Otsu's method. The resultant image was analyzed through the next morphological descriptors: Hu's Moments, Eccentricity, Signature, Compactness, Area and Ramifications; this information becomes in the input vector of a Neural Network (Backpropagation) as a automated classifier. The number of samples used as experimental tests to validate the hypothesis proposed in this work was obtained using the determination of minimum samples of experimental observations, obviating the data's variance estimates. The results were compare with Medical diagnosis made by Radiologist, which true positives, true negatives, false positives and false negatives are obtained to analyzed the performance of this classification algorithm using sensitivity and specificity.

(Key words: microcalcification, classification, descriptors, neural network)

**A mi familia
que siempre me ha dado su apoyo.**

Saber no es suficiente;
tenemos que aplicarlo.
Tener voluntad no es suficiente:
tenemos que implementarla.
(Goethe)

AGRADECIMIENTOS

Quisiera agradecer a Dios por haberme concedido la salud y la fortaleza para que estos 2 años de trabajo rindiera fruto. Quiero agradecer también a mis padres Ana Isabel Valverde Noriega y José Candelario Audelo Martínez, que a pesar de la distancia estuvieron apoyándome, motivándome y esperando que siempre tuviera éxito en cuanto emprendiera; a mis hermanos Noemí Audelo Valverde, Isaac Audelo Valverde, Josué Audelo Valverde y Daniel Audelo Valverde les doy gracias porque yo siempre he sentido que debo superarme a mi misma debido a ustedes, son mi mayor motivación.

A la Universidad Autónoma de Querétaro y al cuerpo académico del posgrado en Control y Automatización Automático de la Facultad de Ingeniería por todos los conocimientos compartidos para la realización de esta tesis. A mi asesor el Dr. Alberto Pastrana Palma que con su asesoramiento y guía el desarrollo de esta tesis fue posible.

Quiero agradecer al Dr. Manuel Carrizo de la Unidad de Especialidades Médicas dedicada a la Detección y Diagnóstico de Cáncer de Mama de la ciudad de Santiago de Querétaro por su apoyo y orientación brindados.

ÍNDICE

	Página
Resumen	i
Summary	ii
Dedicatorias	iii
Agradecimientos	iv
Índice	v
Índice de cuadros	vii
Índice de figuras	x
I. INTRODUCCIÓN	1
Hipótesis	3
Objetivo general	3
Objetivos específicos	3
II. REVISIÓN DE LITERATURA	4
Estándar de clasificación BIRADS	4
Estándar de clasificación por tipos según Le Gal	7
Mamografía digital	9
Procesamiento digital de imágenes	9
Descriptores morfológicos	11
Métodos de clasificación	11
Redes neuronales artificiales	12
Detección y clasificación de lesiones en mamografías	14
Detección de grupos de microcalcificaciones	17
Descripción y clasificación de microcalcificaciones	17
III. METODOLOGÍA	25

Base de datos MIAS	26
MatLAB	26
Obtención de las microcalcificaciones y extracción individual	26
Binarización de la microcalcificación	29
Esqueleto de la microcalcificación	36
Contorno de la microcalcificación	37
Área y compacidad	38
Elongación	39
Momentos invariantes	40
Signatura y momento	41
Descripción del esqueleto	42
Clasificación	43
Red multicapa de retropropagación	44
Determinación de la muestra para la validación del algoritmo	47
Evaluación del algoritmo	49
IV. RESULTADOS Y DISCUSIÓN	51
Clasificación benigno/maligno según MIAS	51
Clasificación benigno/maligno según Le Gal	57
Clasificación por tipos Le Gal	62
Conclusiones	63
LITERATURA CITADA	64
APENDICE	67

ÍNDICE DE CUADROS

Cuadro		Página
2.1	Clasificación BIRADS, American College of Radiology 2003.	4
2.2	Clasificación de microcalcificaciones a partir de las observaciones de Le Gal et al. (1984).	8
2.3	Tipo de procesamiento realizado por las redes neuronales artificiales. Ramírez y Chacón (2011).	13
2.4	Resultados de la clasificación de lesiones por Khademi et al. (2009).	14
2.5	Éxito en la clasificación de imágenes mamográficas usando la familia Daubechies, Haar y Biorthogonal. Rajkumar y Raju (2011).	16
2.6	Lista de características aplicadas para la clasificación individual de las microcalcificaciones. Wroblewska, et al (2003).	19
2.7	Combinación óptima de características para cada grupo de característica de textura. Karahaliou et al. (2008).	22
2.8	Desempeño de la clasificación de los grupos de características de manera individual y el esquema combinado. Karahaliou et al. (2008).	23
2.9.	Análisis de métodos de extracción y clasificación de microcalcificaciones de trabajos previos revisados.	24

3.1	Valores de Z_{α} para diferentes niveles de confianza (Lozano Rivas, 2011).	47
3.2	Valores de Z_{β} para diferentes niveles de poder estadístico (Lozano Rivas, 2011).	47
3.3	Número mínimo de observaciones para diferentes W , con un poder estadístico $(1 - \beta)$ del 80% (Lozano Rivas, 2011).	48
3.4	Comparación de diagnósticos y descripción de variables VP, VN, FP y FN.	49
4.1	Valores de descriptores para las microcalcificaciones con similitud morfológica (Tipo I Le Gal).	51
4.2	Valores de descriptores para las microcalcificaciones con similitud morfológica (Tipo II Le Gal).	52
4.3	Valores de descriptores para las microcalcificaciones con similitud morfológica (Tipo III Le Gal).	55
4.4	Valores de descriptores para las microcalcificaciones con similitud morfológica (Tipo IV Le Gal).	54
4.5	Valores de descriptores para las microcalcificaciones con similitud morfológica (Tipo V Le Gal).	55
4.6	Nomenclatura de los descriptores utilizados en los Cuadros 4.1 al 4.5.	56
4.7	Ejemplos de microcalcificaciones según su tipo (Clasificación Le Gal).	57
4.8	Ejemplos de microcalcificaciones binarizadas según su tipo (Clasificación Le Gal).	58
4.9	Resultados obtenidos usando la red que se muestra en la Figura 3.15.	59

4.10	Variables de evaluación obtenidas por la red neuronal 1 del diseño de la Figura 3.15.	59
4.11	Variables de evaluación obtenidas por la red neuronal 2 del diseño de la Figura 3.15.	59
4.12	Variables de evaluación obtenidas por la red neuronal 3 del diseño de la Figura 3.15.	59
4.13	Variables de evaluación obtenidas por la red neuronal 4 del diseño de la Figura 3.15.	60
4.14	Precisión obtenida para cada una de las redes neuronales.	61
A.1	Microcalcificaciones utilizadas de la base de datos MIAS para la validación del algoritmo.	66
A.2	Pesos 1 para la evaluación del diseño 1 de red neuronal (Capa entrada – capa oculta).	69
A.3	Pesos 1 para la evaluación del diseño 1 de red neuronal (Capa oculta – capa salida).	69
A.4	Pesos 2 para la evaluación del diseño 1 de red neuronal (Capa entrada – capa oculta).	70
A.5	Pesos 2 para la evaluación del diseño 1 de red neuronal (Capa oculta – capa salida).	70
A.6	Pesos 3 para la evaluación del diseño 1 de red neuronal (Capa entrada – capa oculta).	71
A.7	Pesos 3 para la evaluación del diseño 1 de red neuronal (Capa oculta – capa salida).	71
A.8	Pesos 4 para la evaluación del diseño 1 de red neuronal (Capa entrada – capa oculta).	72
A.9	Pesos 4 para la evaluación del diseño 1 de red neuronal (Capa oculta – capa salida).	72

ÍNDICE DE FIGURAS

Figura	Página
2.1 Tipo de microcalcificaciones que insinúan malignidad.	5
2.2 Patrones de distribución que se asocian a un proceso maligno. a) Distribución ductual y b) Distribución segmental. Heywang et al. (2001).	6
2.3 Tipo de Calcificaciones definitivamente benignas.	7
2.4 Clasificación de microcalcificaciones según Le Gal et al. (1984).	8
2.5 Cadena de procesamiento en Procesamiento Digital de Imágenes.	10
3.1 Diagrama de bloques del algoritmo de clasificación de microcalcificaciones.	24
3.2 Mamografía mdb233 de la base de datos MIAS.	27
3.3 Microcalcificación: mdb233.pgm 2 CALC M 623 591 5.	28
3.4 Histograma de la Microcalcificación: mdb233.pgm 2 CALC M 623 591 5.	29
3.5 Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 tras aplicar el oscurecimiento de la imagen.	30
3.6 Histograma de la Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 tras aplicar el oscurecimiento de la imagen.	30

3.7	Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 después de la umbralización mediante umbral duro.	31
3.8	Histograma de la Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 después de la umbralización mediante umbral duro.	32
3.9	Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 tras la expansión del histograma.	33
3.10	Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 binarizada.	34
3.11	Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 tras la expansión del histograma.	35
3.12	Esqueleto de la Microcalcificación: mdb233.pgm 2 CALC M 623 591 5.	36
3.13	Contorno de la Microcalcificación: mdb233.pgm 2 CALC M 623 591 5.	37
3.14	Vecindad 8 de un píxel.	42
3.15	Red multicapa de retropropagación con salidas: benigna y maligna.	45
3.16	Red multicapa de retropropagación con salidas: Tipo I, Tipo II, Tipo III, Tipo IV y Tipo V.	46
4.1	a) Microcalcificación mdb219.pgm 1 CALC B 548 258 10 y b) Microcalcificación mdb245.pgm 1 CALC M 593 772 10. (Nota: M = Maligno, B=Benigno).	51
4.2	a) Microcalcificación mdb222.pgm 2 CALC B 399 607 10 y b) Microcalcificación mdb245.pgm 2 CALC M 666 745 10. (Nota: M = Maligno, B=Benigno).	52
4.3	a) Microcalcificación mdb233.pgm 3 CALC M 673 837 5 y b) Microcalcificación mdb219.pgm 3 CALC B 544 273 5. (Nota: M = Maligno, B=Benigno).	53

- 4.4 a) Microcalcificación mdb233.pgm 4 CALC M 561 627 5 y b) Microcalcificación mdb226.pgm 4 CALC B 292 414 5. (Nota: M = Maligno, B=Benigno). 54
- 4.5 a) Microcalcificación mdb216.pgm 5 CALC M 382 510 5 y b) Microcalcificación mdb219.pgm 5 CALC B 531 281 5. (Nota: M = Maligno, B=Benigno). 55

I. INTRODUCCIÓN

De acuerdo con la Organización Mundial de la Salud (OMS), el cáncer más frecuente entre las mujeres es el de mama, que a nivel mundial representa 16% de todos los cánceres femeninos, se estima que cada año se detectan 1.38 millones de casos nuevos (OMS, 2012). El cáncer de mama es la patología oncológica más frecuente en la mujer occidental y la principal causa de muerte por cáncer femenino en Europa, Estados Unidos, Australia y ciertos países de América Latina (Ricci et al., 2006), por ejemplo, un estudio realizado por el Instituto Nacional de Cáncer (NCI por sus siglas en inglés) en Estados Unidos en el 2010 más de 207,090 nuevos casos fueron reportados y cerca de 39,840 muertes registradas (Smith et al., 2010). Con la introducción de programas de detección en 1988 los números de muertes han decaído, sin embargo, se predice que una de cada siete mujeres desarrollará cáncer de pecho en algún punto de su vida.

En México, en el 2011 la tasa de morbilidad hospitalaria por causa del cáncer de mama en mujeres fue del 29.6% con respecto a la totalidad de casos registrados de morbilidad hospitalaria por causa de algún tumor maligno y la coloca como la principal causa de morbilidad en mujeres, mientras que la distribución porcentual de defunción por tumor maligno en la mama en mujeres fue de 14.7% y es la cuarta causa de muerte en mujeres. En este contexto, adquiere especial importancia el estudio de las lesiones no palpables de mama y su probable asociación con cáncer, destacando las microcalcificaciones donde la mamografía constituye la principal herramienta diagnóstica (Ricci et al., 2006).

La detección temprana del cáncer de pecho es una estrategia fundamental para el diagnóstico, tratamiento y control oportuno de esta enfermedad ya que reduce su impacto entre la población. El método principal de detección es la radiografía de alta resolución en el pecho, llamada mamografía. El

proceso de identificar cáncer en una mamografía requiere de radiólogos con alta experiencia ya que es una tarea difícil, muy subjetiva y susceptible a errores. Dos indicadores importantes de cáncer que son evaluados en las mamografías son las masas y microcalcificaciones. Las masas son el mayor indicador, son objetos tridimensionales que pueden ser ubicados en la mamografía, pueden ser descritos por su localización, tamaño, forma, características de contorno, densidad, efecto en el tejido circundante y efectos en la piel, las masas pueden ser benignas o malignas.

Las microcalcificaciones son depósitos de calcio pequeños que van de 0.1mm a 1mm de diferentes formas y densidad (Engan et al., 2006), que se producen como resultado de secreciones y pueden tener una variedad de causas benignas, pero también puede ser un indicio de malignidad, entre el 30% al 50% del cáncer de mama en el mundo es diagnosticado debido a las microcalcificaciones. El detectar calcificaciones es de suma importancia ya que cerca del 30% - 40% de los carcinomas albergan microcalcificaciones y a estos carcinomas, gracias a las microcalcificaciones, son comúnmente descubiertos a una etapa temprana, frecuentemente en una etapa pre-invasiva. Sin embargo, existen un considerable número de cambios benignos que contienen microcalcificaciones.

El análisis meticuloso de las microcalcificaciones es el pre-requisito para la detección precisa de carcinomas, una mejor valoración posible del alcance del carcinoma y evitar un alto número de biopsias innecesarias por cambios benignos. El objetivo de este análisis es la categorización o clasificación de las microcalcificaciones en: altamente sospechosas de malignidad, definitivamente benigna e indeterminado; las características para distinguir entre microcalcificaciones benignas de las malignas son: forma, densidad, distribución y definición.

Los Diagnósticos Asistidos por Computadora (CAD por sus siglas en inglés) pueden ser utilizados para proveer una opinión alternativa al diagnóstico o enfocar la atención del radiólogo hacia un área de interés y ayudan de manera importante en dos aspectos: la identificación de áreas de interés y la clasificación de dichas áreas en cancerígenas (malignas) o no cancerígenas (benignas). El CAD para la clasificación de microcalcificaciones ha sido una tarea difícil debido a la similitud morfológica entre los grupos de microcalcificaciones.

Hipótesis

Mediante el uso de descriptores morfológicos y técnicas de clasificación supervisada se puede categorizar, en benignas y malignas, a las microcalcificaciones encontradas en mamografías con un 90% de sensibilidad.

Objetivo general

El objetivo de este proyecto de investigación es el de realizar un algoritmo que clasifique a las microcalcificaciones según su tipo utilizando descriptores morfológicos y métodos de clasificación supervisada.

Objetivos específicos

- Identificar las características de forma de las microcalcificaciones mediante la implementación de descriptores morfológicos.
- Implementar clasificador supervisado utilizando los descriptores morfológicos como vector de entrada.
- Evaluar la eficiencia del algoritmo basado en la sensibilidad y especificidad.

II. REVISIÓN DE LITERATURA

Estándar de clasificación BIRADS

BI-RADS (*Breast Imaging Reporting and Data Systems*) es un sistema desarrollado para permitir una mejor estandarización de reportes mamográficos y diagnóstico, fue publicado por el *American College of Radiology* y ha sido actualizado en 4 ediciones (1993, 1996, 1998, 2003) (Ricci et al., 2006), el diagnóstico final debe ser categorizado como se muestra en el Cuadro 2.1.

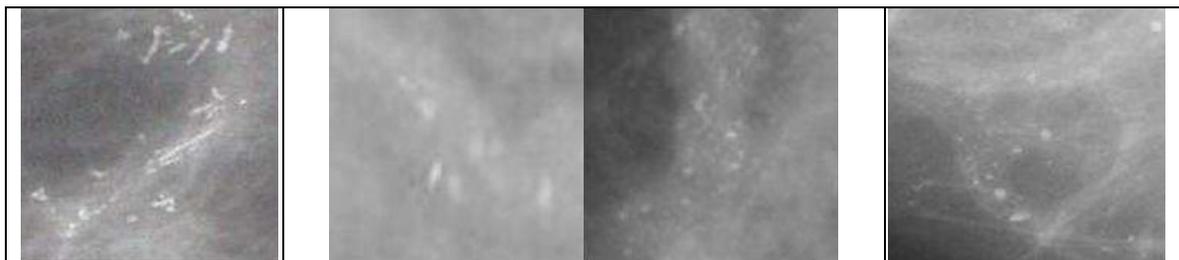
Cuadro 2.1. Clasificación BIRADS, *American College of Radiology* 2003.

