



UNIVERSIDAD TÉCNICA DE AMBATO

FACULTAD DE INGENIERÍA EN SISTEMAS ELECTRÓNICA E INDUSTRIAL

CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES E

INFORMATICOS

Tema:

"Detección automática de lesiones en mamografías usando técnicas de análisis de imágenes"

Proyecto de Trabajo de Graduación Modalidad: TEMI Trabajo Estructurado de Manera Independiente, presentado previo la obtención del título de Ingeniero en Sistemas Computacionales e Informáticos.

SUBLINEA DE INVESTIGACIÓN: Tratamiento de Imágenes

AUTOR: Carlos Bolívar Fiallos Quinteros

PROFESOR REVISOR: Ing. María Gabriela Pérez, PhD.

Ambato – Ecuador

Junio 2015

APROBACIÓN DEL TUTOR

En mi calidad de tutor del trabajo de investigación sobre el tema: “**DETECCIÓN AUTOMÁTICA DE LESIONES EN MAMOGRAFÍAS USANDO TÉCNICAS DE ANÁLISIS DE IMÁGENES**”, del Sr. Carlos Bolívar Fiallos Quinteros, estudiante de la Carrera de Ingeniería en Sistemas Informáticos y Computacionales, de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial, de la Universidad Técnica de Ambato, considero que el informe investigativo reúne los requisitos suficientes para que continúe con los tramites y consiguiente aprobación de conformidad el Art. 16 del Capítulo II, del Reglamento de Graduación para Obtener el Título Terminal de Tercer Nivel de la Universidad Técnica de Ambato.

Ambato, Junio 2015

TUTOR

Ing. María Gabriela Pérez, PHD

AUTORÍA

El presente trabajo de investigación titulado: “**DETECCIÓN AUTOMÁTICA DE LESIONES EN MAMOGRAFÍAS USANDO TÉCNICAS DE ANÁLISIS DE IMÁGENES**”. Es absolutamente original, autentico y personal, en tal virtud el contenido, efectos legales y académicos que se desprenden del mismo son de exclusiva responsabilidad del autor.

Ambato, Junio-2015

Carlos Bolívar Fiallos Quinteros

CI. 1600433971

APROBACIÓN DE LA COMISIÓN CALIFICADORA

La comisión Calificadora del presente trabajo conformada por los señores docentes Ing. Renato Urbina B., Mg e Ing. Silvia Zurita M., Mg, revisó y aprobó el informe Final del trabajo de graduación titulado “**DETECCIÓN AUTOMÁTICA DE LESIONES EN MAMOGRAFÍAS USANDO TÉCNICAS DE ANÁLISIS DE IMÁGENES**”, presentado por el Sr Carlos Bolívar Fiallos Quinteros de acuerdo al Art. 18 del Reglamento de Graduación para obtener el título terminal del tercer Nivel de la Universidad Técnica de Ambato.

Ing. Vicente Morales L., Mg
PRESIDENTE DEL TRIBUNAL

Ing. Renato Urbina B., Mg
DOCENTE CALIFICADOR

Ing. Silvia Zurita M., Mg
DOCENTE CALIFICADOR

DERECHOS DE AUTOR

Autorizo a la Universidad Técnica de Ambato para que haga uso de este trabajo de titulación como un documento disponible para su lectura, consulta y procesos de investigación.

Cedo los derechos de mi trabajo de titulación, con fines de difusión pública, además autorizo su reproducción dentro de las regulaciones de la Universidad Técnica de Ambato.

Carlos Bolívar Fiallos Quinteros

CI: 1600433971

DEDICATORIA

La presente Tesis se la dedico a mi familia que gracias a su apoyo pude concluir mi carrera.

A mis Padres y hermanos por su apoyo y confianza en todo lo necesario para cumplir mis objetivos como persona y estudiante.

A todo el resto de mi familia y amigos que de una u otra manera me han llenado de sabiduría para terminar la tesis. A todos en general por darme el tiempo para realizarme profesionalmente.

Carlos Bolívar Fiallos Q

AGRADECIMIENTO

Primero y antes de nada, dar gracias a Dios por estar conmigo en cada paso que doy, por iluminar mi mente y por haber puesto en mi camino a aquellas personas que han sido mi soporte y compañía durante todo el periodo de estudio como es lo es mi familia. A la Universidad Técnica de Ambato Facultad de Ingeniería en Sistemas ya que en esta institución he logrado realizarme como profesional con sus conocimientos. A todos quienes estuvieron durante mi carrera universitaria, Director y Asesor de Tesis.

Carlos Bolívar Fiallos Quinteros.

ÍNDICE GENERAL

Tema	i
Aprobación del Tutor	ii
Autoría	iii
Aprobación de la Comisión Calificadora	iv
Derechos de Autor	v
Dedicatoria	vi
Agradecimiento	vii
Índice General	viii
Índice de Tablas	xi
Índice de Figuras	xii
Índice de Ecuaciones	xiii
Resumen	xiv
Summary	xvi
Glosario de Términos	xviii
Introducción	xx

CAPÍTULO I El Problema

1.1 Tema de Investigación.....	1
1.2 Planteamiento del Problema.....	1
1.3 Delimitación del problema.....	3
1.4 Justificación.....	4
1.5 Objetivos.....	6
1.5.1 General.....	6

1.5.2 Específicos.....	6
------------------------	---

CAPÍTULO II Marco Teórico

2.1 Antecedentes Investigativos.....	7
2.2 Fundamentación Teórica.....	7
2.2.1 Definición de Mamografía.....	7
2.2.2 Mamografía Selectiva de Detección y una Mamografía de Diagnóstico...	8
2.2.3 Beneficios de las mamografías selectivas de detección.....	9
2.2.4 Daños Posibles de las mamografías de detección.....	9
2.2.5 Sistema de Base de Datos e Informes de Imágenes (BI-RADS).....	10
2.2.6 ¿Qué es la Mamografía Digital? ¿En qué se diferencia de la mamografía convencional?.....	13
2.2.7 Ecografía Mamaria.....	13
2.2.8 Mamografía en 3D.....	14
2.2.9 Otras Tecnologías utilizadas para la detección del cáncer de seno.....	14
2.2.10 Transformada Wavelet.....	15
2.2.11 Transformadas wavelet discreta 2D.....	17
2.2.12 Imágenes de Mamografía.....	18
2.2.13 Paradigma de Desarrollo en Espiral.....	20
2.2.14 MatLab.....	22

CAPÍTULO III Metodología

3.1 Modalidad de la investigación.....	24
3.2 Recolección de Información.....	24
3.3 Procesamiento y análisis de los datos.....	25
3.4 Desarrollo del Proyecto.....	25

CAPÍTULO IV Desarrollo de la Propuesta

4.1 Técnica de análisis de Imágenes.....	27
4.2 Revisión de la Literatura.....	29
4.3 Descripción del Método Propuesto.....	30
4.3.1 Objetivo 1.....	32
4.3.2 Objetivo 2.....	35
4.3.3 Objetivo 3.....	46
4.3.4 Objetivo 4.....	49
4.4 Demostración del Algoritmo.....	54

CAPÍTULO V Conclusiones y Recomendaciones

5.1 Conclusiones.....	57
5.2 Recomendaciones.....	58
Bibliografía o Referencias.....	60
Anexos.....	63
Anexo1: Certificado de Validación del Especialista medico.....	64
Anexo2: Código de Programación y Manual de Usuario.....	65-71

ÍNDICE DE TABLAS

2.1 Clasificación de las etapas de la lesión cancerígena en mamografías.....	11
2.2 Clasificación BI-RADS.....	12
2.3 Tipos de Imágenes.....	20
4.1 Imágenes de la base de datos MI-AS utilizadas en los experimentos.....	33-34
4.2 Clasificación de Microcalcificaciones.....	47
4.3 Resultados de las métricas de validación aplicados a las imágenes de prueba...	53

ÍNDICE DE FIGURAS

1.1 Estructura anatómica de la mama femenina.....	2
2.1 Algoritmo Piramidal de la Transformada Discreta Wavelet Bidimensional.....	17
2.2 Etapas de descomposición wavelet patrón dimensional con 5 niveles de resolución.....	18
2.3 La imagen plana (2D) de la derecha puede presentarse como una superficie. En ella la coordenada z para el punto (x, y) corresponde al brillo que tiene en la imagen plana.....	19
2.4 Modelo en Espiral.....	22
4.1 Diagrama de Bloque del método Aplicado.....	26
4.2 Etapas en el análisis de imágenes Médicas.....	29
4.3 Método Propuesto.....	31
4.4 Resultado de Binarización y Segmentación del are de interés.....	40
4.5 Máscaras empleadas para la máscara Unsharp.....	41
4.6 Resultado de aplicar la máscara Unsharp(Filtro Laplaciano).....	42
4.7 Resultado de aplicar Histograma Lineal y Exponencial.....	43
4.8 Algoritmo de Descomposición y Reconstrucción TW2D.....	44
4.9 Resultado de aplicar TW2D.....	45
4.10 Imágenes Resultantes de aplicar las Familias Wavelet.....	45
4.11 Resultado de aplicar TWD coif2.....	46
4.12 Corte ampliado de una mamografía con múltiples microcalcificaciones (marcadas con recuadros)	48
4.13 Resultado etiquetar las microcalcificaciones detectadas.....	50
4.14 Muestra de la detección de las microcalcificaciones superpuesta en la Imagen filtrada.....	51