Evaluación incompleta	
Categoría 0:	Se requiere evaluación por imágenes (Magnificación, focalización, ultrasonido, etc.)
Evaluación completa	
Categoría 1:	Negativa, mama normal.
Categoría 2:	Hallazgo benigno (fiboadenomas, lesiones grasas, linfonodos intramamarios, etc.).
Categoría 3:	Hallazgo probablemente benigno. Sugiere intervalo corto de seguimiento.
Categoría 4:	Anomalías sospechosas. Debe considerarse biopsia debido a riesgo de cáncer.
Categoría 4A:	Lesiones de baja probabilidad de ser malignas, las cuales sin embargo van a biopsia.
Categoría 4B:	Lesiones de sospecha intermedia de malignidad. La conducta dependerá de la correlación radio-histológica.
Categoría 4C:	Lesiones de sospecha intermedia de malignidad. La conducta dependerá de la correlación radio-histológica.
Categoría 5:	Altamente sugerente de malignidad.
Categoría 6:	Se denominan a los cánceres ya confirmados con biopsia, antes de su tratamiento definido (cirugía, radioterapia, quimioterapia, hormonoterapia).

Existen 3 categorías en las que las calcificaciones pueden ser clasificadas según el Colegio Americano de Radiología y BIRADS (Panda et al., 2009), las cuales son:

- a) Típicamente benignas.
- b) Indeterminadas.
- c) Altamente sospechosa de malignidad.

Las características técnicas establecidas para el análisis y la descripción de microcalcificaciones son su localización, forma individual y distribución. Las microcalcificaciones que indican malignidad están siempre intramamarias, calcificaciones lineales irregulares con ramificaciones, granuladas pleomórficas, microcalcificaciones amorfas o indiferenciadas y granulado fino si es combinado con un patrón de distribución sospechosa; de acuerdo a los patrones de distribución que se asocian a un proceso de malignidad nos encontramos a los de disposición ductual (Figura 2.2 a) y segmental (Figura 2.2 b). Las microcalcificaciones que indican malignidad son de tamaño menor a 0.5 mm y aparecen en grupo de 5 o más.



a) Ramificadas b) Amorfas o indiferenciadas c) Pleomórficas
Figura 2.1. Tipo de microcalcificaciones que insinúan malignidad.

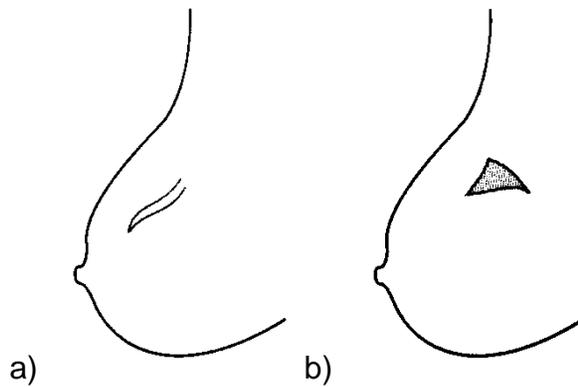


Figura 2.2. Patrones de distribución que se asocian a un proceso maligno. a) Distribución ductal y b) Distribución segmental. Heywang et al. (2001).

Las microcalcificaciones que tienen tamaño mayor a 1mm y las que se encuentran subcutáneas insinúan benignidad. También se consideran benignas todas aquellas calcificaciones redondeadas que tengan o no centro traslúcido, las calcificaciones semicirculares, en forma de cáscara de huevo, como palomita de maíz, puntiformes y de forma irregular de más de 2mm. Los patrones de distribución de las microcalcificaciones con proceso benigno son las siguientes: calcificaciones solitarias, difusas y simétricas.

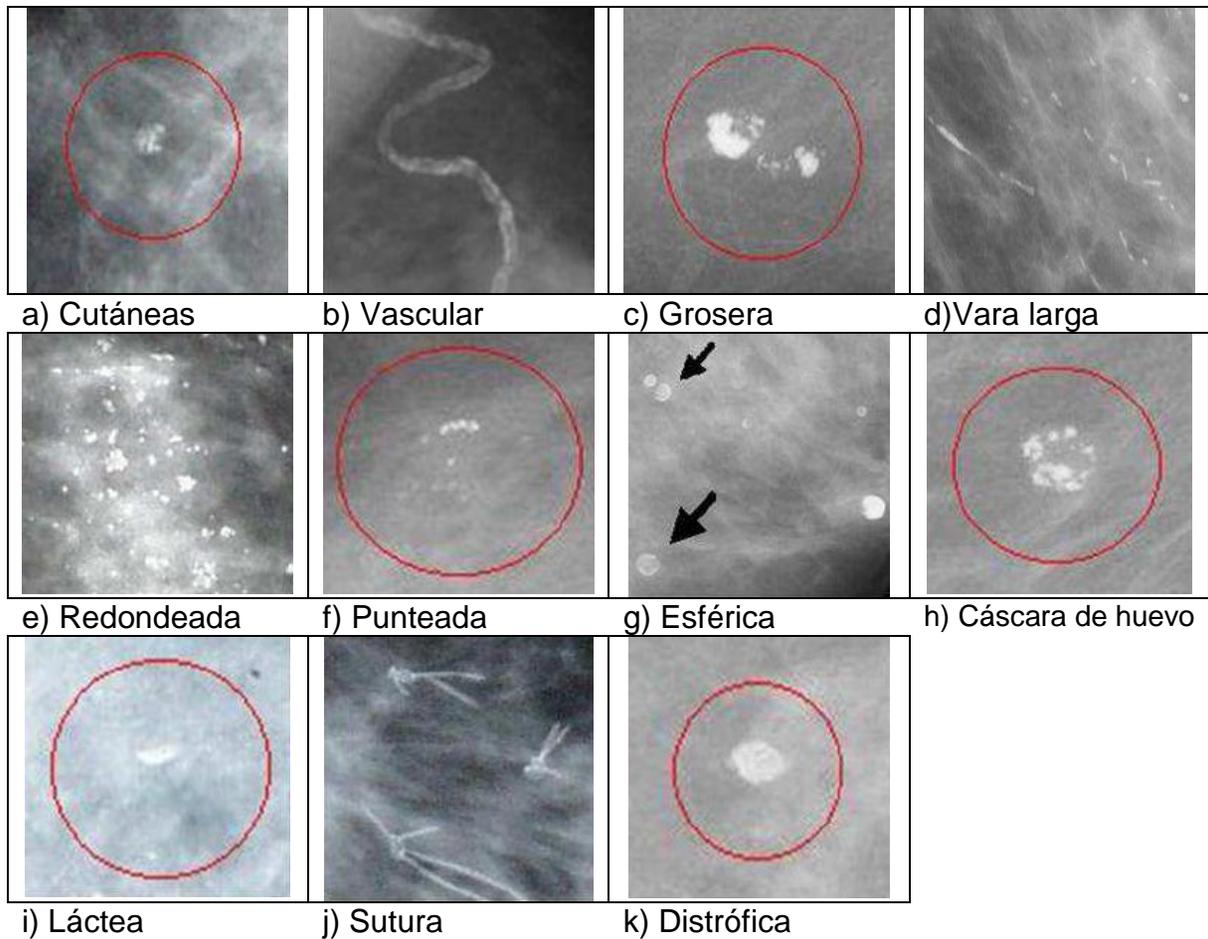


Figura 2.3. Tipo de Calcificaciones definitivamente benignas.

Estándar de clasificación por tipos según Le Gal

En 1984, Le Gals divide las microcalcificaciones en 5 tipos, determinando porcentajes de malignidad en cada caso (Cuadro 2.2 y Figura 2.4). Se asocian con mayor frecuencia a cáncer las microcalcificaciones descritas con forma vermicular, distribución lineal o en ramas de árbol sin hojas, y las de tamaño irregulares (Ricci et al., 2006).

Cuadro 2.2. Clasificación de microcalcificaciones a partir de las observaciones de Le Gal et al. (1984).

Tipo 1:	Microcalcificaciones anulares, redondeadas, de centro claro.	0% Malignidad.
Tipo 2:	Microcalcificaciones puntiformes regulares, redondeadas, con el centro y la superficie de la microcalcificación de idéntico tono cálido.	10% Malignidad.
Tipo 3:	Microcalcificaciones “en polvo”, muy finas sin poder precisar su forma ni su número, en el límite de la visibilidad.	19% Malignidad.
Tipo 4:	Microcalcificaciones puntiformes irregulares, poliédricas, en grano de sal.	29% Malignidad.
Tipo 5:	Microcalcificaciones vermiculares, alargadas, en bastoncillos (como árbol sin hojas).	72% Malignidad.

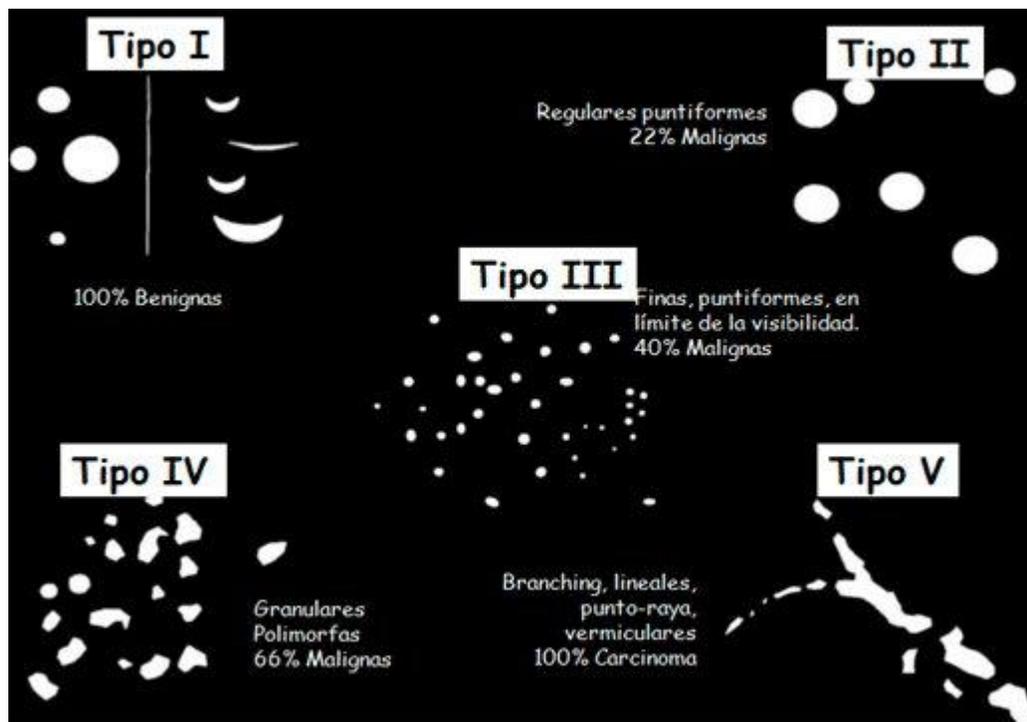


Figura 2.4. Clasificación de microcalcificaciones según Le Gal et al. (1984).

La morfología de las calcificaciones es considerada para ser el indicador más importante en la diferenciación entre benigno y maligno. Según de Lafontan et al. (1994) otras características que son estadísticamente significativas para sospecha de cáncer son ocho: 1) Número de microcalcificaciones por cm^2 , 2) Diámetro medio de las microcalcificaciones, 3) Número total de microcalcificaciones, 4) Irregularidad en la densidad de las microcalcificaciones, 5) Irregularidad en los tamaños, 6) Formas lineales o ramificadas, 7) Alta densidad, 8) Calcificaciones vermiculares.

El Diagnóstico Asistido por Computadora (CAD) para la clasificación de microcalcificaciones ha sido una difícil tarea debido a la similitud morfológica entre los grupos de microcalcificaciones, además de su baja distinción con el tejido mamario.

Mamografía digital

La mamografía digital constituye una técnica computarizada que es capaz de retirar los artefactos de la mamografía convencional, constituyendo un buen método para aislar las microcalcificaciones y así analizarlas con mayor facilidad, son un método de diagnóstico confiable que puede detectar calcificaciones que insinúan malignidad.

Procesamiento digital de imágenes

La imagen digital está formada por una serie de matrices numéricas que definen sus características. Una imagen digital de $M \times N$ píxeles en escala de grises L es una función en el que cada píxel de la imagen en las coordenadas (x, y) contiene un nivel de gris:

$$f : [0, M - 1]_x [0, N - 1]_y \rightarrow [0, L - 1] \quad \text{Ecuación 2.1}$$

Donde:

M y N : Tamaño de la imagen en píxeles.

L : Niveles de gris.

x y y : Coordenadas de largo y ancho en la imagen.

El procesamiento digital de imágenes puede definirse como la operación de imágenes mediante computadora. Existen hoy en día sistemas de procesamiento de imagen que operan sobre imágenes generadas a partir del sensado de rayos X, rayos gamma, resonancia magnética, microondas entre otras. Las técnicas para analizar una imagen digital se agrupan en un área llamado Procesamiento Digital de Imágenes, las herramientas de procesamiento digital se organizan según el nivel de procesamiento y éstos a su vez forman una cadena (Véase Figura 2.5).



Figura 2.5. Cadena de procesamiento en Procesamiento Digital de Imágenes.

En el pre-procesamiento se realizan operaciones para adaptar la información de la imagen y tener mejor análisis en pasos posteriores, ejemplos de procesamiento son las operaciones de brillo y contraste. La segmentación consiste en hacer una partición de la imagen en varias regiones que representen la información necesaria para el problema a resolver. En el nivel de detección de objetos y clasificación se determinan y clasifican los objetos contenidos en la imagen. Cuando se obtiene información de alto nivel acerca de lo que la imagen muestra entonces se está trabajando en el análisis de imagen (Ramírez and Chacón, 2011).-

El interés de los métodos de procesamiento de imágenes digitales se fundamenta en dos áreas principales de aplicación, la primera de ellas es el de mejorar la calidad de la imagen para la interpretación por parte del humano, y la segunda es el procesamiento de los datos de la escena para la percepción de las máquinas de forma automática.

Descriptores morfológicos

Los descriptores son características que se expresan más o menos estables bajo la influencia de diferentes condiciones del medio y permiten identificar los individuos, mientras que la morfología en medicina es el examen cualitativo y cuantitativo de ciertos componentes o sustancias del organismo según métodos especializados, con un fin de diagnóstico.

Si se desea describir a un objeto es necesaria la utilización de las características de los objetos presentes, estas características son una determinada medida cuantitativa las cuales se obtienen mediante el cálculo directo de los píxeles que lo componen. Para describir a un objeto normalmente se combinan varias de las características en un vector para ser clasificados mediante la discriminación de las diferencias existentes entre los vectores de características de los distintos objetos. Para ello las características a utilizar deben tener las propiedades de: inmunidad a cambios de rotación, de desplazamiento y de escalado, sensibilidad para diferenciar objetos casi iguales, unicidad (cada objeto debe tener una única representación) y la capacidad de representar los rasgos característicos básicos de los objetos y abstraer los detalles.

Métodos de clasificación

Los seres humanos pueden reconocer patrones familiares u objetos independientemente de sus diferencias en orientación y tamaño. Esto es debido a nuestro sistema inteligente de percepción el cual ha sido entrenado para

reconocer los objetos a través del tiempo. El reconocer patrones está definido como “la acción de tomar un vector de datos y tomar una acción basada en la categoría del patrón de identificación”. La identificación de objetos en imágenes se realiza por formas, color, textura, temperatura y otros atributos. Las redes multicapa de retropropagación son una herramienta poderosa para el reconocimiento de patrones y la clasificación (Bharathi and Subashini, 2011).

Redes neuronales artificiales

El sistema nervioso de los animales se estructura con neuronas que forma redes para el procesamiento de información generada por los diversos estímulos de entrada de los sistemas sensoriales. Estas redes han servido de inspiración para modelar redes neuronales artificiales basadas en distintas teorías de funcionamiento de las redes neuronales biológicas.

Las redes neuronales artificiales se aplican en distintas áreas de ingeniería como su aplicación para la solución de problemas en procesamiento de imágenes, sistemas financieros, técnicas biométricas, criptosistemas, control automático, etc. Desde hace décadas las redes neuronales artificiales han sido herramientas para resolver problemas en los diferentes niveles de procesamiento de imágenes, debido a su potencial para reconocer patrones de información la mayoría de las aplicaciones para las redes neuronales artificiales en imágenes digitales son segmentación de forma y clasificación. En el Cuadro 2.3 se muestran los trabajos realizados en la última década con las redes neuronales artificiales en las distintas tareas en el procesamiento de imágenes (Ramírez and Chacón, 2011).

Cuadro 2.3. Tipo de procesamiento realizado por las redes neuronales artificiales.
Ramírez y Chacón (2011).