ÍNDICE DE ECUACIONES

Ecuación 1 Función Wavelet DWT.....	16
Ecuación 2 Multiresolución Wavelet.....	16
Ecuación 3 Coeficientes Wavelet en dos escalas.....	16
Ecuación 4 Términos en función de escalado	16
Ecuación 5 Filtros paso bajo y paso alto.....	17
Ecuación 6 Función basada en una máscara unsharp.....	41
Ecuación 7 Histograma Lineal.....	42
Ecuación 8 Histograma Exponencial.....	42
Ecuación 9 Función para encontrar k.....	43
Ecuación 10 Función para encontrar la Exactitud.....	52
Ecuación 11 Función para encontrar la Precisión.....	52
Ecuación 12 Función para encontrar la Covarianza.....	52
Ecuación 13 Función para encontrar el Error Cuadrático.....	53

RESUMEN

Este trabajo de investigación presenta una técnica basada en el uso de imágenes de mamografía para extraer automáticamente ciertos descriptores relevantes, que permitan reconocer las patologías mamarias mediante técnicas de análisis de imágenes médicas; en particular, se da solución a problemas existentes en la eliminación de ruido, segmentación de las imágenes, extracción de la región de interés (ROI) y las características que describan las posibles asimetrías entre las regiones de interés (ROI) de ambas mamas. Dicho análisis se puede realizar en múltiples pruebas mastológicas.

La detección eficaz y confiable de microcalcificaciones es el objetivo principal del proceso de clasificación y caracterización de la ROI a partir del uso de las mamografías, ya que su idea fundamental es ofrecer una herramienta de ayuda al diagnóstico médico en etapas tempranas de la aparición de estas patologías de la mama. El conocimiento de los primeros signos y síntomas en la población, permite identificar la enfermedad en estados tempranos, lo que dará lugar a una mejor calidad de vida sobre dicha enfermedad.

Así, el estándar *Breast Imaging Report and Database System* (BI-RADS), desarrollado por el Colegio Americano de Radiología (ACR), constituye un estándar y una herramienta radiográfica para el análisis y la interpretación de lesiones en mamografías. Este estándar permite clasificar la malignidad de las lesiones en mamografías en 7 categorías (el valor 0 indica una mamografía incompleta; y los valores del 1 al 6 indican el incremento en el nivel de gravedad desde el valor 1 para la evaluación negativa hasta el valor 6, que se refiere a malignidad comprobada mediante biopsia). Cada nivel lleva asociadas un conjunto de recomendaciones para los doctores y los pacientes. Tradicionalmente, la interpretación y categorización de las mamografías es realizada de forma manual por parte de los médicos radiólogos.

Existen numerosas técnicas de procesamiento y análisis de imágenes de forma automática que pueden aplicarse en la detección de anomalías en mamografías, como pueden ser nódulos y microcalcificaciones, usadas en la clasificación BI-RADS para el diagnóstico de la gravedad de las lesiones. La Visión Artificial es un área de investigación que tiene por objetivo el reconocimiento de objetos o elementos en imágenes. Las etapas típicas de un sistema de este tipo son: la captura de imágenes, el pre-procesamiento para realzar y filtrar la imagen con objeto de facilitar la posterior detección de regiones de interés en las imágenes (segmentación). Las regiones de interés segmentadas de las imágenes se caracterizan mediante unos descriptores que las representan a la hora de reconocerlas o clasificarlas.

Además, para garantizar la eficiencia de los resultados y la calidad de las mismas, se propone incorporar en el análisis de imágenes, el ciclo de vida de desarrollo de software en espiral, por sus características particulares de múltiples iteraciones del proceso de ciclo de vida. Cada vuelta en la espiral representa una fase del proceso, dicho de otra manera, supone un avance en el proceso de desarrollo. No existen etapas fijas tradicionales, ligadas a actividades como la especificación o diseño, sino que cada vuelta en la espiral determina las actividades a realizar.

Para entender el problema se ha llevado a cabo un amplio proceso de investigación, basado en comprender las técnicas de tratamiento de imágenes, las patologías y los diagnósticos médicos. Una vez comprendido el área en el que se iba a desenvolver el sistema, se llevó a cabo el desarrollo del mismo.

SUMMARY

This research work a technique based on the use of mammography images to automatically extract certain relevant descriptors from the same to recognizing these breast conditions by techniques of medical image analysis is proposed, in particular solution is given to existing problems in noise removal and segmentation of images, extracting the region of interest (ROI) and the characteristics that describe the possible asymmetries between the regions of interest (ROI) of both breasts. Such analysis can be performed on multiple mastologic tests.

The efficient and reliable detection of microcalcifications is the principal objective of the ROI classification and characterization process as a starting stage the mammography's use, since the main idea is to provide a tool to aid medical diagnosis at early stages of the breast pathology. The first signs and symptoms of the symptomatic population allow identify the disease earlier, and durability of life on this disease.

Therefore, the standard *Breast Imaging Report and Database System* (BI-RADS), developed by the American College Radiography (ACR), consist in a radiographic standard and a tool use in the interpretation and analysis in the interpretation of the mammograms diseases. This standard allows to identifier the madness of the disease in seven different categories (level 0 means an incomplete mammography; the 1 to 6 values means the disease increment from the level 1 a negative evaluation to level 6 that points out the proved disease throughout a biopsy). Each level suggest a set of recommendations for doctors and patients. Traditionally, interpretation and classification of mammograms is performed manually by radiologists.

There are a lot of technics for the processing and analysis of automatic generated images that could be applied in the detection of anomalies in the mammographic images, as can be nodules and microcalcifications, these are used in the BI-RADS classification for the level of the diagnostic of sickness level. The machine vision is a research area that has the objective of recognize the objects and elements on images.

The typical stages of this type of systems are: the image capture, pre-processing in order to enhance and filter the image as a way to facilitate the latter detection of anomalies segmentation. The interest images on the segmented regions are characterized by a sort of descriptors that represent the time to recognize and categorize them.

Besides, in order to guaranty the quality and veracity of the results it is being propose to incorporate the spiral life cycle development software, since its particular characteristics as multiple interacting on the live cycle process. Each turn on the spiral represents one stage on the process, in other words, this suppose an advance on the process development. There are not traditional fixed stages, involved in the characterization or design, since each turn on the spiral dictates the tasks to be followed.

In order to understand the problem this research has taken a wide investigation process, based on the understanding of the techniques of images treatment, the pathologies and medic diagnostic. Once understood the field where the system is performs, it took the development process.

GLOSARIO DE TERMINOS Y ACRONIMOS

A continuación se indica la terminología más utilizada en el campo de la imagen médica objeto de este estudio, mediante una relación de la notación y el significado de los acrónimos, que corresponden al nombre original en inglés del que provienen:

CDI	Carcinoma ductal infiltrate
ACR	American College of Radiology
BI-RADS	Breast Imaging Reporting and Database System)
FDA	Food and Drug Administration
3D	3 Dimensiones
MRI	Imagen de Resonancia Magnética
PET	Tomografía por Emisión de Positrones
WT	Transformada de Wavelet
DWT	Transformada Discreta de Wavelet
2D	2 Dimensiones
ROI	Región de Interés
RM	Resonancia Magnética
MIAS	<i>Mammographic Image Analysis Society</i>
RGB	Rojo, Verde, Azul
GROUND TRUTH	Imagen de referencia creada manualmente o con alguna herramienta, por un experto.
GOLD ESTÁNDAR	Técnica diagnóstica que define la presencia de la condición con la máxima certeza.

RUIDO	Información no deseada presente en las imágenes que hacen que se observen de forma visual determinadas características
ARTEFACTOS	Serie de problemas q contienen las imágenes
MAMOGRAFIA	Radiografía de los senos
SOLCA	Sociedad de lucha contra el Cáncer
RAYOS X	Es una forma de Radiación electromagnética
ULTRASONIDO	Señal de audio no captada por el oído humano
NÓDULOS	Pequeña Agrupación de Células
INEC	Instituto Nacional de Estadísticas y Censos
BREAST IMAGING	Cáncer de Mama
OMS	Organización Mundial de la Salud
GLOBOCAN	Proyecto dedicado al estudio en la incidencia en la mortalidad y la prevalencia de los tipos de cáncer
FP	Falsos Positivos
FN	Falsos Negativos
CANNY	Detector de bordes y lleva el nombre de su creador

INTRODUCCIÓN

El cáncer de mama es una de las principales causas de mortalidad en las mujeres tanto en los países del primer mundo como en los que están en desarrollo y representa un gran problema dentro de la salud pública, es así que contar con un programa de detección precoz es muy importante para mantener esta patología bajo control y poder prevenirla, siendo la mamografía una de las opciones más tradicionales para la detección temprana del cáncer de mama en la mujer.

Las calcificaciones mamarias corresponden a depósitos de calcio al interior del tejido mamario. Son hallazgos frecuentes en la mamografía, especialmente en mujeres postmenopáusicas. Si bien, la mayoría de las microcalcificaciones son originadas por patología benigna, algunos patrones agrupados específicos pueden ser causados por patología maligna. Aproximadamente el 55% de los cánceres de mama no palpables presentan microcalcificaciones visibles.