Procesamiento	Número de artículos encontrados
Eliminación de ruido	1
Segmentación de textura	10
Segmentación de movimiento	8
Detección de bordes	13
Segmentación de forma	53
Aplicaciones específicas	15
Segmentación de color	13
Clasificación	43
Comparación entre redes neuronales artificiales	3
Otros	8
Total de artículos revisados	167

Hoy en día existen diferentes tipos de redes neuronales que debido a sus características son utilizadas en algún área en específico en el procesamiento de imágenes. Según Ramírez y Chacón (2011) de los 43 trabajos que utilizan las redes neuronales para la clasificación la Teoría de Resonancia Adaptativa (ART por sus siglas en inglés) es la red neuronal más utilizada para la tarea de clasificación encontrando 15 trabajos, seguido de la red neuronal con funciones de base radial con 10 trabajos, la red multicapa y las redes neuronales probabilísticas se utilizaron en 5 trabajos cada uno.

En el siguiente apartado se hablará acerca de los trabajos que han sido realizados para la detección y clasificación automática de lesiones encontradas en mamografías mediante el uso de técnicas de procesamiento de imágenes y reconocimiento de patrones.

Detección y clasificación de lesiones en mamografías

Una técnica para determinar la homogeneidad de la forma de la lesión en las mamografías es introducido por Khademi et al. (2009) cuyo análisis se basa en la complicación de la frontera es decir, en los límites. Para lograr esto, se realiza primeramente una detección de bordes a través de la transformación espacio-frecuencia (Wavelet) y se extraen las características en el dominio Wavelet, ya que cada coeficiente en este dominio individualiza la información del borde. Un conjunto de valores de coeficientes para una lesión benigna es relativamente homogéneo, mientras que el de las lesiones malignas es heterogéneo.

Después de la extracción una técnica de reconocimiento de patrones es necesaria para la clasificación correcta de la información, Khademi et al. (Khademi et al., 2009) utilizaron dos métodos un clasificador no lineal (K-NN) y un clasificador lineal (Análisis discriminante lineal LDA). De un total de 60 imágenes benignas y 39 malignas el mejor clasificador fue el LDA con un 72.5% de precisión, sin embargo el clasificador K-NN provee una tasa de clasificación más alta en los casos benignos, véase Cuadro 2.4 donde se muestran los resultados de la clasificación.

Cuadro 2.4. Resultados de la clasificación de lesiones por Khademi et al. (2009).

	K-NN (k=3)	K-NN (k=4)	LDA
Benigno	87%	87%	73%
Maligno	33%	33%	72%

Han sido varios los métodos de clasificación de lesiones en las mamografías, entre estos se encuentra el uso de una red múltiple de Programación Genética Cartesiana (*Cartesian Genetic Programming CGP*) en el cual se demostró la clasificación correcta de las microcalcificaciones como benignas y malignas de un pequeño grupo de imágenes mamográficas. Este nuevo método fue implementado por Volk et al. (2009).

Otro método muy frecuentemente utilizado para este objetivo son las redes neuronales artificiales en los cuales se les observa en ocasiones en la combinación con otros métodos tanto de extracción de características como de clasificación para mejorar su desempeño. Tal es el caso presentado por Simily y Balakrishnan (2011), en el que combinan las redes neuronales con descriptores de patrones binarios locales, la transformada wavelet discreta Haar y características de textura Haralick; la precisión promedio que obtuvieron con este método de clasificación para masas en mamografías fue de 98.6% en el entrenamiento, validación y prueba de la red neuronal.

Las wavelet no sólo son usadas para la extracción de características si no que se puede detectar a las microcalcificaciones mediante su empleo, como lo hizo Ocusse et al. (2009) quienes utilizaron diferentes familia wavelet para detectar microcalcificaciones como: Haar, DB4, DB6, DB8, DB10, DB12, DB20 (Daubechies), SY8, SY12, SY16 (Symmlets), CF6, CF12, CF18, CF24, y CF30 (Coiflets), y la mejor eficacia del método la obtuvo usando SY16, CF6, SY8, y SY12.

Rajkumar y Raju (2011) encontraron que una parte fraccional del coeficiente wavelet más grande es suficiente para describir características de textura en una imagen, basados en este concepto desarrollaron un algoritmo de clasificación de imágenes mamográficas en 3 clases normal, benigna y maligna. Esta clasificación se hace a través de diferentes coeficientes y familias Wavelet. Los resultados se muestran en Cuadro 2.5. Donde el número de imágenes utilizadas es de 162, de las cuales 98 son imágenes normales, 38 imágenes benignas y 26 imágenes malignas. Donde todos los métodos utilizados tuvieron su mejor rendimiento cuando se usa el coeficiente al 50% ya que clasifican correctamente a 74 imágenes normales, 38 imágenes benignas y 22 malignas.

Cuadro 2.5. Éxito en la clasificación de imágenes mamográficas usando la familia Daubechies, Haar y Biorthogonal. Rajkumar y Raju (2011).

Coef. %		Daubechies									Haar			Biorthogonal			
		Db4			Db8			Db16									
		Clasificación obtenida (número de imágenes)															
		N	B	M	N	B	M	N	B	M	N	B	M	N	B	M	
25	Diagnóstico radiólogo	N	72	09	17	73	08	17	72	09	17	72	09	17	68	03	27
		B	0	37	01	0	38	0	0	38	0	0	37	01	0	36	02
		M	04	03	19	04	03	19	04	03	19	04	03	19	02	01	23
50	Diagnóstico radiólogo	N	74	06	18	74	06	18	74	06	18	74	06	18	68	03	27
		B	0	38	0	0	38	0	0	38	0	0	38	0	0	36	02
		M	02	02	22	02	02	22	02	02	22	02	02	22	02	01	23
75	Diagnóstico radiólogo	N	69	06	23	69	06	23	69	06	23	69	06	23	68	03	27
		B	0	36	02	0	36	02	0	36	02	0	36	02	0	36	02
		M	02	01	23	02	01	23	02	01	23	02	01	23	02	01	23
100	Diagnóstico radiólogo	N	68	03	27	68	03	27	68	03	27	68	03	27	68	03	27
		B	0	36	02	0	36	02	0	36	02	0	36	02	0	36	02
		M	02	01	23	02	01	23	02	01	23	02	01	23	01	01	02

Nota: N=Imágenes normales, B=Imágenes benignas, M=Imágenes malignas.

Las máquinas de soporte de vectores son muy comunes en clasificación gracias a su eficiencia, Qi Guo et al. (2009) utilizaron este método ya que ellos basaron sus estudios en la textura de las lesiones y el tejido de las mamas encontradas en las imágenes mamográficas. Así mismo, cuando el número de características obtenidas en una mamografía son infinitas, es necesario un método de selección de características, Yoon y Kim (2009) proponen un método de selección basado en una máquina de soporte de vectores recursivo de eliminación de características (MSVM-RFE). Las características que se utilizaron para este método fueron los establecidos por BI-RADS: forma de la masa, margen de la masa, tipo de calcificación, distribución de la calcificación, densidad y valoración. Los resultados obtenidos fueron eficientes, dando un buen rendimiento en 3 de los 4 conjuntos de datos de calcificaciones.

Detección de grupos de microcalcificaciones

Es importante determinar en una mamografía si existe un conjunto de microcalcificaciones ya que dichos grupos pueden indicar malignidad. Siong et al. (2009) determinaron el número de grupos sospechosos de microcalcificaciones encontrados en las mamografías. Dicho método detecta los límites de las microcalcificaciones y determina el número de microcalcificaciones encontradas, este método está basado en la detección de bordes de Sobel.

Chandra et al. (2011) también presentan un método para la detección automática de grupos de microcalcificaciones tanto benignas como malignas utilizando el Índice de Calidad de Imagen Universal (UIQI), que se basa en las medidas de distorsión entre la imagen original e imágenes de referencia que se componen distorsión de correlación, distorsión de iluminación y distorsión de contraste.

Descripción y clasificación de microcalcificaciones

La investigación conducida por Lee et al. (2000) presentó una detección y clasificación de microcalcificaciones *shape cognitron*, el cual está basado en una red neuronal como sistema de reconocimiento de forma, es introducido como una técnica de clasificación de microcalcificaciones. El sistema tuvo una clasificación de 95% en 50 casos de prueba donde 14 eran microcalcificaciones benignas y 36 malignas.

Breast Cancer Detection System (BCDS) fue desarrollado por Verma y Zakos (2001) con el objetivo de investigar la efectividad de la combinación de características de extracción. 14 características fueron seleccionados para formar el vector de características más significativo, basados en la tasa de clasificación en la red neuronal. Inicialmente cada característica alimenta a la red una por una, si la tasa de clasificación se incrementa o se mantiene esta característica es

incluida en el vector. Sólo cinco de las características más significativas fueron seleccionadas: oblicuidad, entropía, número de píxeles, histograma y desviación estándar, obteniendo una tasa de clasificación de 88.9%.

Christoyianni et al. (2002) examinaron la eficiencia de 3 juegos de descriptores: momentos del histograma de niveles de gris (GLHM por sus siglas en inglés), matriz de dependencia espacial en niveles de gris (SGLD por sus siglas en inglés) y análisis de componente independiente (ICA por sus siglas en inglés) con la finalidad de clasificar las microcalcificaciones para el diagnóstico de cáncer de mama. El clasificador de función de base radial es empleado para reconocer cualquier tipo de cáncer en mamografía, el resultado de la clasificación muestra que ICA tuvo la mayor tasa de reconocimiento con 79.31% comparado con 72.41% obtenido por el SGLD y 70.68% del GLHM.

Ballesteros et al. (2002) usaron dos técnicas para clasificar microcalcificaciones geométricamente (características de forma) y clasificar por agrupamiento, el sistema da un rendimiento de 82.3% para la clasificación geométrica y 73.7% por agrupamiento.

Wroblewska, et al. (2003) presentaron la clasificación de microcalcificaciones utilizando una red multicapa de retropropagación a través de las siguientes características (Véase Cuadro 2.6).

Cuadro 2.6. Lista de características aplicadas para la clasificación individual de las microcalcificaciones. Wroblewska, et al (2003).

No.	Nombre	Descripción
1	Área	Medida del objeto
2	Promedio	Promedio de los niveles de gris, desviación estándar de los niveles de gris, niveles de gris del fondo.
3	Desviación estándar	
4	Fondo	
5	Diferencia en proporción del objeto y el fondo (contraste entre el objeto y el fondo)	
6	Compacidad	$(\text{Perímetro del objeto})^2 / \text{Área}$
7	1er Momento de Forma	
8	Siete momentos de Hu	
9	Contraste	Matriz de ocurrencia relacionada a características
10	Entropía	
11	Energía	
12	Momento inverso diferente	
13	Correlación	
14	Varianza	
15	Suma del promedio	
16	Suma de la entropía	
17	Suma de la varianza	
18	Diferencia del promedio	
19	Diferencia de la entropía	
20	Diferencia de la varianza	

Quienes tuvieron un 89% de microcalcificaciones detectadas correctamente, 76% en la clasificación de microcalcificaciones benignas y malignas y un 70% en la clasificación de benignas, malignas y falsas; éstos resultados se obtuvieron al hacer la extracción automática de las características. Al realizar una selección subjetiva de las características mencionadas en el Cuadro 2.1 los resultados arrojaron lo siguiente: 90% detección correcta, 74% clasificación benigna/maligna y 68% clasificación benigna/maligna/falsa.

Mousa et al. (2005) enfatizaron en el análisis wavelet para la extracción de características y sistemas de inferencia adaptativa neuro-difuso (ANFIS por sus siglas en inglés) como método de clasificación propuesto, la investigación mostró

que el método es muy efectivo para la detección automática y clasificación de microcalcificaciones en mamografías digitales.

Delogu et al. (2005) utilizan una descomposición de la región del pecho en NxN píxeles de ancho parcialmente sobreponiendo una sub-imagen y utilizando una red neuronal auto asociativa para la extracción automática de características. El perceptrón multicapa fue elegido para realizar la clasificación, el cual está basado en el algoritmo de retropropagación. El rendimiento de este método obtuvo una sensibilidad de 82.2% y una tasa de falsos positivos de 4.15.

Papadopoulos et al. (2005) caracterizaron cada grupo de microcalcificaciones en benigno o maligno empleando dos diferentes esquemas de clasificación basados en máquina de soporte de vectores y redes neuronales. Se realizaron medidas de rendimiento a los métodos de clasificación en las bases de datos Nijmegen y MIAS, de estos experimentos, la máquina de soporte de vectores (SVM) con función kernel Gaussiana provee un mejor rendimiento comparado con la red neuronal.

Xu et al. (2006) detectan a las microcalcificaciones utilizando la transformada wavelet discreta (DWT por sus siglas en inglés) y la técnica de llenado por dilatación fue usada para la segmentación de las microcalcificaciones del fondo. En el experimento se seleccionaron 10 características para describir a las microcalcificaciones incluyendo: área, intensidad promedio, contraste, coherencia, compacidad, relación de huecos, número de huecos, elongación, dimensión fractal y número de agrupación. Un perceptrón multicapa fue usado para la clasificación, la arquitectura consiste en una capa de entrada con 10 nodos, 20 nodos en la capa oculta y un nodo en la salida. Este experimento tuvo un 96.9% en la tasa de verdaderos positivos y sólo 0.2 falsos positivos por imagen.

Los coeficientes wavelet son utilizados para extraer información importante de las microcalcificaciones, Karahaliou et al. (2008) lo utilizan para la extracción de características de textura del tejido que rodea a las microcalcificaciones. La clasificación es realizada por una red neuronal probabilística PNN (por sus siglas en inglés) cuyo vector de entradas es una combinación de características de textura de niveles de gris y las características de textura de la extracción Wavelet, este sistema clasifica al tejido en dos tipos maligno o benigno.

En el Cuadro 2.7 se muestran las características para cada grupo de extracción de información de textura aplicados a la red neuronal probabilística. Este método fue evaluado usando sus características individuales y en el Cuadro 2.8 se muestra el desempeño obtenido por cada uno de ellos en cuyo cuadro se puede observar que el método de extracción combinado que proponen Karahaliou et al. (2008) tiene un mejor rendimiento.

Cuadro 2.7. Combinación óptima de características para cada grupo de característica de textura. Karahaliou et al. (2008).

Grupo de característica	Mejor grupo de característica
GLFOS	Promedio Oblicuidad
GLCM	Tono promedio Intervalo de contraste Intervalo en la diferencia de entropía Intervalo en la diferencia de varianza
LTEM	Curtosis de S5L5TR Oblicuidad de de S5L5TR Promedio de R5L5TR Promedio de L5L5TR STD de de S5L5TR
WCFOS+WCCM 1er Escala	Entropía Homogeneidad local promedio Diferencia de entropía promedio Intervalo de la diferencia de varianza
WCFOS+WCCM 2da Escala	Entropía Prominencia promedio Intervalo de varianza Suma ponderada promedio
WCFOS+WCCM 3er Escala	Entropía Entropía promedio Suma de entropía promedio Tono promedio Intervalo de la suma ponderada Intervalo IMC2
DDSM	Base de datos que provee una valoración maligna para cada grupo de microcalcificaciones

Cuadro 2.8. Desempeño de la clasificación de los grupos de características de manera individual y el esquema combinado. Karahaliou et al. (2008).

Grupo de característica	Desempeño A_z
GLFOS	0.763
GLCM	0.813
LTEM	0.903
WCFOS+WCCM 1er Escala	0.864
WCFOS+WCCM 2da Escala	0.860
WCFOS+WCCM 3er Escala	0.917
DDSM	0.838
Esquema combinado	0.989

El siguiente Cuadro (Cuadro 2.9) es un análisis sobre los trabajos relacionados a la clasificación de microcalcificaciones vistos en este apartado. Como se puede observar la mayoría de los trabajos se enfocan en las características de forma de las microcalcificaciones como método de extracción, esto debido a que es uno de los elementos importantes para determinar si la microcalcificación anuncia un proceso maligno o benigno. Otra extracción es la que incluye la textura de la microcalcificación en conjunto con la forma, esto también se puede observar en trabajos relacionados con la clasificación de masas encontradas en mamografías.

También en el Cuadro 2.9 se muestran los clasificadores utilizados en dichos trabajos, de los cuales se puede concluir que la tendencia es utilizar redes neurales en sus diferentes formas, debido a su capacidad de clasificar sistemas complejos como los son las microcalcificaciones. Con lo que respecta al tipo de microcalcificación en la clasificación, todos los trabajos utilizan las variables benigno y maligno; excepto Lee et al. (2000) quienes utilizan la clasificación BIRADS.

Cuadro 2.9. Análisis de métodos de extracción y clasificación de microcalcificaciones de trabajos previos revisados.