Tener un diagnóstico precoz de la enfermedad puede ayudar a las personas que lo padecen a una recuperación temprana. El número de microcalcificaciones es arbitrario, los especialistas radiólogos tienden a decir que el número mínimo de microcalcificaciones debe ser de cuatro a seis, para ser relevante, ya que un número menor raramente conlleva a la detección de cáncer de mama.

Diferentes métodos y algoritmos computarizados han sido propuestos, para detectar las microcalcificaciones y mejorar el contraste de las imágenes mamográficas, basados en técnicas de operaciones morfológicas, técnicas de segmentación, técnicas de filtrado y análisis y transformada wavelet. Las wavelet son funciones matemáticas que separan los datos en diferentes componentes de frecuencia. Las wavelet se definen por medio de una o varias funciones iniciales llamadas wavelet madre (*analyzing wavelet o mother wavelet*). La transformada Discreta Wavelet es una función ortogonal la cual puede ser aplicada a un grupo finito de datos. La idea principal de la transformada wavelet, es representar funciones en términos de bloques fijos más simples con diferente escalados y posición. Muchas de las veces las

funciones analizadas no dependen de una única variable. En el caso de que el número de variables a tratar sea dos, podremos utilizar la descomposición wavelet 2D o bidimensional. Un claro ejemplo de funciones digitales bidimensionales de gran importancia, viene dado por las imágenes digitales.

Además por las características específicas del sistema y teniendo en cuenta la necesidad de explicar de forma más clara y precisa el proceso que se llevará a cabo en el desarrollo del presente proyecto de graduación, se consideró necesario proponer un método híbrido que integra dos metodologías, el paradigma de desarrollo de software en espiral y el proceso análisis de las imágenes o de los sistemas de visión artificial.

La organización de esta tesis además del presente Capítulo introductorio, consta de los siguientes capítulos:

En el Capítulo 1 se describe la temática que se tratará en el proyecto y el planteamiento del problema relacionado con el cáncer de mama, así mismo la justificación y los objetivos trazados para realizar este trabajo de investigación.

En el Capítulo 2 se describe los antecedentes, la fundamentación teórica.

En el Capítulo 3 se presenta el marco metodológico con el cual obtendremos información relevante para entender, verificar, corregir o aplicar el conocimiento.

En el Capítulo 4 se presenta la nueva técnica de análisis de imágenes propuesta y una descripción de las técnicas que se propone para garantizar la bondad de los resultados mediante los métodos propuestos que se consigue con el desarrollo de cada uno de los objetivos planteados.

En el Capítulo 5 se presentan las conclusiones y recomendaciones de este trabajo. Finalmente se muestran las referencias bibliográficas utilizadas durante el desarrollo de la Tesis.

CAPÍTULO I

EL PROBLEMA

1.1. Tema de investigación

Detección automática de lesiones en mamografías usando técnicas de análisis de imágenes.

1.2. Planteamiento del problema

El cáncer de mama representa una de las enfermedades de cáncer más frecuentes en las mujeres no solo en Ecuador sino también en todas partes del mundo, es por tal razón que una detección rápida del mismo es de suma importancia ya que un pronto diagnóstico evita que la enfermedad se diagnostique en etapas avanzadas y por lo tanto es más probable que se conozca y se tenga éxito en los tratamientos médicos desde las primeras etapas de su aparición. Esta patología puede estar presente tanto en el hombre como en la mujer. Aunque el cáncer es más frecuente en la mujer (un 23% del total de casos de cáncer en el mundo), pues una de cada diez mujeres es susceptible de desarrollarlo a lo largo de su vida.

La primera causa de muerte en la mujer se da precisamente por el cáncer de mama. Afectando generalmente a la mujer entre los 25 y 45 años de edad. Al mismo tiempo, se trata de una enfermedad cuyo diagnóstico precoz podría resultar relativamente sencillo. Aunque estas dos afirmaciones previas parezcan aparentemente contradictorias, en realidad las mujeres de regiones poco desarrolladas de Iberoamérica carecen de la posibilidad de recibir asistencia sanitaria primaria en centros de salud y, además, por razones culturales sienten pudor para requerir ayuda médica ante dolencias relacionadas con este tipo de enfermedad [2].

Además, diferentes estudios han demostrado que la detección del cáncer en su estado inicial es de gran importancia para que aumente la tasa de supervivencia en las pacientes afectadas.

Aunque se puede lograr cierta reducción del riesgo mediante medidas de prevención, esas estrategias no pueden eliminar la mayoría de los cánceres de mama que se registran en los países de ingresos bajos y medios. Las estrategias de detección precoz recomendadas para estos países son el tamizaje mediante mamografía y la autoexploración, sin embargo el primero es muy costoso y se recomienda para los países que cuentan con una buena infraestructura sanitaria y pueden asumir costes de programas a largo plazo. Así, la detección precoz con vistas a mejorar el pronóstico y la supervivencia del cáncer de mama sigue siendo la piedra angular del control de este tipo de cáncer Anderson et al., [3].

Este tipo de cáncer es de dos tipos: el carcinoma ductal (presente en los conductos transportadores de la leche desde la mama hasta el pezón) y el carcinoma lobulillar, en los lobulillos que producen la leche (Fig. 1.1). Es poco frecuente que aparezca en otras regiones de la mama [4]. En el hombre se puede presentar a cualquier edad, pero generalmente se detecta en hombres de 60 a 70 años de edad. No obstante, solo representa menos del 1% de todos los casos. Los tipos de cáncer presentes en el hombre son: Carcinoma ductal infiltrante (CDI): cáncer que se disemina más allá de las células que revisten los conductos de la mama. Este es el más habitual en la mayoría de los casos de hombres con cáncer de mama.

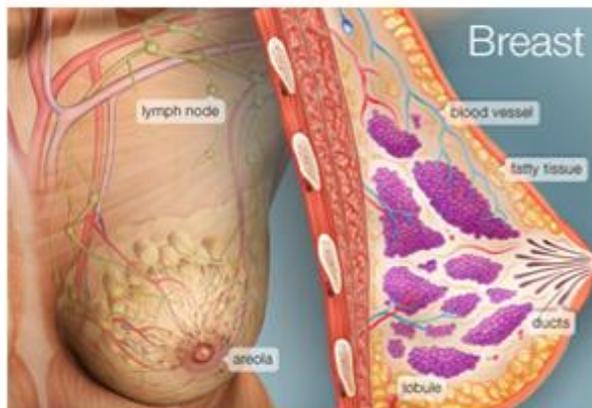


Fig. 1.1. Estructura anatómica de la mama femenina (Tomada de: [5])

En Ecuador el cáncer de mama representa desde hace algunos años, un problema de suma importancia de Salud Pública. Por otro lado, muchas personas no poseen los suficientes recursos económicos para recibir atención temprana y cuando lo hacen es demasiado tarde, muchas veces en estado de emergencia y es ahí cuando con los respectivos análisis, se descubre que la enfermedad de cáncer está en una etapa avanzada. Cada año se presentan 1500 nuevos casos. Según el Ministerio de Salud, los registros hospitalarios de tumores de los hospitales de SOLCA, Quito ocupa el primer lugar después del cáncer de cuello uterino, en cuanto a la incidencia de la patología tumoral maligna en la mujer ecuatoriana [6]. Esta representa una tasa de 35,6% de diagnósticos de cáncer de mama por cada 100.000 habitantes, y para las ciudades de Guayaquil y Cuenca estas tasas son de 26,44% y 31,35%, respectivamente.

El diagnóstico médico basado en diferentes tipos de imágenes como rayos X, ultrasonidos y resonancia magnética, principalmente se ha convertido en una herramienta de gran importancia para la detección a tiempo de las patologías mamarias.

Muchas de las imágenes térmicas permiten analizar los gradientes de temperatura observados en las regiones afectadas de los diferentes pacientes que presentan esta anomalía, es decir, patologías mamarias en su estado precoz. De esta manera, el lugar que se encuentra afectado puede ser detectado con mayor facilidad.

1.3. Delimitación del problema

Área académica: Software

Línea de Investigación: Desarrollo de software.

Sublínea de Investigación: Tratamiento de imágenes.

Delimitación Espacial: La investigación se realizará en la Provincia de Tungurahua en el Hospital SOLCA.

Delimitación Temporal: La investigación se realizará en el periodo de 6 meses después de ser aprobada.

1.4. Justificación

A escala mundial, se diagnostica cada segundo un nuevo caso de cáncer de mama. En Ecuador, cada año se presentan 1500 nuevos casos. En el Congreso Internacional de Senología celebrado en Octubre de 2011 en Guayaquil, se publicó que el cáncer de mama en Ecuador ocupa el segundo lugar después del cáncer de cuello uterino, en cuanto a la incidencia de la patología tumoral maligna en la mujer ecuatoriana [6].

Se pretende contribuir con este proyecto, desarrollando algoritmos que ayuden al especialista radiólogo, a detectar eficazmente las lesiones, usando técnicas de análisis de imágenes en mamografías y de los patrones representativos como pueden ser microcalcificaciones, beneficiando así al hospital SOLCA y a sus médicos, ya que brindará mejores diagnósticos hacia sus pacientes con signos de la enfermedad de cáncer de mama para ser detectada, tratada y prevenirla a tiempo.

Los resultados del proyecto se podrán usar para dar atención a problemas o necesidades locales, como es el caso del diagnóstico precoz de patologías mamarias usando imágenes térmicas. El proyecto contribuirá al fortalecimiento de la línea de investigación e innovación científica sobre el “análisis automático de imágenes mamográficas para detección de patologías mamarias y su aplicación al diagnóstico clínico”.

En relación con el Plan Nacional del Buen Vivir (PNBV) para el periodo 2013-17, este proyecto está directamente relacionado con los siguientes objetivos:

- Auspiciar la igualdad, la cohesión y la integración social y territorial en la diversidad.
- Mejorar las capacidades y potencialidades de la ciudadanía.
- Mejorar la calidad de vida de la población.

A continuación, se señalan algunos datos relacionados con la incidencia del cáncer de mama en Ecuador (Fuente: Diario “El Telégrafo”, edición impresa del día 20/10/12).

“Durante el año 2010 se registró que un total de 470 mujeres fallecieron por cáncer de mama en todo el país. Esto ubica a la patología como la decimocuarta causa de mortalidad

en las mujeres, y como la tercera por cáncer después del de útero (con 722 casos) y del de estómago (con 679), según el registro de Estadísticas Vitales del Instituto Nacional de Estadísticas y Censos (INEC).

Según el Ministerio de Salud, los registros hospitalarios de tumores de los hospitales de SOLCA, donde se recaba toda la información respecto de los tipos de cáncer en el país, documentan que en Quito se tiene una tasa de 35,6% de diagnósticos de cáncer de mama por cada 100.000 habitantes, y para las ciudades de Guayaquil y Cuenca estas tasas son de 26,44% y 31,35%, respectivamente. Actualmente, la mamografía que utiliza la tecnología del “*BreastImaging*” es un sistema de control de calidad que presenta un reporte radiográfico uniforme que puede ser interpretado por cualquier médico. Sin embargo, algunos expertos en mastología, consideran que la única vía de detección temprana de esta patología es la autoexploración y la consulta médica al menos una vez al año. Según el Ministerio de Salud, del total de las 26.786 mujeres fallecidas por diferentes causas en el Ecuador durante el año 2010, el 1,8% correspondió a consecuencia del cáncer de mama, con una tasa de 6,6% por cada 100.000 habitantes.

Los datos recabados por la Organización Mundial de la Salud (OMS), a través de la Agencia Internacional de Investigación del Cáncer, Globocan2008, colocan al cáncer de mama como la tercera causa de mortalidad entre las mujeres en el país, con 628 mujeres fallecidas. “En el país un aproximado de 31 mujeres de cada 100.000 son diagnosticadas de cáncer de mama cada año, y de estas 6 o 7 fallecen como consecuencia de la enfermedad”, afirmó el Ministerio de Salud en un comunicado. [7]

1.5. Objetivos

1.5.1. Objetivo general

- Desarrollar algoritmos para la detección automática de lesiones en mamografías usando técnicas de análisis de imágenes.

1.5.2. Objetivos específicos

- Analizar las imágenes de la base de datos utilizada.
- Definir los algoritmos para identificación de las características a detectar en las imágenes.
- Cuantificar las propiedades de las características identificadas.
- Evaluar la calidad de la detección automática realizada

CAPÍTULO II

MARCO TEÓRICO

2.1. Antecedentes Investigativos

Dra. Esther Ristori Bogajo con un tema referente a Segmentación interactiva y el análisis morfométrico de nódulos en mamografía, investigación realizada en el año 2000, trabajo previo a la obtención de tesis doctoral. [8]

José Manuel Mossi García con un tema correspondiente al Análisis de microcalcificaciones en mamografías, investigación realizada en el año 1998 trabajo que se encuentra en el repositorio de la biblioteca de la Universidad Politécnica de Valencia. [9].

Pedro José García Laencina con el tema detección de microcalcificaciones en mamografías digitalizadas mediante la aplicación de arquitecturas neuronales, investigación realizada en el año 2004, trabajo que se encuentra en el repositorio de la Universidad Politécnica de Cartagena[10].

2.2. Fundamentación Teórica

2.2.1. Definición de mamografía

Una mamografía es una radiografía de los senos. Es una manera de realizar una revisión médica para el control del cáncer de seno. Las mamografías permiten encontrar tumores que son demasiado pequeños como para detectarlos en un examen de los senos. Una mamografía periódica se conoce como mamografía de revisión médica para el control del cáncer de seno. Las mujeres que no presentan síntomas deben hacerse mamografías

periódicas. Una mamografía periódica brinda a los proveedores de atención de la salud una idea de cómo son sus senos normalmente. Cada mamografía ayudará a percibir cualquier alteración que se haya producido desde el último examen.

A veces, se recomienda la realización de una mamografía después de un examen de los senos. Ésta recibe el nombre de mamografía de diagnóstico. Se podría realizar una mamografía de diagnóstico en caso de detectarse un nódulo u otras alteraciones durante un examen de los senos. Las mamografías son efectivas, aunque pueden tener ciertas limitaciones. Es posible obtener resultados imprecisos. En algunos casos, una mancha sospechosa en una mamografía puede terminar siendo algo que no representará mayor preocupación. En otros casos, se puede pasar por alto un diagnóstico de cáncer de seno [11].

2.2.2. Mamografía selectiva de detección y una mamografía de diagnóstico

Una mamografía selectiva de detección es una radiografía del seno que se usa para detectar cambios en el seno en mujeres que no presentan signos o síntomas de cáncer de mama. Se dice selectiva porque ordinariamente se elige tener esta prueba de detección por razones de prevención y no porque haya algún síntoma.

Una mamografía de diagnóstico es una radiografía del seno que se usa para ver si la paciente tiene cáncer de mama después de haberse encontrado un bulto u otro signo o síntoma de cáncer de seno. Los signos de cáncer de mama pueden ser el dolor, engrosamiento de la piel, secreción del pezón, o un cambio en el tamaño o forma del seno. La mamografía de diagnóstico puede usarse también para evaluar los cambios que se hayan encontrado en una mamografía selectiva de detección o para ver el tejido de seno cuando es difícil obtener una mamografía selectiva de detección debido a circunstancias especiales, como la presencia de implantes en el seno. Una mamografía de diagnóstico se lleva más tiempo que una mamografía selectiva de detección porque requiere más radiografías para obtener vistas del seno desde varios ángulos. [12]

2.2.3. Beneficios de las mamografías selectivas de detección

La detección temprana de cáncer de seno con la mamografía selectiva de detección significa que el tratamiento puede empezarse más pronto en el curso de la enfermedad, posiblemente antes de que se haya extendido. Los resultados de estudios clínicos aleatorizados y de otros estudios indican que la mamografía de detección puede ayudar a reducir el número de muertes por cáncer de seno entre mujeres de 40 a 70 años de edad, especialmente para las que tienen más de 50 años de edad. Sin embargo, los estudios llevados a cabo hasta la fecha no han indicado que haya un beneficio de las mamografías de detección regulares en mujeres menores de 40 años de edad o de mamografías de detección como base (mamografías que se usan como punto de comparación) que se toman antes de los 40 años de edad.

2.2.4. Daños posibles de las mamografías de detección

La detección temprana del cáncer no siempre hace que sea menor la posibilidad de una mujer de morir por cáncer de seno. Aunque las mamografías pueden detectar tumores malignos que no se pueden palpar, tratar un tumor pequeño no siempre significa que la mujer no morirá por el cáncer. Un cáncer de crecimiento rápido o muy maligno puede haberse diseminado ya a otras partes del cuerpo antes de ser detectado. Las mujeres con ese tipo de tumores viven un largo período de tiempo sabiendo que probablemente tienen una enfermedad mortal.

Resultados negativos falsos (NF). Estos ocurren cuando las mamografías parecen normales aun cuando el cáncer de seno está presente. En general, las mamografías de detección no detectan hasta 20% de los cánceres de seno presentes cuando se hace esta prueba.

La causa principal de los resultados negativos falsos es la alta densidad del seno. Los senos están formados tanto de tejido denso (es decir, tejido glandular y tejido conjuntivo, lo cual en conjunto se conoce como tejido fibroglandular) como de tejido adiposo. El tejido adiposo aparece de color oscuro en las mamografías, mientras que el tejido fibroglandular aparece como zonas blancas, ya que el tejido fibroglandular y los tumores tienen una

densidad semejante, puede ser más difícil detectar los tumores en mujeres con senos más densos. Los resultados negativos falsos ocurren con más frecuencia entre mujeres jóvenes que entre mujeres de mayor edad ya que las mujeres jóvenes tienen más probabilidad de tener senos densos. Conforme envejece la mujer, sus senos se hacen, en general, más grasos, por lo que son menos probables los resultados negativos falsos. Los resultados negativos falsos pueden conducir a que se retrase el tratamiento y a generar un sentido falso de seguridad en las mujeres afectadas.

Resultados positivos falsos (FP). Los resultados positivos falsos ocurren cuando los radiólogos concluyen que las mamografías son anormales pero en realidad no hay cáncer presente. Después de toda mamografía anormal se deberán hacer pruebas adicionales (mamografías de diagnóstico, ecografías, una biopsia) para determinar la presencia de cáncer.