Característica	Método	Número de trabajos revisados
Extracción	Características morfológicas	3
	Características morfológicas y textura	2
	Análisis componente independiente	1
	Descomposición wavelet	1
	Red neuronal auto-asociativa	1
	Análisis de textura	1
Clasificación	Red neuronal multicapa	4
	<i>Shape cognitron</i>	1
	Clasificador función de base radial	1
	Geometría y por grupos de microcalcificaciones	1
	KNFIS	1
	Máquina de soporte de ectores	1
	Red neuronal probabilística	1

III. METODOLOGÍA

El trabajo se desarrolló en las instalaciones de la Universidad Autónoma de Querétaro utilizando las mamografías de la base de datos mamográfica MIAS (*Mammographic Image Analysis Society*) y el software MatLAB. El algoritmo está compuesto por tres etapas (ver Figura 3.1), la primera consiste en la binarización de la microcalcificación contenida en la mamografía, a partir de esta imagen se obtiene el esqueleto y el contorno de la microcalcificación. Posteriormente al objeto contenido en las imágenes se le extraen las características morfológicas mediante los descriptores mencionados a continuación: Elongación, compacidad, momentos de Hu, área, ramificación y momento del contorno; ésta información pasa por el clasificador red multicapa de retropropagación (tercera etapa) y finalmente determinar el tipo de microcalcificación.

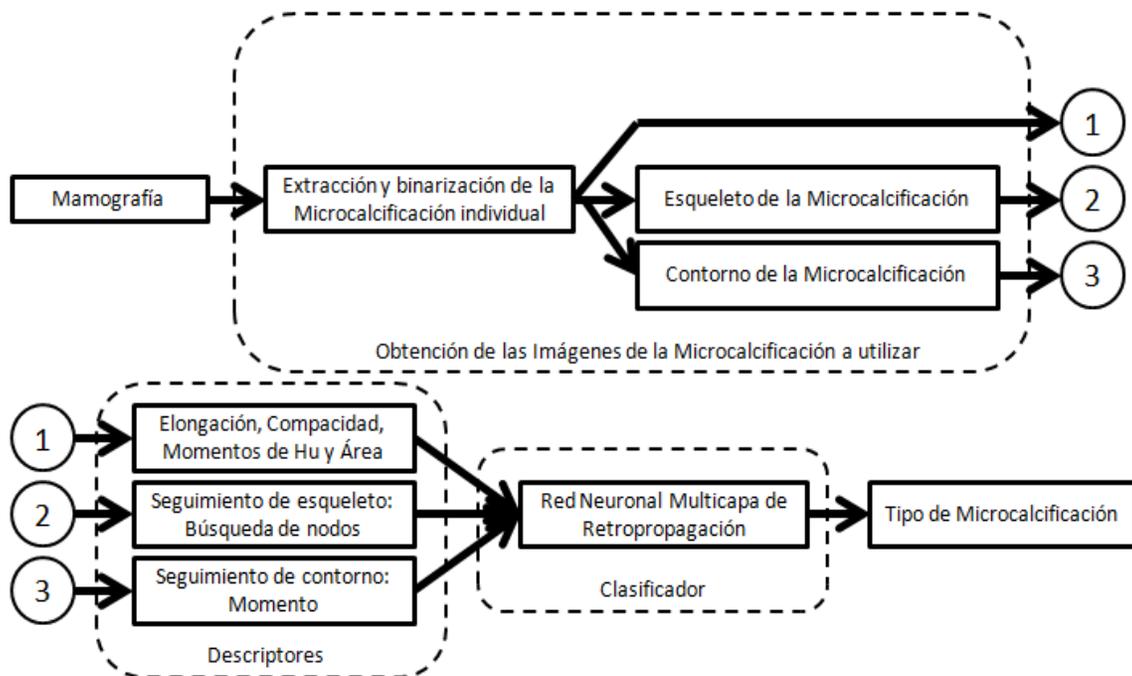


Figura 3.1. Diagrama de bloques del algoritmo de clasificación de microcalcificaciones.

Base de datos MIAS

La base de datos mamográfica MIAS (Suckling et al., 1994) consiste de un total de 322 imágenes de las cuales 22 contienen microcalcificaciones benignas y malignas, son imágenes de una resolución de 1024 píxeles por 1024 píxeles y cada píxel corresponde a 50 micrómetros.

MatLAB

MatLAB es un programa que permite realizar cálculos técnicos y científicos, con ya varios años en el mercado y con una gran cantidad de funciones implementadas para diferentes disciplinas científicas se ha convertido en un estándar de programación y desarrollo de aplicaciones. La unidad de procesamiento de MatLAB es la matriz, por lo que su utilización en el procesamiento de imágenes (matrices) es una extensión natural de su lenguaje de programación.

Obtención de las microcalcificaciones y extracción individual

Las imágenes formato pgm son leídas a partir del software MatLAB y la ubicación de la microcalcificación es proporcionada por el diagnóstico de la base de datos MIAS. Es necesaria la extracción de las microcalcificaciones de manera individual, a partir de la siguiente información:

- Nombre de la mamografía (Número de referencia en la Base de Datos MIAS).
- Tipo de microcalcificación según Le Gal (1, 2, 3, 4 o 5).
- Clase de anormalidad presente (CALC: Calcificación).
- Severidad de la anormalidad (B: Benigno, M: Maligno).
- Coordenada y del centro de la anormalidad.
- Coordenada x del centro de la anormalidad.
- Radio aproximado del círculo que envuelve a la microcalcificación en píxeles.

A continuación se muestra el formato dado para la información detallada de la base de datos MIAS.

mdb233.pgm 2 CALC M 623 591 5

A partir del nombre de la mamografía, las coordenadas (x, y) del centro y el radio del círculo que envuelve a la microcalcificación se dibuja un cuadrado de tamaño:

$$y1 = Cy - \text{radio} \quad \text{Ecuación 3.1}$$

$$y2 = Cy + \text{radio} \quad \text{Ecuación 3.2}$$

$$x1 = Cx - \text{radio} \quad \text{Ecuación 3.3}$$

$$x2 = Cx + \text{radio} \quad \text{Ecuación 3.4}$$

Donde:

$y1$ = Coordenada y inicial del cuadrado,

$y2$ = Coordenada y final del cuadrado,

$x1$ = Coordenada x inicial del cuadrado,

$x2$ = Coordenada x final del cuadrado,

Cy = Es la coordenada en y del centro de la microcalcificación,

Cx = Es la coordenada en x del centro de la microcalcificación, y

radio = Es el radio que envuelve a la microcalcificación.

En la siguiente imagen (Figura 3.2) se muestra la mamografía mdb233 con la microcalcificación localizada en el punto 623 591 y radio 5.



Figura 3.2. Mamografía mdb233 de la base de datos MIAS.

La extracción de la microcalcificación de manera individual parte del cuadrado definido y la nueva imagen se crea como se muestra en la siguiente ecuación:

$$I(x, y) = \begin{bmatrix} M(x1, y1) & M(x1+1, y1) & \dots & M(x2, y1) \\ M(x1, y1+1) & M(x1+1, y1+1) & \dots & M(x2, y1+1) \\ \vdots & \vdots & \ddots & \vdots \\ M(x1, y2) & M(x1+1, y2) & \dots & M(x2, y2) \end{bmatrix} \quad \text{Ecuación 3.5}$$

Donde:

$I(x, y)$ = es la nueva imagen.

M = corresponde al nombre de la mamografía

$x1, y1, x2, y2$ = son las coordenadas del rectángulo que envuelve a la microcalcificación.

La Figura 3.3 muestra la microcalcificación de la mdb233 localizada en $Cy = 623$, $Cx = 591$ y $radio = 5$ ya extraída formando una nueva imagen.

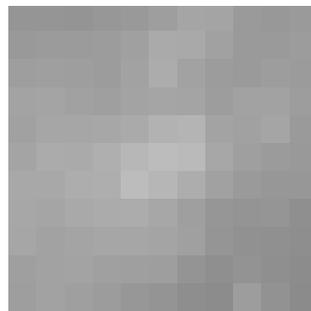


Figura 3.3. Microcalcificación: mdb233.pgm 2 CALC M 623 591 5.

Binarización de la microcalcificación

Para conocer la forma de la microcalcificación es necesario separar el fondo de la imagen y la microcalcificación para ello se aplica la binarización, donde 0 es el fondo de la imagen y 1 es el valor para los píxeles correspondientes a la microcalcificación.

En esta etapa se hace un acoplamiento a la imagen la cual consiste en oscurecerla, para posteriormente a la imagen original restarle la imagen oscurecida, de manera que se obtengan los valores con más diferencia como los píxeles más claros correspondientes a la microcalcificación y los de menor diferencia para los píxeles del fondo. El efecto de oscurecimiento se logra empleando el filtro de mínimos en una vecindad 8 como corresponde (Ecuación 3.6).

$$I'(x, y) = \min \{I(x + i, y + j) | (i, j) \in \mathfrak{R}\} = \min(\mathfrak{R}(x, y)) \quad \text{Ecuación 3.6}$$

$$I'' = I - I' \quad \text{Ecuación 3.7}$$

Aplicando la ecuación 3.7 se obtienen las diferencias que permiten reconocer la microcalcificación del fondo. Una vez realizado el realce se necesita conocer el umbral requerido para binarizar la imagen de una manera óptima.

En el siguiente ejemplo se ilustran los pasos anteriormente mencionados, primeramente en la Figura 3.4 se muestra el histograma de la microcalcificación mostrada en la Figura 3.3, en donde claramente se muestra que los valores del fondo y la microcalcificación no están completamente definidos.

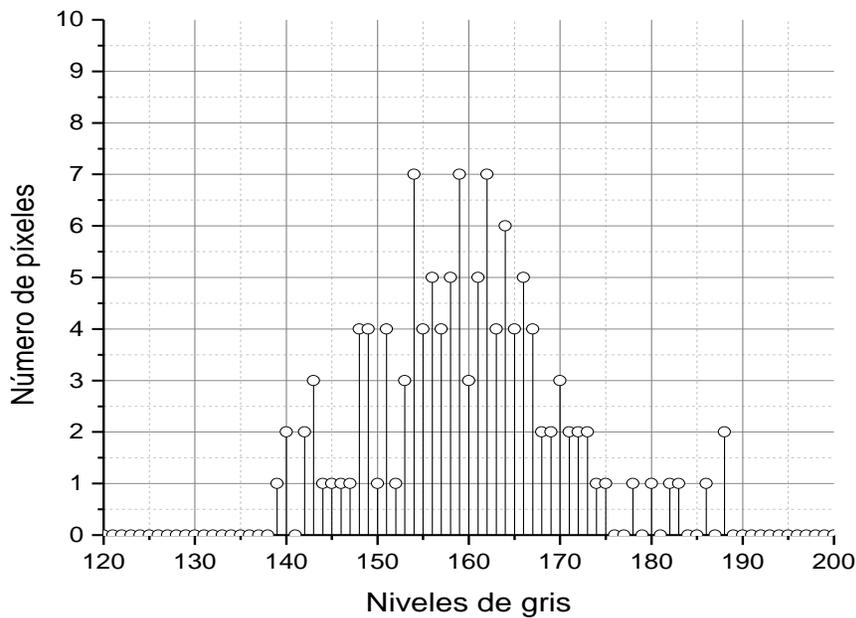


Figura 3.4. Histograma de la Microcalcificación: mdb233.pgm 2 CALC M 623 591 5.

Al aplicar las ecuaciones 3.6 y 3.7 a la microcalcificación mdb233.pgm 2 CALC M 623 591 5 obtenemos lo que se muestra en la Figura 3.5 y su respectivo histograma en la Figura 3.6.

0	0	0	0	0	0	0	0	0	0	0
0	7	7	8	9	17	13	7	2	2	0
0	6	3	1	6	12	6	2	0	3	0
0	6	5	3	7	3	8	3	8	8	0
0	5	7	8	11	15	23	6	6	12	0
0	10	4	9	16	18	25	13	8	4	0
0	4	8	5	21	23	23	13	6	8	0
0	3	7	7	8	8	10	5	4	7	0
0	3	4	7	8	16	17	6	2	4	0
0	4	7	9	11	16	8	3	7	4	0
0	0	0	0	0	0	0	0	0	0	0

Figura 3.5. Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 tras aplicar el oscurecimiento de la imagen.

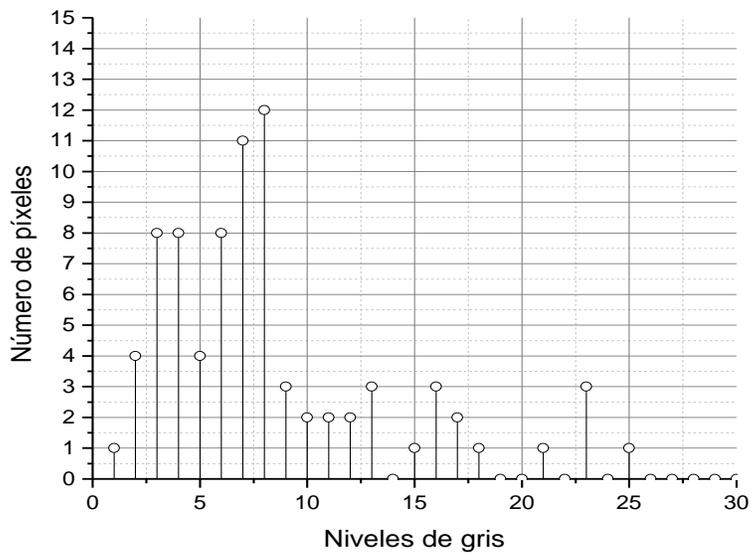


Figura 3.6. Histograma de la Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 tras aplicar el oscurecimiento de la imagen.

Se aplica una umbralización a la imagen acoplada a partir de la desviación estándar, dicha umbralización se realiza mediante el método de *Hard thresholding* (Umbral duro), cuya ecuación se muestra a continuación:

$$\delta_{\lambda}(w) = \begin{cases} w, & \text{si } |w| > \lambda \\ 0, & \text{en caso contrario} \end{cases} \quad \text{Ecuación 3.8}$$

Donde:

δ_{λ} = Valor de la umbralización.

w = Nivel de gris del píxel.

λ = Valor de umbral.

Como se muestra en la Figura 3.7 los píxeles correspondientes a la microcalcificación y otros puntos brillantes mantienen un valor distinto de 0, mientras que todos aquellos píxeles que corresponden al fondo, es decir, los menos brillantes son llevados al valor 0. En la Figura 3.8 se ilustra el histograma de la Figura 3.7 donde se observa que los valores de fondo son 0 y los distintos a este corresponden a los puntos más brillantes de la imagen original (Figura 3.3).

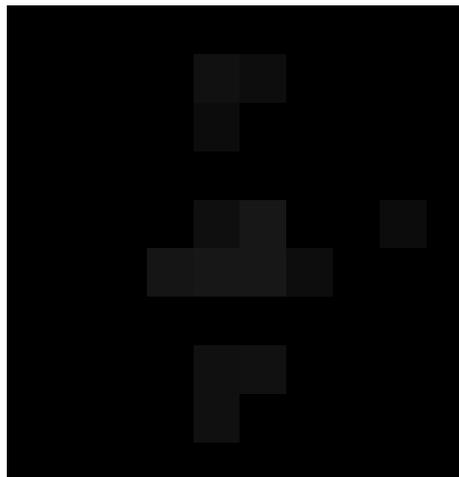


Figura 3.7. Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 después de la umbralización mediante umbral duro.

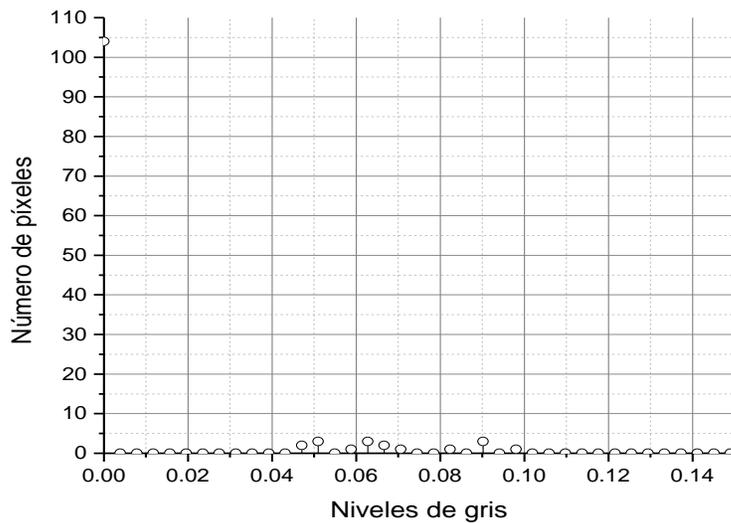


Figura 3.8. Histograma de la Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 después de la umbralización mediante umbral duro.

Se realiza una expansión del histograma basado en la ecuación 3.9 antes de llevarlo a la binarización con el objetivo de realzar el contraste.

$$ecualización(x, y) = \left[\frac{C-D}{A-B} \right] \cdot [I(x, y) - B] + D \quad \text{Ecuación 3.9}$$

Donde:

A = Nivel de gris máximo en la imagen

B = Nivel de gris mínimo en la imagen

C = Máximo nivel de gris disponible

D = Mínimo nivel de gris disponible

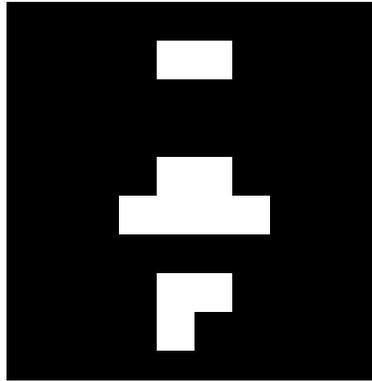


Figura 3.9. Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 tras la expansión del histograma.

A continuación, obtenemos una imagen binaria basándonos en el hecho de que los píxeles de una región presentan una distribución de intensidad similar, es por eso que a partir del histograma se determina la zona a la que corresponde. El método de Otsu supone que la medida de las Gaussianas (suma de dos densidades de probabilidad) se asemeja al histograma real, las desviaciones estándar deben disminuir y como consecuencia se debe elegir un umbral que minimice la suma de las varianzas de los dos lóbulos del histograma (Ecuación 3.10).

$$\sigma_w^2(t) = w_1(t)\sigma_1^2 + w_2(t)\sigma_2^2 \quad \text{Ecuación 3.10}$$

$$w_1(t) = \sum_{z=1}^T P(z) \quad \text{y} \quad w_2(t) = \sum_{z=T+1}^L P(z) \quad \text{Ecuación 3.11}$$

$$\sigma_1^2(t) = \sum_{z=1}^T (z - \mu_1(t))^2 \frac{P(z)}{w_1(t)} \quad \text{y} \quad \sigma_2^2(t) = \sum_{z=T+1}^L (z - \mu_2(t))^2 \frac{P(z)}{w_2(t)} \quad \text{Ecuación 3.12}$$

$$\mu_1(t) = \sum_{z=1}^T zP(z) \quad \text{y} \quad \mu_2(t) = \sum_{z=T+1}^L zP(z) \quad \text{Ecuación 3.13}$$

Donde:

σ_w^2 = Es la varianza ponderada,
 w_1, w_2 = Probabilidades acumuladas
 σ_1, σ_2 = Varianzas asociadas,
 z = Nivel de intensidad
 T = Umbral buscado,
 L = Nivel de intensidad máximo,
 $P(z)$ = Función de densidad de probabilidad, y
 μ_1, μ_2 = Medias.

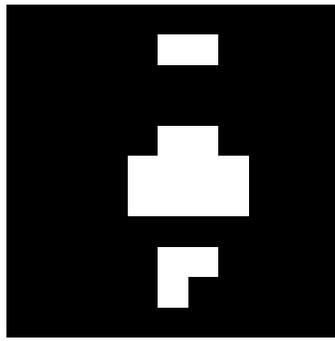


Figura 3.10. Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 binarizada.

Después de la binarización es posible encontrar píxeles que con valor 1 (blancos) no pertenecen a la microcalcificación pero debido a que estaba rodeado de puntos oscuros prevaleció su diferencia alta. Para eliminar el ruido tipo sal (puntos blancos en la imagen) comúnmente se utiliza el filtro de mínimos, la desventaja de utilizarlo para las microcalcificaciones es que afecta directamente a la forma de éstas, ya que están compuestas por un pequeño grupo de píxeles.

En este caso se utiliza el etiquetado de objetos, de tal manera que se localiza un píxel perteneciente a la microcalcificación y a todos los píxeles conectados a partir de éste, finalmente todos aquellos píxeles que no fueron etiquetados se consideran de fondo.

El algoritmo de etiquetado consiste en:

- 1) Encontrar el centroide de la microcalcificación.
- 2) Etiquetado a partir del centroide de la imagen total.
- 3) Todos aquellos píxeles fuera de la etiqueta principal se considera con valor 0.

Finalmente la binarización sin ruido de la microcalcificación mdb233.pgm 2 CALC M 623 591 5 se muestra en la Figura 3.11.

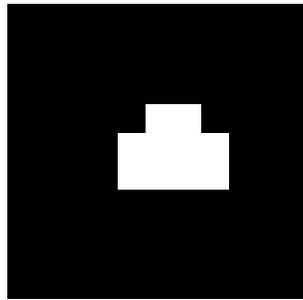


Figura 3.11. Microcalcificación: mdb233.pgm 2 CALC M 623 591 5 tras la eliminación de ruido.

Esqueleto de la microcalcificación

Una característica de los objetos que nos proporciona información sobre su topología y estructura es el esqueleto, este representa la estructura del objeto con un número menor de píxeles pero conservando su conectividad, agujeros y extensión. Una manera de conocer el esqueleto morfológico de una imagen está dada por la Ecuación 3.14 (Fórmula de Lantuéjoul), la cual consiste en una sucesión de erosiones y aperturas con un elemento estructurante.

$$S(I) = \bigcup_{k=0}^K S_k(I), \text{ donde } S_k(I) = \bigcup_{k=0}^K \{(I \ominus k B) - [(I \ominus k B) \circ B]\} \quad \text{Ecuación 3.14}$$

Aplicando la ecuación 3.14 a la Figura 3.11 se obtiene el esqueleto de la microcalcificación mdb233.pgm 2 CALC M 623 591 5 (Ver Figura 3.12).

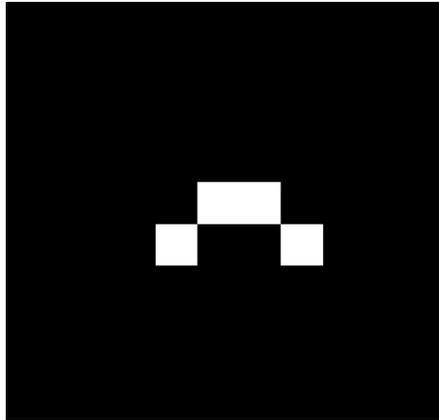


Figura 3.12. Esqueleto de la Microcalcificación: mdb233.pgm 2 CALC M 623 591
5.

Contorno de la microcalcificación

Tanto la imagen binarizada, como el esqueleto de la microcalcificación y su contorno son imágenes base para la extracción de las características de forma. El gradiente morfológico (Ecuación 3.15 y 3.16) permite en imágenes binarias obtener los límites de las imágenes, ya sea contorno interior o exterior. Ya que las microcalcificaciones son objetos muy pequeños se opta por utilizar el límite externo ya que provee mayor información de forma.

$$\beta(I) = I - (I \ominus B) \quad \text{Ecuación 3.15}$$

$$\beta(I) = (I \oplus B) - I \quad \text{Ecuación 3.16}$$

La dilatación (\oplus) es el máximo de la suma realizada entre los valores de una estructura de referencia (B) y la región de la imagen con la que se corresponde. Mientras que la erosión (\ominus) es definida como el mínimo de la resta realizada entre los valores de la estructura de referencia (B) y la región de la imagen con la que se corresponde.

La dilatación y la erosión pueden ser combinadas para realizar una variedad de efectos. Por ejemplo, substraer la erosión de una imagen a su versión

dilatada se encuentra un tipo de gradiente morfológico. La erosión seguida de una dilatación crea una transformación morfológica llamada apertura y la dilatación seguida de una erosión es conocida como cierre, la operación de apertura suprime detalles luminosos más pequeños que la estructura de referencia y la operación de cierre elimina objetos oscuros, el objetivo de estas operaciones es suavizar y eliminar ruido en la imagen.

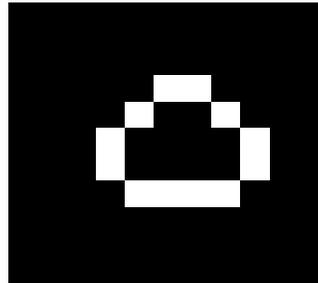


Figura 3.13. Contorno de la Microcalcificación: mdb233.pgm 2 CALC M 623 591 5.

La Figura 3.13 se obtiene al aplicar la ecuación 3.16 a la Figura 3.11 en el que se observa que el contorno a utilizar para aplicar descriptores es el contorno exterior de la microcalcificación.

Área y compacidad

Para describir una imagen se pueden utilizar las características de los objetos, como el perímetro, área, número de objetos, etc. Usar la combinación de dichas características en un vector nos permite la clasificación de objetos mediante la discriminación de diferencias.

Una manera de identificar la forma del objeto es mediante el perímetro y el área. El perímetro (p) consiste en calcular el número de píxeles en el borde de la imagen y el área (A) es el número de píxeles del objeto. La compacidad (circularidad) es un parámetro que no depende del tamaño de la región y viene dado por la división entre el área y su perímetro al cuadrado.

$$compacidad = \frac{A}{p^2} \quad \text{Ecuación 3.17}$$

La medida de compactación tiene el valor $\left(\frac{1}{4\pi}\right)$ para un objeto redondo de cualquier diámetro. Si es normalizado, la característica redondez/circularidad se define como se muestra en la Ecuación 3.18

$$compacidad = 4\pi \frac{A}{p^2} \quad \text{Ecuación 3.18}$$

Cuando se trata de un objeto redondo el valor máximo de la compacidad es 1, mientras que para otro tipo de formas el valor va entre 0 y menor de 1.

Elongación

Para obtener las características de objetos independientemente de la posición se puede considerar el centroide como referencia, de tal manera que el cálculo de los momentos se denominarán momentos centralizados. Para las imágenes binarias los momentos centralizados se expresan como:

$$\mu_{pq}(I) = \sum_{x,y \in I} (x - \bar{x})^p \cdot (y - \bar{y})^q \quad \text{Ecuación 3.19}$$

Este descriptor es invariante ante desplazamientos, pero si depende de la escala a la que se encuentre la imagen, para ello resulta necesaria la normalización y obtener momentos invariantes ante el tamaño. Dicha modificación se puede observar en la Ecuación 3.20, donde el factor correspondiente al área es normalizado.

$$\mu_{pq}(I) = \mu_{pq} \cdot \left(\frac{1}{\mu_{00}(I)} \right)^{\frac{(p+q+2)}{2}} \quad \text{Ecuación 3.20}$$

Donde: $(p + q) \geq 2$

Algunas medidas y características derivadas del uso de los momentos pueden ser directamente aplicables para caracterizar medidas geométricas del objeto. El momento permite determinar el ángulo de orientación del objeto, la elongación (Ecuación 3.21) ayuda a conocer la información de la forma, es decir, la relación entre longitud y anchura del objeto.

$$En(I) = \frac{[\mu_{20}(I) - \mu_{20}(I)]^2 + 4[\mu_{11}(I)]^2}{[\mu_{20}(I) - \mu_{20}(I)]^2} \quad \text{Ecuación 3.21}$$

Momentos invariantes

A partir de los momentos invariantes propuestos por Hu (1962) se pretende reconocer un objeto independientemente de la posición, escalado y orientación. Y se obtienen a través del siguiente cálculo:

$$H_1 = \mu_{20} + \mu_{02} \quad \text{Ecuación 3.22}$$

$$H_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11} \quad \text{Ecuación 3.23}$$

$$H_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \quad \text{Ecuación 3.24}$$

$$H_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \quad \text{Ecuación 3.25}$$

$$H_5 = (\mu_{30} - 3\mu_{12}) \cdot (\mu_{30} + \mu_{12}) \cdot \left[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + 3\mu_{03})^2 \right] \\ + (3\mu_{21} - \mu_{03}) \cdot (\mu_{21} + \mu_{03}) \cdot \left[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + 3\mu_{03})^2 \right] \quad \text{Ecuación 3.26}$$

$$H_6 = (\mu_{20} - \mu_{02}) \cdot [(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11} \cdot (\mu_{30} + \mu_{12}) \cdot (\mu_{21} + \mu_{03}) \quad \text{Ecuación 3.27}$$

$$H_7 = (3\mu_{21} - \mu_{03}) \cdot (\mu_{30} + \mu_{12}) \cdot [(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + 3\mu_{03})^2] + (3\mu_{12} - \mu_{30}) \cdot (\mu_{21} + \mu_{03}) \cdot [3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + 3\mu_{03})^2] \quad \text{Ecuación 3.28}$$

Para que todos los momentos invariantes contribuyan por igual en el momento de la discriminación, y estén en el mismo rango de magnitud, es necesario aplicar un proceso de normalización, el cual se expresa en Ecuación 3.29.

$$H_n = \text{abs} \ln(\text{abs}(H_n)) \quad \text{Ecuación 3.29}$$

Signatura y momento

La signatura (firma de forma) es una representación unidimensional del contorno. Una manera de obtenerla es mediante la representación de la distancia del centro al límite de la figura. Es invariante a traslaciones pero no para rotaciones y cambios de escala, cuando se encuentra un punto característico del contorno a partir del cual comenzar, por ejemplo, el punto más cercano al centro (que sea único) se puede conseguir ser invariante a las rotaciones. También se puede usar una normalización de tamaño para que el valor máximo de la representación sea la unidad.

Para la descripción de la signatura se utilizan los momentos propuestos por Fu et al. (1988), los cuales están definidos en la Ecuación 3.30, para figuras o formas bien diferenciadas sólo se necesitan los primeros momentos para describir su firma de forma.

$$\mu_n(a) = \sum_{i=1}^K (a_i - m)^n p(a_i) \quad \text{Ecuación 3.30}$$

Donde:

$$m = \sum_{i=1}^K a_i p(a_i): \text{ es el valor medio de } a .$$

a : es una variable aleatoria discreta que representa las variaciones de amplitud.

$p(a_i)$: corresponde al histograma.

μ_2 : su varianza.

Descripción del esqueleto

Las microcalcificaciones que son ramificadas indican malignidad, para ello se opta por hacer un seguimiento punto a punto al esqueleto de la microcalcificación y determinar si existe un punto-nodo donde se presenta la ramificación.

El algoritmo consiste en seguir la conectividad del esqueleto utilizando vecindad 8 (Figura 3.14), en una lista se almacenan las posiciones adyacentes (correspondientes a la microcalcificación) al punto origen tal como se muestra a continuación:

- i) Se localiza un primer píxel perteneciente a la microcalcificación y se etiqueta.
- ii) Se verifican cada uno de los 8 vecinos circundantes al punto actual, si el vecino tiene valor de 1 su posición x, y es guardado en una lista.
- iii) La última posición guardada en la lista se toma como la posición actual y es borrado de la lista, se etiqueta y el paso ii) se repite. Este paso se repetirá hasta que no haya un vecino circundante existente con valor 1 para la posición actual.
- iv) Si el tamaño de la lista es 0 entonces es una microcalcificación no ramificada. Si el tamaño de la lista es mayor a 0 entonces se obtienen las distancias de las dos últimas posiciones guardadas, si la diferencia entre ellas es mayor a 1 en x y/o y se dice que la microcalcificación es ramificada.

(-1,-1)	(-1,0)	(-1,1)
(0,-1)	(0,0)	(0,1)
(1,-1)	(1,0)	(1,1)

Figura 3.14. Vecindad 8 de un píxel.

Clasificación

En los siguientes apartados es necesario utilizar los datos obtenidos de los descriptores de forma para determinar a qué grupo pertenece, para ello se utilizan los siguientes métodos que permiten el reconocimiento de patrones de información.

El vector de entrada tiene una longitud de 12 las cuales se enumeran a continuación:

1. Compacidad
2. Excentricidad
3. Momento invariante H_1
4. Momento invariante H_2
5. Momento invariante H_3
6. Momento invariante H_4
7. Momento invariante H_5
8. Momento invariante H_6
9. Momento invariante H_7
10. Signatura μ_1
11. Ramificación
12. Área

Las clases de salida para este trabajo son 2 según base de datos MIAS: benignas y malignas; y 5 salidas según la clasificación Le Gal: Tipo I, Tipo II, Tipo III, Tipo IV y Tipo V.

Red multicapa de retropropagación

Para realizar el diseño del Perceptrón Multicapa son importantes la toma de dos decisiones con respecto a las capas ocultas: cuántas capas ocultas debe tener y cuántas neuronas habrá en cada una de las capas. Para muchos casos prácticos no hay una razón para usar más de una capa oculta, sin embargo, redes neuronales con dos capas ocultas pueden representar funciones con cualquier tipo de forma y son requeridas para modelar datos con discontinuidades como una onda diente de sierra. Usar dos capas ocultas raramente mejora el modelo y puede introducir un mayor riesgo de conversión a locales mínimos.

La primera red multicapa de retropropagación propuesta para la clasificación de microcalcificaciones es la siguiente (Figura 3.15), una capa de entrada (12 entradas), una capa oculta (7 neuronas) y una capa de salida (1 neurona), la cual corresponde a la clasificación proporcionada por la base de datos MIAS. Dicha clasificación consiste en 2 estados de la neurona de salida, 0 para microcalcificaciones benignas y 1 para las microcalcificaciones malignas.