Los resultados positivos falsos son más comunes en mujeres jóvenes, en mujeres que han tenido antes biopsias de seno, en mujeres con antecedentes familiares de cáncer de seno y en mujeres que toman estrógeno (por ejemplo, como terapia hormonal para la menopausia). Los resultados positivos falsos de las mamografías pueden conducir a ansiedad y a otras formas de angustia psicológica en las mujeres afectadas. Las pruebas adicionales que se requieren para descartar la presencia de cáncer pueden también ser costosas, requerir tiempo y causar molestias físicas.

2.2.5. Sistema de Base de Datos e Informes de Imágenes del Seno (BI-RADS)

El Colegio Estadounidense de Radiología (*American College of Radiology, (ACR)*) ha establecido un método uniforme para que los radiólogos describan los resultados de las mamografías. El sistema, llamado BI-RADS (*Breast Imaging Reporting and Database System*), consiste en siete categorías o grados generalizados. Cada categoría del BI-RADS tiene un plan de seguimiento asociado para ayudar a los radiólogos y a otros médicos a manejar adecuadamente el cuidado de las pacientes. Se puede obtener información adicional del BI-RADS en el sitio web: [13].

La mamografía, cuando detecta una lesión sospechosa de cáncer, la clasifica dentro de una categoría de BI-RADS, las primeras etapas I y II son benignas, la III es probablemente benigna, mientras la IV y V aumenta la probabilidad de que sean malignas (Tabla 2.1).

Tabla 2.1. Clasificación de las etapas de la lesión cancerígena en mamografías

Categoría	Evaluación	Seguimiento
0	Requiere una evaluación adicional con imágenes	Requiere más estudios con imágenes para poder asignar una categoría
1	Negativo	Examen normal. Continuar las mamografías regulares de detección (para mujeres mayores de 40 años de edad)
2	Resultado benigno (no canceroso)	Continuar las mamografías regulares de detección (para mujeres mayores de 40 años de edad)
3	Probablemente benigno	Hacerse una mamografía de seguimiento a los seis meses
4	Anomalía sospechosa	Puede requerir biopsia
5	Muy probable que sea maligno(cáncer)	Requiere biopsia
6	Malignidad reconocida, comprobada por biopsia (cáncer)	Biopsia confirma la presencia de cáncer antes de iniciar tratamiento

En este proyecto se propone el uso de imágenes mamográficas para poder extraer automáticamente ciertos criterios BI-RADS. [14]. En muchas ocasiones la mamografía puede revelar lesiones malignas sin que estas se palpen clínicamente. Los criterios más comunes a clasificar serán los siguientes (Tabla 2.2):

Tabla2.2 Clasificación BI-RADS

Criterios BI-RADS	
Nódulos	
	Forma: <ul style="list-style-type: none"> - Redonda - Oval - Lobulada - Irregular
	Márgenes: <ul style="list-style-type: none"> - Circunscritos - Oscurecidos - Mal definidos (indistintos) - Micro lobulados - Especulados
Microcalcificaciones	
	Morfología: <ul style="list-style-type: none"> - Redondas u ovales - Granulares - Punctatas - Amorfas - Lineales en lanza en V ó en Y - Mono o pleomorfismo
	Distribución: <ul style="list-style-type: none"> - Al azar - Lineal - Segmental - Regional - Multifocal o multicéntrica - Difusa

2.2.6. Qué es la mamografía digital? ¿En qué se diferencia de la mamografía convencional en láminas?

Tanto la mamografía digital como la convencional utilizan radiación de rayos X para producir una imagen del seno; sin embargo, la mamografía convencional registra la imagen directamente en una lámina, mientras que en la mamografía digital, una imagen electrónica del seno se almacena como archivo de computadora. Esta información digital se puede mejorar, ampliar o manipular para hacer una evaluación ulterior con más facilidad que la información almacenada en película.

Ya que la mamografía digital permite que el radiólogo ajuste, guarde y recupere electrónicamente las imágenes digitales, la mamografía digital puede ofrecer las siguientes ventajas que no ofrece la mamografía convencional:

- Los profesionales médicos pueden compartir imágenes archivadas electrónicamente, lo cual facilita las consultas a larga distancia entre radiólogos y cirujanos del seno.
- Es posible notar con mayor facilidad diferencias sutiles entre los tejidos normales y los anormales.
- El número de procedimientos de seguimiento necesarios puede ser menor.
- Quizás sea necesario repetir menos imágenes, lo cual reduce el tiempo de exposición a la radiación.

Las mamografías digitales pueden hacerse solo en establecimientos certificados para mamografías convencionales y que han recibido aprobación de la FDA para ofrecer mamografías digitales. El procedimiento para hacerse una mamografía con un sistema digital es el mismo que con una mamografía convencional [14].

2.2.7. Ecografía mamaria

La ecografía es una técnica que usa los ultrasonidos para la formación de una imagen de la parte del organismo a estudio. No utiliza rayos X. Esta exploración no duplica la información obtenida en una mamografía sino que es una evaluación complementaria. Es útil para distinguir estructuras líquidas de sólidas y por lo tanto a menudo se usa como

recurso para detectar quistes (estructuras pequeñas rellenas de líquido) dentro de la mama. La ecografía mamaria no necesita preparación previa. Durante el examen, la mujer descansa en una camilla. Se le coloca una cantidad de gel en la mama, para poder ser examinada. Con un transductor, el radiólogo examina la mama. De este estudio se obtienen varias imágenes [15].

2.2.8. Mamografías en 3D

Las mamografías tridimensionales (3D), también conocidas como tomosíntesis del seno, es una nueva revolución en imágenes mamarias, que consiste en la captura de múltiples imágenes de cada mama, a partir de cortes milimétricos. Esto permite mejorar significativamente los resultados de los estudios, reduciendo la posibilidad de falsos positivos. La mamografía tal como la conocíamos hasta ahora, captura una única imagen de cada mama, en la que la superposición de tejidos puede requerir la realización de estudios complementarios para descartar o confirmar una lesión.

La Tomosíntesis 3D permite analizar los tejidos capa por capa, con separaciones milimétricas entre sí, despejando dudas y detectando la aparición de lesiones muy sutiles e incipientes y caracterizar microcalcificaciones. La Tomosíntesis 3D utiliza un equipamiento muy poderoso para convertir imágenes mamarias digitales en una sucesión de capas muy finas, que constituyen una imagen en 3 dimensiones de la mama. [16]

2.2.9. Otras tecnologías utilizadas para la detección del cáncer de seno

El Instituto Nacional del Cáncer está apoyando la creación de varias tecnologías nuevas para detectar tumores en el seno. Esta investigación abarca desde métodos que se están formulando en laboratorios de investigación, hasta los que se están evaluando en estudios clínicos. Los esfuerzos para mejorar la mamografía convencional incluyen exploraciones de mamografía digital, resonancia magnética (MRI), tomografía por emisión de positrones (PET) y la tomografía óptica difusa, la cual utiliza luz en lugar de rayos X para crear imágenes del seno [17].

2.2.10 Transformada Wavelet

La Transformada Wavelet (WT) permite conocer qué frecuencias componen una señal en cada instante de tiempo (a diferencia de la transformada de Fourier la cual realiza una descomposición de las señales en sus componentes frecuenciales pero sin facilitar información temporal de éstas), ya que la WT permite el análisis dentro de intervalos grandes de tiempo en aquellos segmentos en los que se requiere mayor precisión en baja frecuencia, y regiones más pequeñas donde se requiere información en alta frecuencia [18].

En el análisis de imágenes por medio de la transformada Wavelet es necesario hacer una discretización de la misma surgiendo por esto el método *Discreet Wavelet Transform* (DWT) que es una técnica que permite el procesamiento digital de señales discretas e imágenes. Cabe resaltar estas características:

- Aísla y manipula distintos tipos de patrones específicos ocultos en grandes cantidades de datos.
- Comprime o elimina ruido sin degradación apreciable [19].

Los tipos de wavelet son llamadas wavelet madre, entre las cuales tenemos: *Wavelet de Haar, Daubechies, Symlets, Coiflets, BiorSplines, Reverse Bior, Meyer, DMeyer, Gaussian, Mexican_hat, Morlet, Complex Gaussian, Shannon, Frequency B-Spline y Complex Morlet.*

Según la wavelet madre utilizada, se puede seleccionar el grado de la wavelet madre Ej.

Tipo: Daubechies:

Grado: db1, db2, db3, db4, db5, db6, db7, db8, db9, db10, db**.

Según el Artículo “La transformada de Wavelet Discreta” de la universidad de Oviedo [20]:
“Si a una imagen le aplicamos la DWT obtenemos cuatro tipos de coeficientes: aproximaciones, detalles horizontales, detalles verticales y detalles diagonales. La

aproximación contiene la mayor parte de la energía de la imagen, es decir, la información más importante, mientras que los detalles tienen valores próximos a cero.”