La segunda red propuesta (Figura 3.16) tiene el mismo número de neuronas en la capa de entrada y la capa oculta, mientras que la capa de salida tiene 5 neuronas correspondientes a los tipos de microcalcificaciones según la forma individual propuesta por Le Gal (1984). Cuando las microcalcificaciones correspondan al Tipo I la primera neurona se activará (valor de 1) mientras que las demás permanecerán desactivadas (con valor 0), cuando pertenezcan al Tipo II se activará la segunda neurona mientras las demás estarán desactivadas y así respectivamente para las del Tipo III, IV y V.

El número de neuronas en la capa oculta se calculó que debía estar entre el tamaño de la capa de entrada y de la salida. Para el primer diseño cada una de las neuronas de la capa oculta contiene la Ecuación 3.31 y una función de activación Gaussiana (Ecuación 3.32), mientras que en la capa de salida sus neuronas contienen la Ecuación 3.33 y la función de activación mostrada en la Ecuación 3.34.

$$S(j) = \sum_{i=1}^{12} X(i)W(i, j) \quad \text{Ecuación 3.31}$$

$$h(j) = e^{-S(j)^2} \quad \text{Ecuación 3.32}$$

$$T = \sum_{j=1}^7 h(j)V(j) \quad \text{Ecuación 3.33}$$

$$f = e^{-T^2} \quad \text{Ecuación 3.34}$$

Donde:

X = Vector de entrada a la Red Neuronal.

W = Pesos correspondientes de las neuronas asociadas de la capa de entrada - capa oculta.

V = Pesos correspondientes de las neuronas asociadas de la capa oculta - capa de salida.

f = Salida de la Red Neuronal.

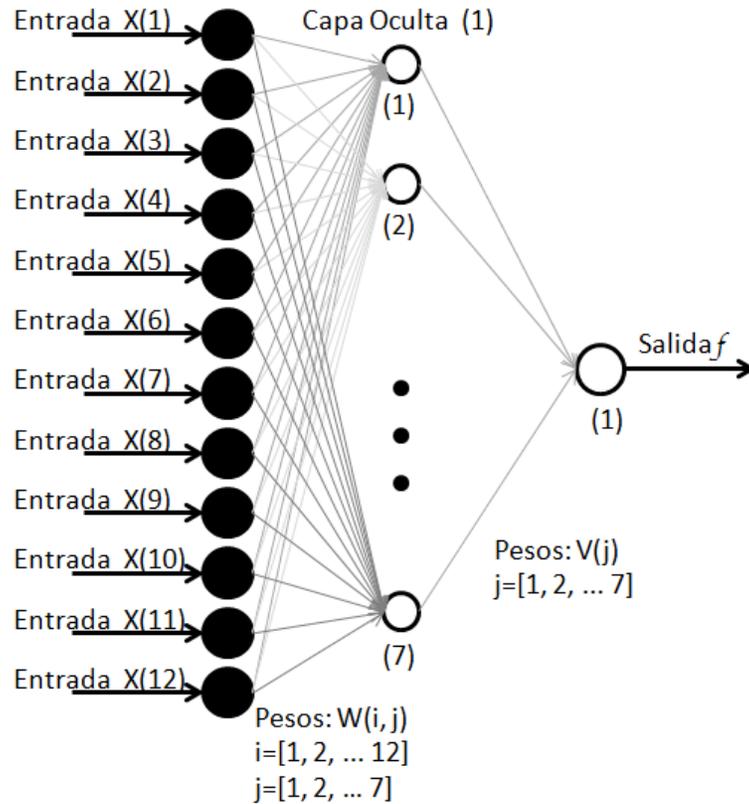


Figura 3.15. Red multicapa de retropropagación con salidas: benigna y maligna.

Para la segunda red neuronal cada una de las neuronas de la capa oculta contiene la Ecuación 3.31 y una función de activación Gaussiana (Ecuación 3.32), mientras que en la capa de salida sus neuronas contienen la Ecuación 3.35 y la función de activación mostrada en la Ecuación 3.36.

$$T(k) = \sum_{j=1}^7 h(j)V(j) \quad \text{Ecuación 3.35}$$

$$f(k) = e^{-T(k)^2} \quad \text{Ecuación 3.36}$$

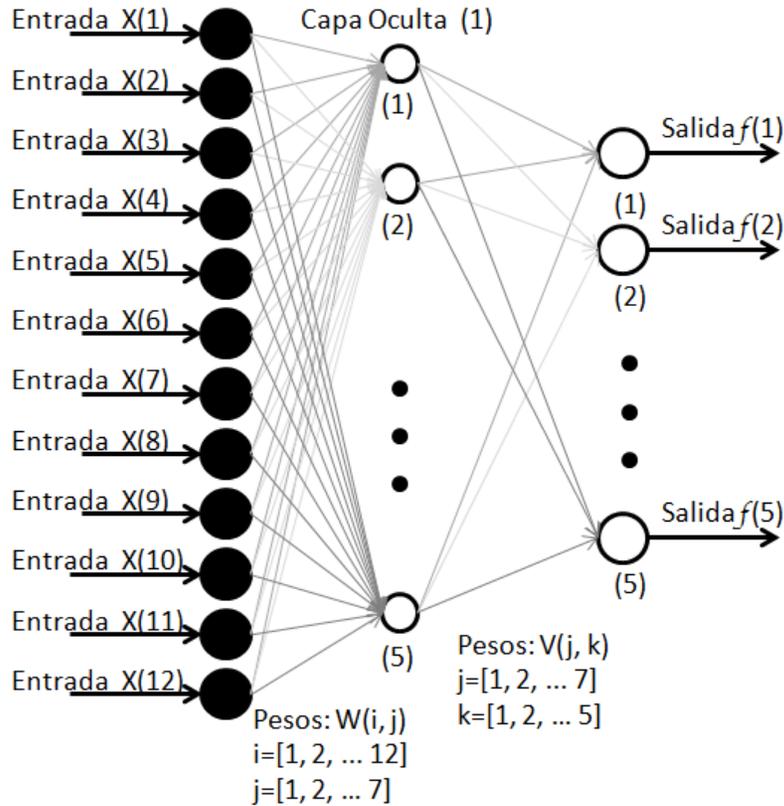


Figura 3.16. Red multicapa de retropropagación con salidas: Tipo I, Tipo II, Tipo III, Tipo IV y Tipo V.

Determinación de la muestra para la validación del algoritmo

Quando se evalúa un método, tratamiento, técnica, tecnología y sistema nuevo de manera aislada, con el objeto de determinar su comportamiento y verificar si funciona de la manera esperada o, por el contrario, no surge efecto alguno, se está frente a un experimento de un solo factor (Lozano Rivas, 2011). La Ecuación 3.37 es el modelo para la determinación del número mínimo de observaciones, de repeticiones o de datos que deben efectuarse u obtenerse para experimentos de un solo factor donde no es posible estimar la varianza de los datos.

$$n = \frac{(W - W^2)(Z_{\beta} + 1.4)(Z\alpha^2)}{W^2} \tag{Ecuación 3.37}$$

Donde:

n = Número mínimo de muestras.

Z_{α} = Valor correspondiente al nivel de confianza.

Z_{β} = Valor correspondiente al poder estadístico.

W = Diferencia mínima observable

En el Cuadro 3.1 y 3.2 se observan los valores de Z_{α} y Z_{β} para los diferentes niveles de confianza y de poder estadístico respectivamente, mientras que en el Cuadro 3.3 se muestra el número mínimo de repeticiones para diferentes diferencias mínimas observables (W) con un poder estadístico del 80%.

Cuadro 3.1. Valores de Z_{α} para diferentes niveles de confianza (Lozano Rivas, 2011).

Nivel de Confianza ($1-\alpha$)		
α	(%)	Z_{α}
0.050	95.0	1.960
0.025	97.5	2.240
0.010	99.0	2.576

Cuadro 3.2. Valores de Z_{β} para diferentes niveles de poder estadístico (Lozano Rivas, 2011).

Poder estadístico ($1-\beta$)		
β	(%)	Z_{β}
0.20	80.0	0.842
0.15	85.0	1.036
0.010	90.0	1.282

Cuadro 3.3. Número mínimo de observaciones para diferentes W , con un poder estadístico $(1-\beta)$ del 80% (Lozano Rivas, 2011).

Número mínimo de observaciones		
Diferencia mínima observable, W (valor porcentual)	N. Confianza del 95% ($1-\alpha$)=0.95	N. Confianza del 99% ($1-\alpha$)=0.99
0.05	245	376
0.1	116	179
0.2	52	80
0.3	30	47
0.4	20	30
0.5	13	20
0.6	9	14
0.7	6	9
0.8	4	5
0.9	3	3
1.0	3	3

Lo que se puede observar en el Cuadro 3.3 es que a un nivel de confianza mayor se demanda un mayor número de muestras, lo mismo ocurre cuando se espera que las diferencias a detectar en la experimentación sean pequeñas. Para la validación del algoritmo propuesto en este trabajo se propone un nivel de confianza del 99%, un poder estadístico del 80% y ya que la sensibilidad que se quiere demostrar es del 90% la diferencia mínima observable es de 0.1, por lo tanto el número de muestras requeridas es de 116.

Evaluación del algoritmo

Para evaluar este trabajo de tesis se requiere de conocer de dos parámetros importantes los cuales son: sensibilidad y especificidad. El primero es la probabilidad de clasificar correctamente un caso como enfermo y el segundo es la probabilidad de clasificar correctamente a un individuo como sano. Dichas probabilidades se estiman de según las ecuaciones 3.38 y 3.39 respectivamente.

$$sensibilidad = \frac{VP}{VP + FN} \quad \text{Ecuación 3.38}$$

$$especificidad = \frac{VN}{VN + FP} \quad \text{Ecuación 3.39}$$

En el Cuadro 3.4 se detallan las variables utilizadas para las ecuaciones 3.34 y 3.35 según sean los diagnósticos obtenidos por el algoritmo en comparación con el diagnóstico MIAS. Cuando el algoritmo diagnostique una microcalcificación maligna cuando en realidad lo es entonces se habla de un Verdadero Positivo, en caso contrario, sería un Falso Positivo; cuando el algoritmo diagnostique microcalcificación benigna y al compararla con el diagnóstico MIAS resulte una verdadera esta afirmación será un Verdadero Negativo, de otro modo será un Falso Negativo.

Cuadro 3.4. Comparación de diagnósticos y descripción de variables VP, VN, FP y FN.

		Diagnóstico MIAS	
		Microcalcificación maligna	Microcalcificación benigna
Diagnóstico algoritmo	Microcalcificación maligna	VP – Verdadero Positivo	FP – Falso Positivo
	Microcalcificación benigna	FN – Falso Negativo	VN – Verdadero Negativo

IV. RESULTADOS Y DISCUSIÓN

Clasificación benigno/maligno según MIAS

Primeramente se evaluó el rendimiento del primer modelo de clasificación que se muestra en la Figura 3.15, el cual consiste en clasificar en dos estados a las microcalcificaciones, esto es en benignas y malignas según el diagnóstico MIAS. Al realizar la etapa de aprendizaje de la red neuronal se pudo observar que no podía converger ya que existen en la base de datos MIAS microcalcificaciones de la misma morfología clasificadas como benignas y malignas en diferentes casos, esto debido a que el diagnóstico está dado para toda la mamografía y no para cada microcalcificación de manera individual.

De la Figura 4.1 a la Figura 4.5 se muestran microcalcificaciones con similitud morfológica y que el diagnóstico MIAS las clasifica en diferentes categorías, en los Cuadros correspondientes a las Figuras anteriormente mencionadas se publican los resultados obtenidos por los descriptores para cada uno de los casos presentados, en el que se observa la similitud en los valores de las variables y por lo tanto resulta imposible la diferenciación entre cada uno.

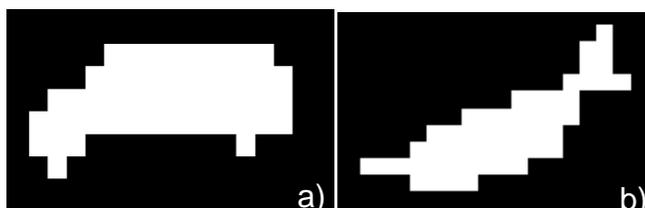


Figura 4.1. a) Microcalcificación mdb219.pgm 1 CALC B 548 258 10 y b) Microcalcificación mdb245.pgm 1 CALC M 593 772 10. (Nota: M = Maligno, B=Benigno).

Cuadro 4.1. Valores de descriptores para las microcalcificaciones con similitud morfológica (Tipo I Le Gal). (Véase Cuadro 4.6 para la nomenclatura de los descriptores y el apartado Obtención de las microcalcificaciones y extracción individual del Capítulo 3 de esta tesis).

52

	C	E	H1	H2	H3	H4	H5	H6	H7	Á	R
mdb219.pgm 1 CALC B 548 258 10	0.3742	0.6866	0.8926	0.2408	5.0148	6.9028	12.2104	8.8159	13.5305	43	0
mdb245.pgm 1 CALC M 593 772 10	0.3058	0.6926	0.8673	0.6182	4.0723	5.4009	10.5059	7.5678	10.1878	45	0

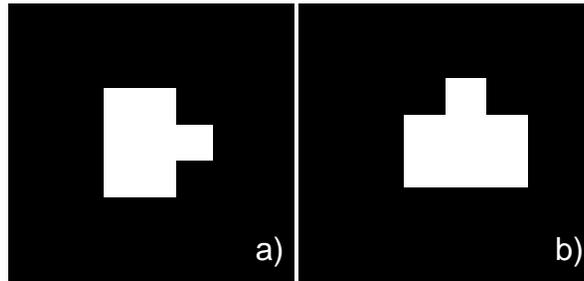


Figura 4.2. a) Microcalcificación mdb222.pgm 2 CALC B 399 607 10 y b) Microcalcificación mdb245.pgm 2 CALC M 666 745 10. (Nota: M = Maligno, B=Benigno).

Cuadro 4.2. Valores de descriptores para las microcalcificaciones con similitud morfológica (Tipo II Le Gal). (Véase Cuadro 4.6 para la nomenclatura de los descriptores y el apartado Obtención de las microcalcificaciones y extracción individual del Capítulo 3 de esta tesis).

53

	C	E	H1	H2	H3	H4	H5	H6	H7	Á	R
mdb222.pgm 2 CALC B 399 607 10	0.8796	0.0059	1.8865	8.9029	7.1273	12.5434	17.7735	16.9948	20.0000	7	0
mdb245.pgm 2 CALC M 666 745 10	0.8796	0.0059	1.8865	8.9029	7.273	12.5434	22.337	16.9948	20.0000	7	0

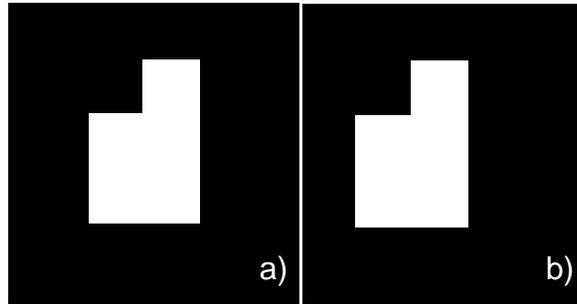


Figura 4.3. a) Microcalcificación mdb233.pgm 3 CALC M 673 837 5 y b) Microcalcificación mdb219.pgm 3 CALC B 544 273 5. (Nota: M = Maligno, B=Benigno).

Cuadro 4.3. Valores de descriptores para las microcalcificaciones con similitud morfológica (Tipo III Le Gal). (Véase Cuadro 4.6 para la nomenclatura de los descriptores y el apartado Obtención de las microcalcificaciones y extracción individual del Capítulo 3 de esta tesis).

54

	C	E	H1	H2	H3	H4	H5	H6	H7	Á	R
mdb233.pgm 3 CALC M 673 837 5	0.7757	0.2500	1.8326	2.3870	6.7663	8.9635	17.0274	11.4892	18.4016	5	0
mdb219.pgm 3 CALC B 544 273 5	0.7757	0.2500	1.8326	2.3870	6.7663	8.9635	17.0274	11.4892	18.4016	5	0

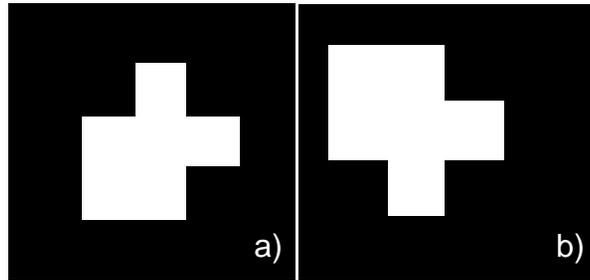


Figura 4.4. a) Microcalcificación mdb233.pgm 4 CALC M 561 627 5 y b) Microcalcificación mdb226.pgm 4 CALC B 292 414 5. (Nota: M = Maligno, B=Benigno).

55 Cuadro 4.4. Valores de descriptores para las microcalcificaciones con similitud morfológica (Tipo IV Le Gal). (Véase Cuadro 4.6 para la nomenclatura de los descriptores y el apartado Obtención de las microcalcificaciones y extracción individual del Capítulo 3 de esta tesis).

	C	E	H1	H2	H3	H4	H5	H6	H7	Á	R
mdb233.pgm 4 CALC M 561 627 5	0.9308	0.0865	1.8489	2.3795	8.0549	11.2738	18.3732	14.3465	17.7601	6	0
mdb226.pgm 4 CALC B 292 414 5	0.9308	0.0865	1.8489	2.3795	8.0549	11.2738	18.3732	14.3465	17.7601	6	0

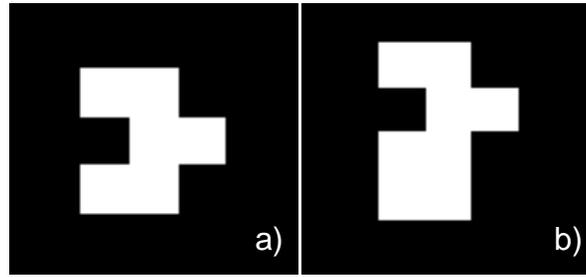


Figura 4.5. a) Microcalcificación mdb216.pgm 5 CALC M 382 510 5 y b) Microcalcificación mdb219.pgm 5 CALC B 531 281 5. (Nota: M = Maligno, B=Benigno).

Cuadro 4.5. Valores de descriptores para las microcalcificaciones con similitud morfológica (Tipo V Le Gal). (Véase Cuadro 4.6 para la nomenclatura de los descriptores y el apartado Obtención de las microcalcificaciones y extracción individual del Capítulo 3 de esta tesis).

	C	E	H1	H2	H3	H4	H5	H6	H7	Á	R
mdb216.pgm 5 CALC M 382 510 5	0.7540	0.0291	1.6617	6.8587	5.9755	9.1944	20.0000	12.6237	20.0000	6	0
mdb219.pgm 5 CALC B 531 281 5	0.6981	0.2538	1.5562	2.9530	6.5299	9.8025	19.0445	12.1554	20.3890	8	0

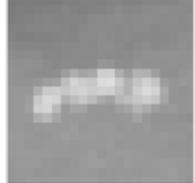
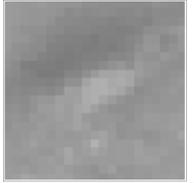
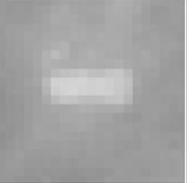
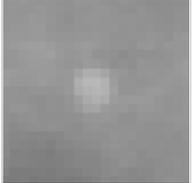
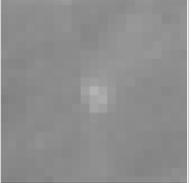
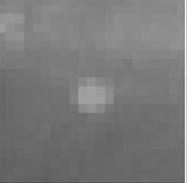
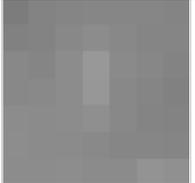
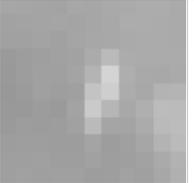
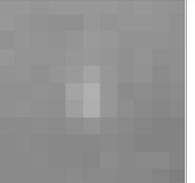
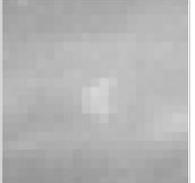
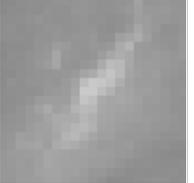
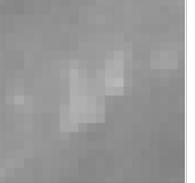
Cuadro 4.6. Nomenclatura de los descriptores utilizados en los Cuadros 4.1 al 4.5.

Símbolo	Nomenclatura
C	Compacidad
E	Excentricidad
H1 a H7	Momentos de Hu
A	Área
R	Ramificación

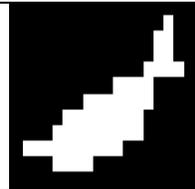
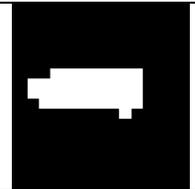
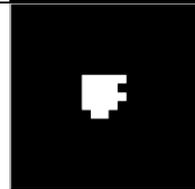
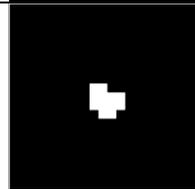
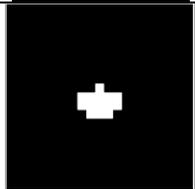
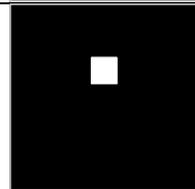
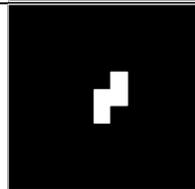
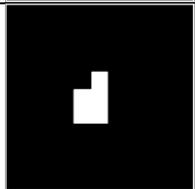
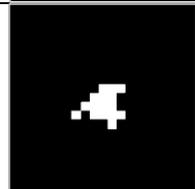
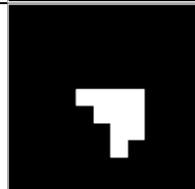
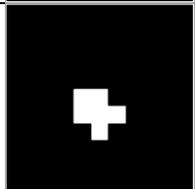
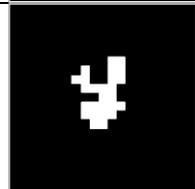
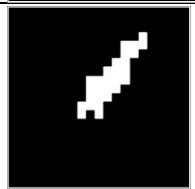
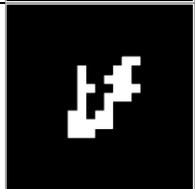
Clasificación benigno/maligno según Le Gal

Ahora la clasificación se realizó mediante la similitud de forma de las microcalcificaciones de acuerdo a como se presenta en el Cuadro 2.2 y en la Figura 2.4 (clasificación Le Gal), donde las microcalcificaciones del Tipo III al Tipo V son las más sospechosas de malignidad, así que se consideran a estos tipos como malignas y a las del Tipo I y Tipo II como benignas por ser poco propensas a procesos de malignidad. De lo anterior se muestran 3 microcalcificaciones de ejemplo de cada tipo según la clasificación (Cuadro 4.7) y en el Cuadro 4.8 son mostradas dichas microcalcificaciones en su forma binarizada para su procesamiento.

Cuadro 4.7. Ejemplos de microcalcificaciones según su tipo (Clasificación Le Gal).

Clasificación según Le Gal	Ejemplo 1	Ejemplo 2	Ejemplo3
TIPO I			
TIPO II			
TIPO III			
TIPO IV			
TIPO V			

Cuadro 4.8. Ejemplos de microcalcificaciones binarizadas según su tipo (Clasificación Le Gal).

Clasificación según Le Gal	Ejemplo 1	Ejemplo 2	Ejemplo 3
TIPO I			
TIPO II			
TIPO III			
TIPO IV			
TIPO V			

Al utilizar el primer diseño de red neuronal basado en el párrafo anterior donde solamente tenemos 2 salidas de acuerdo al grado de sospecha de malignidad se obtuvieron diferentes resultados de acuerdo a los pesos utilizados, de los cuales se destacan las pruebas donde la sensibilidad es mayor del 90% con lo que se aprueba la hipótesis. Los resultados de sensibilidad y especificidad para

esta red se muestran en el siguiente Cuadro (Cuadro 4.9) y sus respectivas variables de evaluación se encuentran del Cuadro 4.10 al Cuadro 4.13.

Cuadro 4.9. Resultados obtenidos usando diferentes pesos en la red que se muestra en la Figura 3.15.

Red neuronal	Sensibilidad	Especificidad
1 (Véase Cuadro A.2 y A.3)	95.51%	44.44%
2 (Véase Cuadro A.4 y A.5)	94.38%	48.15%
3 (Véase Cuadro A.6 y A.7)	93.26%	51.85%
4 (Véase Cuadro A.8 y A.9)	80.90%	70.37%

Cuadro 4.10. Variables de evaluación obtenidas por la red neuronal 1 del diseño de la Figura 3.15.

Variables	Resultados
Verdaderos positivos	85
Verdaderos negativos	12
Falsos positivos	15
Falsos negativos	4
Total microcalcificaciones	116

Cuadro 4.11. Variables de evaluación obtenidas por la red neuronal 2 del diseño de la Figura 3.15.

Variables	Resultados
Verdaderos positivos	84
Verdaderos negativos	13
Falsos positivos	14
Falsos negativos	5
Total microcalcificaciones	116

Cuadro 4.12. Variables de evaluación obtenidas por la red neuronal 3 del diseño de la Figura 3.15.

Variables	Resultados
Verdaderos positivos	83
Verdaderos negativos	14
Falsos positivos	13
Falsos negativos	6
Total microcalcificaciones	116

Cuadro 4.13. Variables de evaluación obtenidas por la red neuronal 4 del diseño de la Figura 3.15.

Variabes	Resultados
Verdaderos positivos	72
Verdaderos negativos	19
Falsos positivos	8
Falsos negativos	17
Total microcalcificaciones	116

De los cuadros anteriores se puede observar el comportamiento de las variables en cada uno de los casos, para obtener una sensibilidad alta es necesario tener el mayor número de verdaderos positivos y pocos falsos negativos como ocurren en los casos 1, 2 y 3, mientras que en el caso 4 donde se tiene una mayor especificidad en comparación con los otros 3 casos se tiene que el número de verdaderos negativos es mayor y se presentan menos casos de falsos positivos. También se observa la relación estrecha que mantienen la sensibilidad y la especificidad en estas pruebas, ya que si se busca una mayor sensibilidad la especificidad se ve afectada de manera importante; el coeficiente de correlación entre estos 2 indicadores obtenido es -0.99151 que es aproximadamente -1 lo que significa que existe una dependencia total entre las dos variables en una relación inversa, cuando una de ellas aumenta la otra disminuye de forma constante.

Como conclusión, el algoritmo cuando utiliza los primeros 3 diseños de red neuronal clasifica a las microcalcificaciones en benignas y malignas con una sensibilidad mayor al 90%, mientras que el algoritmo solo puede diagnosticar correctamente a una persona como sana (especificidad) con un aproximado del 40%-50%, por lo que se comprueba la hipótesis, utilizando los descriptores morfológicos se clasifica a las microcalcificaciones con un 90% de sensibilidad. El mejor resultado considerado por el rendimiento en sensibilidad y especificidad es la red neuronal número 4, ya que puede clasificar a una persona como enferma con un 80.90% y a una persona como sana con un 70.37% lo cual indica que la

precisión de dicha red es del 75.64%, el cual es mayor que las otras 3 redes neuronales (Véase Cuadro 4.14).

Cuadro 4.14. Precisión obtenida para cada una de las redes neuronales.

Red neuronal	Precisión
1	69.98%
2	71.27%
3	72.56%
4	75.64%

Al realizar el análisis de la influencia de cada uno de los descriptores en la clasificación se obtiene que al no incluir cualquiera de las características el algoritmo no mejora su rendimiento, por lo que se concluye junto con el coeficiente de correlación obtenido que es necesaria la utilización de más descriptores para utilizar la red neuronal de retropropagación como clasificador, ya que no hay suficiente información para discriminar a las microcalcificaciones con un 80%-90% de precisión. También se puede optar por cambiar el clasificador utilizando los descriptores actuales y determinar su precisión.

Clasificación por tipos según Le Gal

El segundo diseño de red neuronal que se muestra en la figura 3.16 no discrimina correctamente a las microcalcificaciones en 5 tipos debido a la alta similitud entre los grupos de microcalcificaciones y a la falta de información, dando constantemente una sensibilidad de 100% y una especificidad del 0%. Es decir, existe un número de verdaderos positivos diferente de cero (clasificación correcta como paciente enfermo) y existen cero casos de falsos negativos (diagnóstico incorrecto para una persona enferma), lo que proporciona una sensibilidad del 100%; pero también se encuentra un número de casos diferente de cero de falsos positivos (diagnóstico incorrecto para una persona sana) y cero verdaderos

negativos (clasificación correcta como un paciente sano) dando una especificidad del 0%.

Conclusiones

Se cumple la hipótesis a partir del criterio de clasificación benigno/maligno según Le Gal con una sensibilidad mayor al 93.26%, la desventaja de tener una sensibilidad alta en esta clasificación repercute en la especificidad, por tal motivo, se optó que la red neuronal número 4 utilizado en dichas pruebas obtuvo el mejor rendimiento.

Los descriptores morfológicos que se implementaron fueron de importancia, ya que al excluir a alguno de ellos en la clasificación benigno/maligno según Le Gal produjo una disminución en las variables de evaluación del algoritmo; también a partir del análisis de los descriptores se pudo observar la similitud morfológica de las microcalcificaciones con diferente diagnóstico según MIAS, y las semejanzas según la clasificación por tipos de Le Gal (Véase Cuadros 4.1 al 4.5). Por lo anterior el clasificador en el caso de diagnóstico MIAS no pudo ser implementado con éxito, mientras que en la clasificación basada en Le Gal pudo ser implementado en 2 redes neuronales presentadas en esta tesis, de los cuales el que tuvo 5 salidas no tuvo resultados eficientes al ser evaluado con la especificidad.

Para la clasificación surge también la necesidad de evaluar a las microcalcificaciones morfológicamente con la estandarización BIRADS (estandarización comúnmente utilizada para el diagnóstico de las mamografías), para ello es necesaria la obtención de una nueva base de datos con sus respectivos diagnósticos.

LITERATURA CITADA

- Ballesteros, F., A. Oropesa, L. Martin and D. Andina. 2002. Mammography classification using wavelets, vol. 13. In 5th Biannual World Automation Congress, 293–300.
- Bharathi, P. and P. Subashini. 2011. Optimization of Image Processing Techniques using Neural Networks—A Review. *Wseas Transactions on Information Science and Applications*, ISSN. 8, 300–328.
- Christoyianni, I., A. Koutras and E. Dermatas. 2002. Computer aided classification of mammographic tissue using independent component analysis, vol. 1. In 14th International Conference on Digital Signal Processing ({DSP}), 163–166 vol.1.
- Delogu, P., M. E. Fantacci, A. P. Martinez, et al. 2005. A scalable system for microcalcification cluster automated detection in a distributed mammographic database. *IEEE Nuclear Science Symposium Conference Record*, 2005. 3.
- Docusse, T. 2009. Microcalcification border characterization. *IEEE Engineering in Medicine and Biology Magazine*. 41–43.
- Engan, K., T. O. Gulsrud, K. F. Fretheim, B. F. Iversen and L. Eriksen. 2006. A Computer Aided Detection (CAD) System for Microcalcifications in Mammograms - MammoScan. *World Academy of Science, Engineering and Technology*. 13, 279 – 290.
- Fu, K. S., R. C. González and C. S. G. Lee. 1988. *Robótica: Control, detección, Visión e Inteligencia*. Madrid: McGraw-Hill.
- Le Gal, M., G. Chavanne and P. D. 1984. Diagnostic value of clustered microcalcifications discovered by mammography (apropos of 227 cases with histological verification and without a palpable breast tumor. *Bull Cancer*. 57–64.
- Guo, Q., J. Shao and V. Ruiz. 2009. Characterization and classification of tumor lesions using computerized fractal-based texture analysis and support vector machines in digital mammograms. *International journal of computer assisted ...* 4, 11–25.
- Heywang - Köbrunner, S., D. D. Dershaw and I. Scheer. 2001. *Diagnostic Breast Imaging. Mammography, Sonography, Magnetic Resonance Imaging, and Interventional Procedures*. (2da edition). Thieme New York.