La transformada discreta wavelet (DWT) es una función ortogonal que puede ser aplicada a un grupo finito de datos. Este grupo finito de datos puede ser modelado como una función $f(t)$ en términos de contracciones y dilataciones de una función de escalado (*scaling function*) $\phi(t)$ y una función wavelet $\Psi(t)$ Strang y Nguyen [21]; González y Woods [22].

$$f(t) = \sum_k c_k^j 2^{-\frac{j}{2}} \phi(2^{-j}t - k) + \sum_{j=8}^J \sum_k d_k^j 2^{-\frac{j}{2}} \psi(2^{-j}t - k) \quad (1)$$

Una característica importante de las wavelets es la multiresolución: el primer término de la ecuación (1) es una aproximación a $f(t)$ en una escala o resolución J , mientras que el segundo término de la ecuación corresponde a un refinamiento de la función en una escala $j \leq J$. Los demás elementos que aparecen en la ecuación (1) se describen a continuación:

Los c_k^j conocidos como coeficientes de escalado (*scaling coefficients*)

$$c_k^j = 2^{-\frac{j}{2}} \int f(t) \phi(2^{-j}t - k) \quad (2)$$

Los d_k^j , conocidos como coeficientes wavelet, vienen dados por:

$$d_k^j = 2^{-\frac{j}{2}} \int f(t) \psi(2^{-j}t - k) \quad (3)$$

Expresando los términos de función de escalado de la ecuación (1) en dos escalas diferentes, Strang y Nguyen [21], se define;

$$\phi\left(\frac{t}{2}\right) = \sqrt{2} \sum_n h[n] \phi(t - n) \quad (4)$$

$$\psi\left(\frac{t}{2}\right) = \sqrt{2} \sum_n g[n] \psi(t-n) \quad (5)$$

Donde $h[n]$ y $g[n]$ son los filtros paso bajo y paso alto, respectivamente (Strang y Nguyen, [21]).

2.2.11. Transformadas wavelet discretas 2D

La transformada wavelet de dos dimensiones separa las líneas de las columnas, considerando a cada fila y cada columna como una señal uni-dimensional. Desde un punto de vista del procesamiento de imágenes, las wavelets se definen como un conjunto de funciones base que por medio de translaciones y dilataciones operan como un banco de filtros sobre la imagen.

Para obtener la descomposición de una imagen, primero se aplica la transformada wavelet en cada línea unidimensional de la imagen. Esta operación proporciona un coeficiente de medios y un coeficiente de detalle para cada línea. Entonces el tratamiento de estas filas se transforma como si fueran una nueva imagen y se aplica de nuevo la transformada unidimensional a cada columna de la imagen. El resultado de estas operaciones serán todos los coeficientes de detalle, excepto para el primer píxel que corresponde a un único coeficiente de aproximación. En este caso, estamos considerando la aplicación de la transformación para obtener el menor nivel de resolución posible. [23] En la Fig. 2.1, se muestra la gráfica del algoritmo de descomposición wavelet bidimensional.

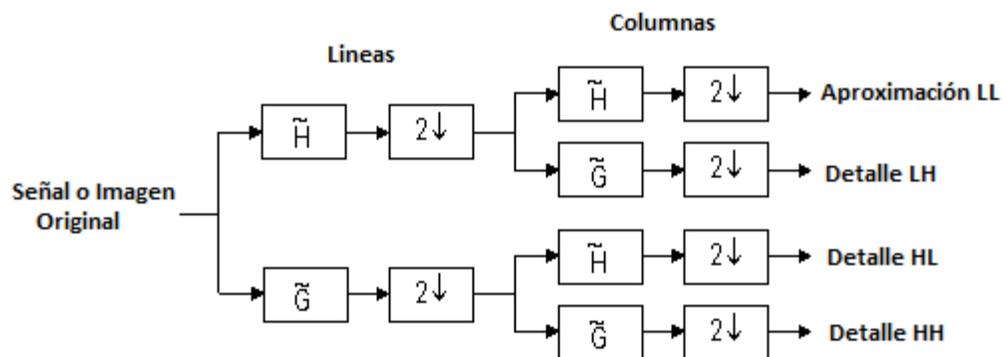


Fig. 2.1 Algoritmo Piramidal de la Transformada Discreta wavelet Bidimensional

El algoritmo primero aplica una descomposición unidimensional sobre una de las dimensiones generando dos subbandas a las que inmediatamente a continuación se les aplica otra descomposición unidimensional sobre la otra dimensión. Así en los casos en los que la señal a tratar sea una imagen, se obtiene otra imagen descompuesta en subbandas como resultado de aplicar la descomposición unidimensional sobre el eje x y al resultado de esta, donde:

- LL (de paso bajo-paso bajo) que corresponde a la banda de paso bajo en ambas direcciones.
- LH (de paso bajo-paso alto) banda correspondiente a la de paso bajo y de paso alto dirección vertical en la dirección horizontal.
- HL (de paso alto-de paso bajo) correspondiente a la de paso de banda alto y de paso bajo dirección vertical en la dirección horizontal, y
- HH (de paso alto-de paso alto) correspondiente a la banda de paso alto en ambas direcciones, respectivamente Fig. 2.2.[23]

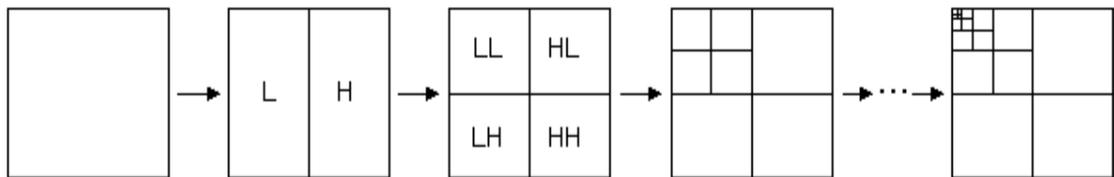


Fig. 2.2 Etapas de descomposición wavelet patrón dimensional con 5 niveles de resolución.

2.2.12. Imágenes de Mamografía

¿Qué es una imagen?

Una imagen bidimensional es una función que a cada par de coordenadas (x, y) asocia un valor relativo a alguna propiedad del punto que representa (por ejemplo su brillo o su matiz). Una imagen acromática, sin información de color, en la que a cada punto se le

asocia información relativa al brillo, se puede representar como una superficie (ver Fig. 2.3), en la cual la altura de cada punto indica su nivel de brillo.

Una imagen en color RGB se puede representar asociando a cada punto una terna de valores que indica la intensidad de tres linternas (una roja, otra verde y otra azul). Una imagen de color de espectro completo se puede representar asociando a cada punto un diagrama espectral de emisión de color [24].

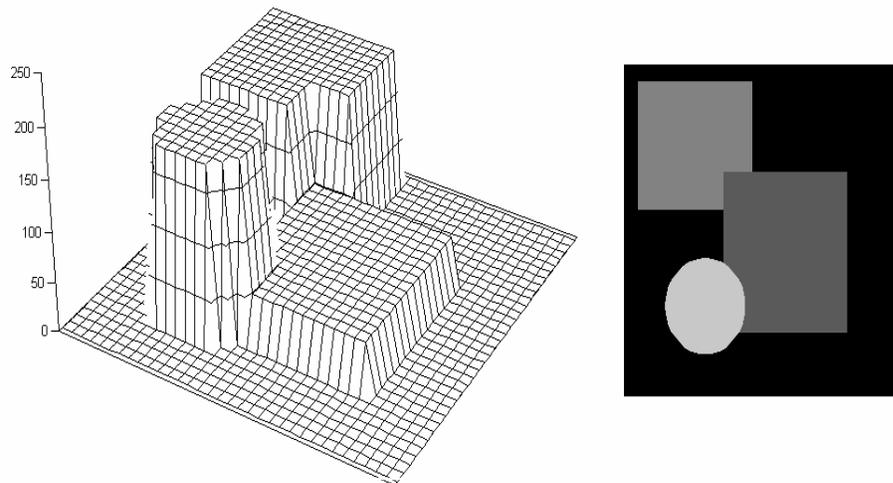


Fig. 2.3. La imagen plana (2D) de la derecha puede presentarse como una superficie. En ella la coordenada z para el punto (x, y) corresponde al brillo que tiene en la imagen plana

Las imágenes de mamografías (Tabla 2.3), se proponen para extraer la información clínica que permitan realizar el diagnóstico y seguimiento de enfermedades de mama. Su objetivo es proporcionar la información asociada a los estados del cáncer de mama. Después de la adquisición, estas imágenes pueden ser sometidas a un riguroso análisis cuantitativo que a menudo requiere la extracción automática robusta, precisa, y fiable de las distintas etapas de patologías.

Este trabajo de investigación aborda varios problemas presentes en este contexto del análisis y cuantificación de las imágenes de mamografía, siendo la microcalcificación unas de las principales razones para poder detectar el cáncer a tiempo.

La segmentación de las distintas patologías es necesaria para las actividades clínicas por parte del especialista en los estudios de patologías presentes. Las imágenes son sometidas a un procesamiento que busca mejorar su calidad antes de avanzar a las siguientes etapas clásicas de un sistema de visión artificial. [25]. Existen muchos tipos de imágenes y de formatos, pero los más comunes son los siguientes (Tabla 2.3).