- Hu, M. K. 1962. Visual Pattern Recognition by Moment Invariants. IEEE Trans. Inform. Theory. 8, 179 – 187.
- Karahaliou, A. N., I. S. Boniatis, S. G. Skiadopoulos, et al. 2008. Breast cancer diagnosis: analyzing texture of tissue surrounding microcalcifications. IEEE transactions on information technology in biomedicine: a publication of the IEEE Engineering in Medicine and Biology Society. 12, 731–8.
- Karahaliou, A. 2008. Breast cancer diagnosis: analyzing texture of tissue surrounding microcalcifications. IEEE Transactions on Information Technology in Biomedicine. 12, 731–738.
- Khademi, A., F. Sahba and A. Venetsanopoulos. 2009. Multiresolution Analysis of Mammogram Lesions. ICIAR 2009, LNCS 5627. 802–813.
- De Lafontan, B., J. P. Daures, B. Salicru, et al. 1994. Isolated clustered microcalcifications: diagnostic value of mammography--series of 400 cases with surgical verification. Radiology. 190, 479–483.
- Lee, S., C. Lo, C. Wang, et al. 2000. A computer-aided design mammography screening system for detection and classification of microcalcifications. International journal of medical informatics. 60, 29–57.
- Lozano Rivas, W. A. 2011. DETERMINACIÓN DEL NÚMERO MÍNIMO DE OBSERVACIONES EN INVESTIGACIÓN, OBVIANDO LAS ESTIMACIONES DE LA VARIANZA DE DATOS. Revista de Didáctica Ambiental. 10, 54–61.
- Mousa, R., Q. Munib and A. Moussa. 2005. Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. Expert Systems with Applications. 28, 713–723.
- Murty, P., T. Sudheer and E. Reddy. 2011. Segmentation of Micro Calcification Clusters in Digital Mammograms Using UIQI. DPPR 2011, CCIS 205. 165–172.
- OMS, O. M. de la S. 2012. Octubre: Mes de Sensibilización sobre el Cáncer de Mama.
- Panda, R., B. Panigrahi and M. Patro. 2009. Feature extraction for classification of microcalcifications and mass lesions in mammograms. IJCSNS International Journal of Computer Science and Network Security. 9, 255–265.
- Papadopoulos, A., D. I. Fotiadis and A. Likas. 2005. Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines. Artificial intelligence in medicine. 34, 141–150.

- Rajkumar, K. K. and G. Raju. 2011. Classification of Mammogram Images Using Discrete. ACC 2011, Part III, CCIS 192. 435–443.
- Ramírez, J. and M. Chacón. 2011. Redes neuronales artificiales para el procesamiento de imágenes, una revisión de la última década. RIEE&C. 9.
- Ricci, P., A. Cruz, M. Rodríguez, et al. 2006. Trabajos Originales MICROCALCIFICACIONES BIRADS 4 : EXPERIENCIA DE 12 AÑOS. REV CHIL OBSTET GINECOL. 71, 388–393.
- Simily, J. and K. Balakrishnan. 2011. Local Binary Patterns, Haar Wavelet Features and Haralick Texture Features for Mammogram Image Classification Using Artificial Neural Networks. ACITY 2011, CCIS 198. 107–114.
- Siong, T. S., N. Ashidi, M. Isa and Z. M. Nordin. 2009. The Determination of the Number of Suspicious Clustered Micro Calcifications on ROI of Mammogram. IVIC 2009, LNCS 5857. 232–242.
- Smith RA, Cokkinides V, Brooks D and Brawley OW. 2010. Cancer screening in the United State, 2010: a review\nof current American Cancer Society guidelines and issues\nin cancer screening., vol. 60(2). *CA Cancer J Clin*, 90–119.
- Suckling, J., S. Astley, B. D, et al. 1994. The Mammographic Image Analysis Society Digital Mammogram Database. International Congress Series 1069. Excerpta Medica. 375–378.
- Verma, B. and J. Zakos. 2001. A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques. IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society. 5, 46–54.
- Völk, K., J. Miller and S. Smith. 2009. Multiple network CGP for the classification of mammograms. Applications of Evolutionary Computing. 5484, 405–413.
- Wroblewska, A., P. Boniński, A. Przelaskowki and M. Kazubek. 2003. Segmentation and feature extraction for reliable classification of microcalcifications in digital mammograms. Optoelectronics. 11, 227–235.
- Xu, W., S. Xia and D. Duan. 2006. A Novel Computer-Aided Diagnosis System of the Mammograms. Intelligent Computing in Signal Processing and Pattern Recognition. 345/2006, 639–644.
- Yoon, S. and S. Kim. 2009. AdaBoost-based multiple SVM-RFE for classification of mammograms in DDSM. BMC medical informatics and decision making. 9 Suppl 1, S1.

APÉNDICE

APÉNDICE

Cuadro A.1. Microcalcificaciones utilizadas de la base de datos MIAS para la validación del algoritmo.

1	mdb209.pgm 3 CALC M 689 532 5
2	mdb209.pgm 4 CALC M 646 516 5
3	mdb209.pgm 5 CALC M 621 512 5
4	mdb209.pgm 5 CALC M 627 563 10
5	mdb211.pgm 5 CALC M 676 709 10
6	mdb213.pgm 3 CALC M 540 469 3
7	mdb213.pgm 3 CALC M 556 544 5
8	mdb213.pgm 4 CALC M 524 479 5
9	mdb216.pgm 2 CALC M 502 148 5
10	mdb216.pgm 2 CALC M 507 557 10
11	mdb216.pgm 2 CALC M 709 423 10
12	mdb216.pgm 3 CALC M 504 184 3
13	mdb216.pgm 4 CALC M 392 261 5
14	mdb216.pgm 4 CALC M 443 497 5
15	mdb216.pgm 4 CALC M 688 387 5
16	mdb216.pgm 5 CALC M 382 510 5
17	mdb216.pgm 5 CALC M 531 504 5
18	mdb216.pgm 5 CALC M 545 534 5
19	mdb218.pgm 3 CALC B 523 399 5
20	mdb219.pgm 3 CALC B 549 282 5
21	mdb219.pgm 4 CALC B 534 264 5
22	mdb226.pgm 4 CALC B 282 419 5
23	mdb226.pgm 4 CALC B 292 414 5
24	mdb226.pgm 4 CALC B 532 310 10
25	mdb231.pgm 3 CALC M 604 492 3
26	mdb233.pgm 1 CALC M 579 683 7
27	mdb233.pgm 2 CALC M 589 763 10
28	mdb233.pgm 2 CALC M 604 595 5
29	mdb233.pgm 2 CALC M 623 591 5
30	mdb233.pgm 2 CALC M 646 682 10
31	mdb233.pgm 2 CALC M 699 833 10
32	mdb233.pgm 3 CALC M 527 715 5
33	mdb233.pgm 3 CALC M 565 628 5
34	mdb233.pgm 3 CALC M 668 809 3
35	mdb233.pgm 3 CALC M 681 832 5
36	mdb233.pgm 3 CALC M 727 771 5

37	mdb233.pgm 4 CALC M 566 640 5
38	mdb233.pgm 4 CALC M 595 719 10
39	mdb233.pgm 4 CALC M 629 836 10
40	mdb233.pgm 4 CALC M 638 661 5
41	mdb233.pgm 4 CALC M 664 656 10
42	mdb233.pgm 4 CALC M 713 837 5
43	mdb233.pgm 5 CALC M 523 698 5
44	mdb233.pgm 5 CALC M 609 749 10
45	mdb233.pgm 5 CALC M 652 693 8
46	mdb233.pgm 5 CALC M 653 846 10
47	mdb233.pgm 5 CALC M 656 844 5
48	mdb233.pgm 5 CALC M 709 822 5
49	mdb236.pgm 2 CALC B 280 212 10
50	mdb238.pgm 3 CALC M 511 469 5
51	mdb241.pgm 2 CALC M 467 329 5
52	mdb241.pgm 3 CALC M 435 370 3
53	mdb241.pgm 3 CALC M 452 358 5
54	mdb241.pgm 3 CALC M 462 349 5
55	mdb241.pgm 3 CALC M 463 358 5
56	mdb241.pgm 5 CALC M 454 342 5
57	mdb241.pgm 5 CALC M 463 350 10
58	mdb245.pgm 2 CALC M 508 820 5
59	mdb245.pgm 2 CALC M 571 873 10
60	mdb245.pgm 2 CALC M 590 847 10
61	mdb245.pgm 2 CALC M 634 799 10
62	mdb245.pgm 2 CALC M 636 915 5
63	mdb245.pgm 2 CALC M 647 922 10
64	mdb245.pgm 2 CALC M 666 745 10
65	mdb245.pgm 2 CALC M 696 771 10
66	mdb245.pgm 3 CALC M 554 848 5
67	mdb245.pgm 3 CALC M 561 868 3
68	mdb245.pgm 3 CALC M 567 865 5
69	mdb245.pgm 3 CALC M 624 893 5
70	mdb245.pgm 3 CALC M 644 765 3
71	mdb245.pgm 3 CALC M 652 733 3
72	mdb245.pgm 3 CALC M 690 729 5
73	mdb245.pgm 3 CALC M 705 762 5
74	mdb245.pgm 4 CALC M 577 853 10
75	mdb245.pgm 4 CALC M 583 821 10
76	mdb245.pgm 4 CALC M 610 761 5
77	mdb245.pgm 4 CALC M 613 788 10
78	mdb245.pgm 4 CALC M 614 759 5
79	mdb245.pgm 4 CALC M 643 830 5

80	mdb245.pgm 4 CALC M 654 740 5
81	mdb245.pgm 4 CALC M 662 700 10
82	mdb245.pgm 4 CALC M 668 813 10
83	mdb245.pgm 4 CALC M 691 900 5
84	mdb245.pgm 5 CALC M 614 897 10
85	mdb245.pgm 5 CALC M 639 870 10
86	mdb245.pgm 5 CALC M 646 772 5
87	mdb245.pgm 5 CALC M 649 825 10
88	mdb245.pgm 5 CALC M 649 839 10
89	mdb245.pgm 5 CALC M 656 755 10
90	mdb245.pgm 5 CALC M 657 753 10
91	mdb245.pgm 5 CALC M 668 894 10
92	mdb245.pgm 5 CALC M 693 725 10
93	mdb248.pgm 1 CALC B 381 425 10
94	mdb248.pgm 4 CALC B 373 429 5
95	mdb249.pgm 1 CALC M 537 502 10
96	mdb249.pgm 1 CALC M 560 492 5
97	mdb249.pgm 2 CALC M 544 406 10
98	mdb249.pgm 4 CALC M 542 531 5
99	mdb249.pgm 4 CALC M 550 514 5
100	mdb249.pgm 5 CALC M 546 430 10
101	mdb249.pgm 5 CALC M 557 519 10
102	mdb249.pgm 5 CALC M 560 527 4
103	mdb249.pgm 5 CALC M 564 421 10
104	mdb249.pgm 5 CALC M 588 343 10
105	mdb252.pgm 2 CALC B 440 653 7
106	mdb252.pgm 3 CALC B 426 642 3
107	mdb252.pgm 3 CALC B 431 659 5
108	mdb252.pgm 3 CALC B 436 661 5
109	mdb252.pgm 3 CALC B 444 654 3
110	mdb252.pgm 4 CALC B 429 646 5
111	mdb252.pgm 5 CALC B 454 663 10
112	mdb253.pgm 2 CALC M 728 459 5
113	mdb253.pgm 3 CALC M 762 456 10
114	mdb256.pgm 1 CALC M 388 524 10
115	mdb256.pgm 2 CALC M 382 521 5
116	mdb256.pgm 3 CALC M 420 511 5

Cuadro A.2. Pesos 1 para la evaluación del diseño 1 de red neuronal (Capa entrada – capa oculta).

W(i,j)	1	2	3	4	5	6	7
1	0.018766	0.011936	0.007024	0.006814	0.009621	0.014772	0.015178
2	0.008636	0.01327	0.010441	0.007716	0.001934	0.012483	0.021008
3	0.02717	0.016293	0.012806	0.012385	0.009776	0.011998	0.036417
4	0.003284	0.012401	0.014088	0.018425	-0.00473	0.021216	8.98E-05
5	7.61E-05	0.043038	0.042125	0.046119	-0.00569	0.041094	0.001172
6	-0.00553	0.060072	0.053357	0.053326	0.004045	0.054857	0.001318
7	-0.00353	0.084118	0.082426	0.088737	0.002159	0.087061	0.002947
8	0.002838	0.067884	0.068426	0.068402	0.013113	0.067585	-0.01013
9	0.000665	0.088444	0.094096	0.086341	-0.01281	0.092349	0.000458
10	-0.00096	0.065769	0.070302	0.071051	0.000729	0.064848	-0.00097
11	0.016191	0.010866	0.013584	0.010262	0.006612	0.006187	0.029224
12	0.01762	0.011897	0.006986	0.006776	0.006749	0.014732	0.010252

Cuadro A.3. Pesos 1 para la evaluación del diseño 1 de red neuronal (Capa oculta – capa salida).

V(j)	1
1	0.339714
2	0.557424
3	0.678095
4	0.705170
5	0.144072
6	0.598827
7	0.349059

Cuadro A.4. Pesos 2 para la evaluación del diseño 1 de red neuronal (Capa entrada – capa oculta).

W(i,j)	1	2	3	4	5	6	7
1	0.018771	0.011936	0.007024	0.006814	0.009621	0.014772	0.015181
2	0.008638	0.01327	0.010441	0.007716	0.001939	0.012483	0.021011
3	0.027171	0.016293	0.012806	0.012385	0.009783	0.011998	0.036428
4	0.003323	0.012401	0.014088	0.018425	-0.0048	0.021216	7.43E-05
5	6.27E-05	0.043038	0.042125	0.046119	-0.00571	0.041094	0.001145
6	-0.00555	0.060072	0.053357	0.053326	0.004051	0.054857	0.001299
7	-0.00353	0.084119	0.082426	0.088737	0.002185	0.087061	0.002971
8	0.00283	0.067884	0.068426	0.068402	0.013145	0.067585	-0.01014
9	0.000661	0.088444	0.094096	0.086341	-0.01282	0.092349	0.000467
10	-0.00096	0.065769	0.070302	0.071051	0.000758	0.064848	-0.00102
11	0.016193	0.010866	0.013584	0.010262	0.006611	0.006187	0.029221
12	0.017615	0.011897	0.006986	0.006776	0.006749	0.014732	0.010243

Cuadro A.5. Pesos 2 para la evaluación del diseño 1 de red neuronal (Capa oculta – capa salida).

V(j)	1
1	0.339715
2	0.557424
3	0.678095
4	0.705170
5	0.144077
6	0.598827
7	0.349060

Cuadro A.6. Pesos 3 para la evaluación del diseño 1 de red neuronal (Capa entrada – capa oculta).

W(i,j)	1	2	3	4	5	6	7
1	0.018778	0.011936	0.007024	0.006814	0.009618	0.014772	0.015191
2	0.008645	0.01327	0.010441	0.007716	0.00195	0.012483	0.021017
3	0.027162	0.016293	0.012806	0.012385	0.00979	0.011998	0.036457
4	0.003392	0.012401	0.014088	0.018425	-0.00495	0.021216	5.35E-05
5	4.86E-05	0.043038	0.042125	0.046119	-0.00579	0.041094	0.001129
6	-0.00559	0.060072	0.053357	0.053326	0.004006	0.054857	0.001319
7	-0.0036	0.084119	0.082426	0.088737	0.002176	0.087061	0.003119
8	0.002772	0.067884	0.068426	0.068402	0.013149	0.067585	-0.01008
9	0.000587	0.088444	0.094096	0.086341	-0.01289	0.092349	0.000574
10	-0.00091	0.065769	0.070303	0.071051	0.000916	0.064848	-0.00108
11	0.016192	0.010866	0.013584	0.010262	0.006612	0.006187	0.029216
12	0.017603	0.011897	0.006986	0.006776	0.006749	0.014732	0.010218

Cuadro A.7. Pesos 3 para la evaluación del diseño 1 de red neuronal (Capa oculta – capa salida).

V(j)	1
1	0.339748
2	0.557424
3	0.678095
4	0.705170
5	0.144117
6	0.598827
7	0.349095

Cuadro A.8. Pesos 4 para la evaluación del diseño 1 de red neuronal (Capa entrada – capa oculta).

W(i,j)	1	2	3	4	5	6	7
1	0.018795	0.011936	0.007024	0.006814	0.009623	0.014772	0.01521
2	0.008651	0.01327	0.010441	0.007716	0.001968	0.012483	0.021021
3	0.027169	0.016293	0.012806	0.012385	0.009815	0.011998	0.036508
4	0.003525	0.012401	0.014088	0.018425	-0.00512	0.021216	1.92E-05
5	0.00023	0.043038	0.042125	0.046119	-0.00586	0.041094	0.001157
6	-0.00545	0.060073	0.053358	0.053326	0.004031	0.054857	0.001417
7	-0.00347	0.084119	0.082426	0.088737	0.002362	0.087061	0.003389
8	0.002891	0.067884	0.068426	0.068402	0.01328	0.067585	-0.00993
9	0.000706	0.088444	0.094096	0.086341	-0.0128	0.092349	0.000761
10	-0.00101	0.065769	0.070303	0.071051	0.001446	0.064848	-0.00146
11	0.016191	0.010866	0.013584	0.010262	0.006618	0.006187	0.029206
12	0.017584	0.011897	0.006986	0.006776	0.006754	0.014732	0.010185

Cuadro A.9. Pesos 4 para la evaluación del diseño 1 de red neuronal (Capa oculta – capa salida).

V(j)	1
1	0.339835
2	0.557424
3	0.678095
4	0.705170
5	0.144217
6	0.598827
7	0.349185