Tabla 2.3 Tipos de Imágenes

Tipo imagen (<i>Image type</i>)	La relación definida entre los valores de un array y los colores de los píxeles. La toolbox soporta imágenes binarias, indexadas, de intensidad y RGB.
Imagen binaria (<i>Binary image</i>)	Imagen que contiene sólo píxeles de color blanco y negro. En MATLAB, una imagen binaria es representada por una matriz de tipo uint8 o double logical que contiene 0's y 1's (los cuales representan blanco y negro respectivamente). Una matriz es logical cuando sus flags lógicas están habilitadas. De forma habitual se nombra a la variable que contiene una imagen binaria como BW.
Imagen de intensidad (<i>Intensity image</i>)	Es una imagen cuyos valores de píxeles corresponden a una escala de grises. En MATLAB, una imagen de intensidad es representada por un array de clase uint8, uint16, o double. Mientras estas imágenes no sean guardadas con mapas de color, MATLAB usará un mapa del sistema para mostrarla. De forma habitual se nombra a la variable que contiene una imagen de intensidad como I. Este término es sinónimo de escala de grises.
Imagen RGB (RGB image)	Una imagen cuyos píxeles son especificados por 3 valores, uno para cada componente de color (rojo, verde y azul) de cada píxel. En MATLAB, una imagen RGB es representada por un array mxnx3 de clase uint8, uint16, o double. De forma habitual se nombra a la variable que contiene una imagen RGB como RGB.
Clase de almacenamiento (<i>Storage class</i>)	Es la clase numérica de almacenamiento usada para guardar una imagen en MATLAB. Las posibles clases que MATLAB utiliza son uint8, uint16 y double.

2.2.13 Paradigma de Desarrollo en Espiral

Propuesto por Boehm en 1988 con la finalidad de paliar los inconvenientes del modelo en cascada y adecuar el desarrollo por prototipos a problemas complejos. El paradigma de espiral es un modelo de ciclo de vida orientado a riesgos que divide un proyecto software

en miniproyectos y donde cada miniproyecto se centra en uno o más riesgos importantes hasta que todos estos estén controlados. Este modelo se realiza en varias iteraciones; se parte de una escala pequeña la cual comienza con la identificación de objetivos, alternativas y restricciones; en medio de la espiral, se localizan riesgos, se genera un plan para manejarlos, y a continuación se establece una aproximación a la siguiente iteración. El primer circuito de la espiral produce el desarrollo de una especificación de productos; los pasos siguientes en la espiral se podrían utilizar para desarrollar un prototipo y progresivamente versiones más sofisticadas del software. Cada paso de la región de planificación produce ajustes en el plan del proyecto.

El costo y la planificación se ajustan según la reacción ante la evaluación del cliente. Además, el gestor del proyecto ajusta el número planificado de iteraciones requeridas para completar el software.

En esencia, la espiral, cuando se caracteriza de esta forma, permanece operativo hasta que el software se retira. Hay veces en que el proceso está inactivo, pero siempre que se inicie un cambio, el proceso arranca en el punto de entrada adecuado. El modelo en espiral es un enfoque realista del desarrollo de sistemas y de software a gran escala. Como el software evoluciona, a medida que progresa el proceso, el desarrollador y el cliente comprenden y reaccionan mejor ante riesgos en cada uno de los niveles evolutivos.

Mantiene el enfoque sistemático de los pasos sugeridos por el ciclo de vida clásico, pero lo incorpora al marco de trabajo interactivo que refleja de forma más realista el mundo real. El modelo en espiral demanda una consideración directa de los riesgos técnicos en todas las etapas del proyecto, y si se aplica adecuadamente, debe reducir los riesgos antes de que se conviertan en problemáticos. [26].

Las actividades de este modelo se conforman en una espiral (Fig. 2.4.), en la que cada bucle o iteración representa un conjunto de actividades. Las actividades no están fijadas a ninguna prioridad, sino que las siguientes se eligen en función del análisis de riesgo, comenzando por el bucle interior.

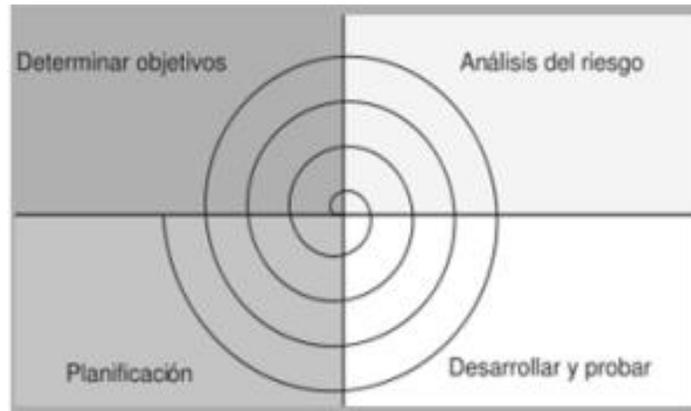


Fig. 2.4. Modelo Espiral

Este enfoque le entrega al proceso una gran capacidad de responder ante eventuales cambios en los requisitos en cualquier etapa del desarrollo del software, y le entrega al análisis de riesgo, un rol fundamental en la toma de decisiones, lo que permitiría mantener acotados los costos y la duración de un proyecto informático. [27]

2.2.14 MatLab

MATLAB (abreviatura de “*MATrix LABORatory*”, laboratorio de matrices) es un software matemático que ofrece un entorno de desarrollo integrado con un lenguaje de programación propio (lenguaje M). Está disponible para las plataformas de Unix, Windows y Apple. Entre sus prestaciones básicas se hallan: la manipulación de matrices, la representación de datos y funciones, la implementación de algoritmos, la creación de interfaces de usuario (GUI) y la comunicación con programas en otros lenguajes. [28]

Es ampliamente usado por Ingenieros de Control en el análisis y diseño, posee además una extraordinaria versatilidad y capacidad para resolver problemas en matemática aplicada, física, química, ingeniería, finanzas y muchas otras aplicaciones. Está basado en un sofisticado software de matrices para el análisis de sistemas de ecuaciones. Permite resolver complicados problemas numéricos sin necesidad de escribir un programa.

MATLAB está siendo utilizado como herramienta de investigación para la resolución de complejos problemas planteados en la realización y aplicación de modelos matemáticos en

ingeniería. Los usos más característicos de la herramienta los encontramos en áreas de computación y cálculo numérico tradicional, prototipaje algorítmico, teoría de control automático, estadística, análisis de series temporales para el proceso digital de señal. [29]

Algunas consideraciones generales:

- MATLAB distingue entre mayúsculas y minúsculas.
- La comilla ' es la que, en un teclado estándar, se encuentra en la tecla de la interrogación.
- Los comentarios deben ir precedidos por % o, lo que es lo mismo, MATLAB ignora todo lo que vaya precedido por el símbolo %.
- La ayuda de MATLAB es bastante útil; para acceder a la misma basta teclear *“help”*. Es recomendable usarlo para obtener una información más precisa sobre la sintaxis y diversas posibilidades de uso de los comandos. [30]

CAPÍTULO III

METODOLOGÍA

3.1. Modalidad de la investigación

La investigación se basará en una metodología de investigación aplicada. Consiste en analizar las imágenes proporcionadas por el especialista médico, para detectar en ellas de forma automática diferentes tipos de síntomas representativos de patologías (ver Tablas 2.1 y 2.2 de este documento). Para ello, una vez diseñados los correspondientes algoritmos, éstos se implementarán en el lenguaje MATLAB. Después, se evaluarán resultados usando un conjunto de imágenes de test. Las modalidades de investigación que se utilizaran son: bibliográfica, de campo y explicativa.

Bibliográfica porque se recurrirá a fuentes de carácter científico para obtener fundamentación teórica. De campo porque el estudio de los hechos es en el lugar donde se producen los acontecimientos. Explicativa porque se dará una solución y se comprobará el funcionamiento de lo que pretende resolver el problema.

3.2. Recolección de Información

Se experimentará con una muestra por 30 imágenes aportadas por el Hospital de SOLCA y la base de datos Mini-MIAS [31].

Para ambos conjuntos de imágenes, se dispone de muestra correspondiente a mujeres sanas y con diferentes grados de patologías. Todas las imágenes tienen asociadas su correspondiente diagnóstico médico.

3.3. Procesamiento y análisis de los datos

En el proyecto se desarrollarán diversos algoritmos de tratamiento de imágenes para el problema planteado (p.ej. algoritmos de filtrado, segmentación y extracción de características). Una vez procesadas las imágenes con dichos algoritmos, se compararán los resultados automáticos con las detecciones manuales proporcionadas por el experto.

3.4. Desarrollo del proyecto

ANALISIS DE IMÁGENES DE LA BASE DE DATOS

1. Para el análisis se pretende realizar la extracción de las características de las imágenes, el proceso consiste en realizar el aumento de contraste con algún filtro, disminuir el ruido, adquisición realce, segmentación, extraer aquellas regiones donde están las anomalías etc.

DEFINICION DE LOS ALGORITMOS PARA LA IDENTIFICACIÓN DE CARACTERISTICAS A DETECTAR EN LAS IMÁGENES.

2. Escoger el algoritmo óptimo para resolver los procesos sobre las imágenes analizadas anteriormente, siguiendo patrones obtenidos por diferentes fuentes.

CUANTIFICAR LAS PROPIEDADES DE LAS CARACTERÍSTICAS IDENTIFICADAS

3. Usar descriptores conocidos en tratamientos de imágenes para saber, por ejemplo, que forma tiene la región. En Matlab, hay algunas funciones para hacer esto. Una es "*regionprops*" que nos da un conjunto de propiedades sobre la forma de una región segmentada en la imagen.

EVALUAR LA CALIDAD DE LA DETECCIÓN AUTOMÁTICA REALIZADA

4. Para la evaluación de la calidad se pueden usar las imágenes del médico y las imágenes que de la Base de Datos Mini-MIAS.

CAPÍTULO IV

DESARROLLO DE LA PROPUESTA

El método que se propone en este trabajo de investigación comprende la fusión de un paradigma de desarrollo de software y las técnicas de análisis de imágenes para brindar solución a problemas reales presentes en cada imagen de la mamografía.

Para ello se desarrollarán los pasos de este nuevo método que se muestran en el diagrama de bloques de la Fig. 4.1. Ambos conjuntos de técnicas podrán aplicarse a los diferentes pasos que se implementaran en el desarrollo de este trabajo. A continuación se introducirán los conceptos fundamentales que intervienen en el presente estudio.

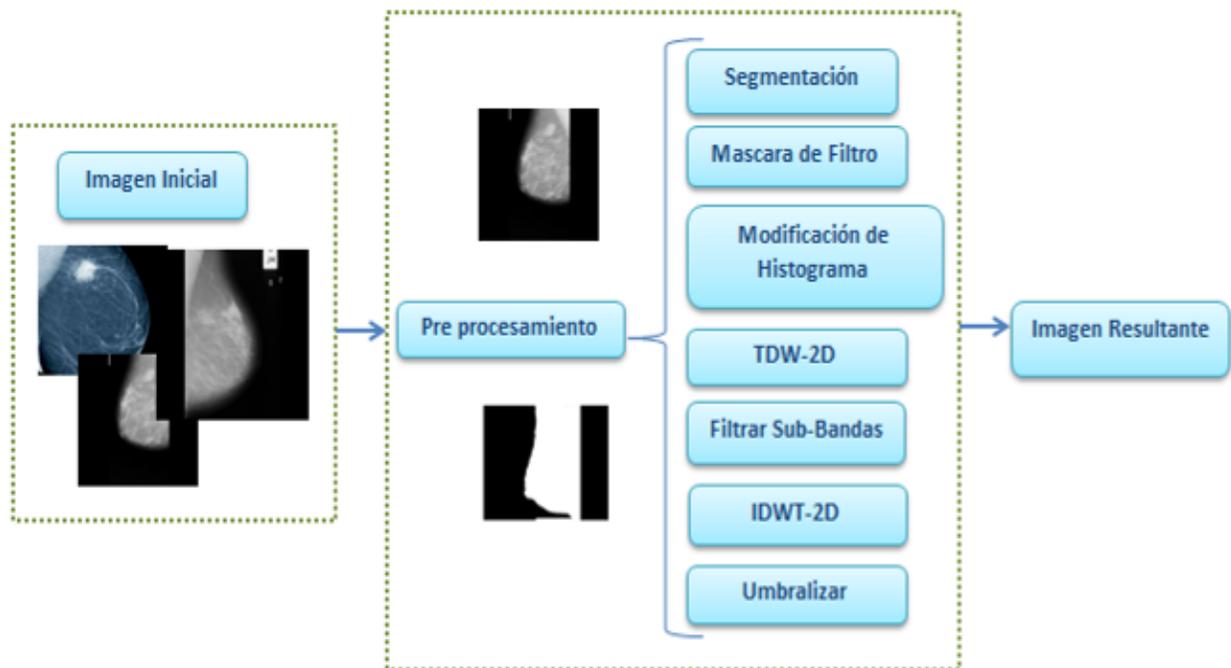


Fig. 4.1. Diagrama de Bloque del método Aplicado

4.1 Técnica de análisis de Imágenes

En el trabajo de investigación se hará una revisión de algunas de las técnicas de procesamiento más utilizadas en el campo de la imagen médica. En particular nos centraremos en el preproceso, la segmentación y la extracción del ROI y las características.

En la actualidad se han desarrollado técnicas que permiten visualizar eficientemente gran cantidad de datos y estudiar la funcionalidad de determinados órganos, como por ejemplo la imagen de mamografía, utilizando las imágenes generadas por Mamógrafos y RM. El desarrollo de programas y procesadores cada vez más rápidos y sofisticados, ha dado lugar a que estas herramientas se estén utilizando cada vez más en los entornos hospitalarios. Así como es el caso del presente trabajo. [25]

En general, las etapas más comunes en el procesamiento de imágenes médicas se encuentran representadas en la Fig. 4.2., y son las siguientes:

- Adquisición de las imágenes de una base de datos o directamente del sistema de adquisición.
- Procesamiento para distintos propósitos entre ellos, mejora o realce de la imagen; extracción del fondo o *background*, realces, filtrado de ruido y artefactos no deseados, normalización, etc.
- Segmentación de las regiones de interés.
- Representación mediante características o caracterización de las regiones segmentadas.
- Clasificación de las características extraídas.
- Verificación o validación de la clasificación para determinar la precisión de la misma.

No siempre se aplican todas, pero en términos generales son las más habituales en este contexto. Estas etapas, pueden agruparse en: el preprocesamiento, el procesamiento y el postprocesamiento (refinamiento). [25]

En la etapa de preprocesamiento se aplican técnicas que contribuyan a mejorar la calidad de la imagen de manera previa a la reconstrucción. Cada modalidad de imagen médica debe ser especialmente tratada con la técnica que más se adecue a sus características específicas Pham [32], y de acuerdo al problema a resolver.

Esta etapa es necesaria pues las imágenes se pueden degradar por múltiples causas (por ejemplo debido a la presencia de ruido) y esto puede ocultar la información que es relevante para el especialista. Cada tipo de posible degradación plantea un problema diferente que requiere un análisis y planteamientos de solución distintos (Basu [33]; Worsley y Friston, [34], [25]).

El éxito en la etapa de preprocesamiento es clave para garantizar el éxito de las subsiguientes etapas del análisis de las imágenes médicas. En particular, la calidad de la segmentación está directamente relacionada con el adecuado tratamiento de los artefactos, por lo que es preciso realizar una validación de las propuestas efectuadas en esta etapa. [25].

Reducir el ruido de las imágenes actualmente sigue siendo un problema fundamental en el campo del procesamiento de las mismas. En la literatura aparecen múltiples propuestas, una de las más populares son las técnicas multirresolución basadas en transformadas Wavelet.

Las wavelets son herramientas que ofrecen un rendimiento superior en el *denoising* de imágenes debido a las propiedades de ortogonalidad, mínima redundancia y dispersión de la información y a su estructura multirresolución [25].

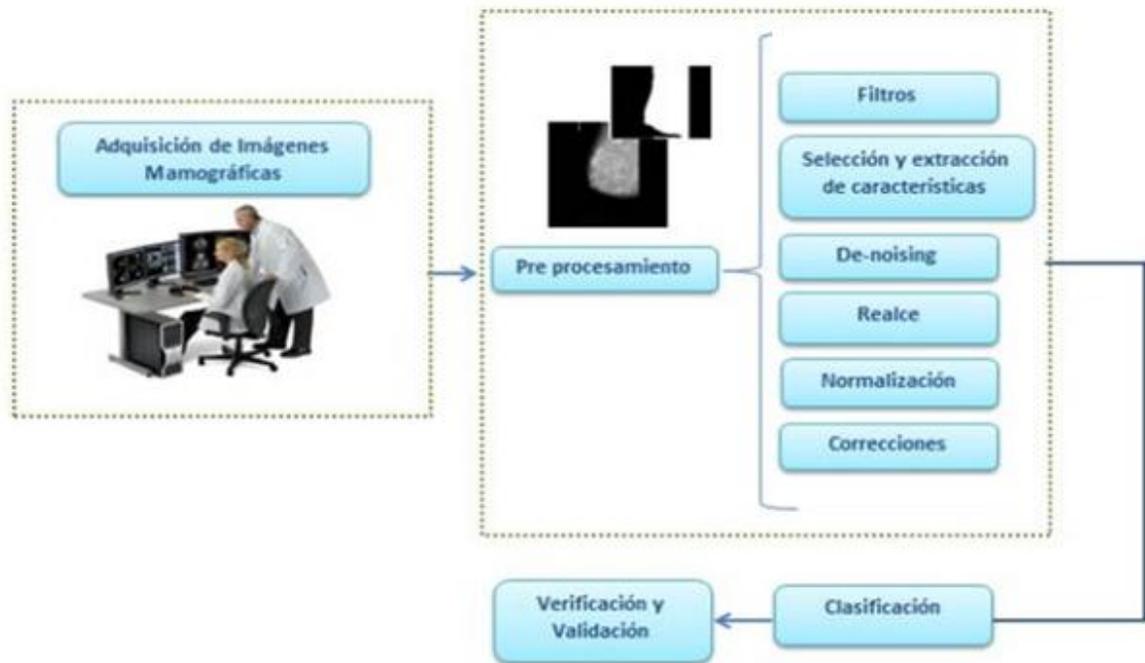


Fig. 4.2. Etapas en el análisis de Imágenes Medicas

4.2 Revisión de la literatura

Una metodología computarizada para identificar microcalcificaciones fue propuesto por Nishikawa et al. [35]. En este artículo se menciona la metodología usada para tal proceso que se compone de tres fases: En primer lugar se disminuye el ruido de la imagen mediante la aplicación de un filtrado. En segundo lugar, se obtiene un umbral de nivel de gris global, para encontrar el umbral adaptivo de nivel de gris se basa en la técnica propuesta por Zhao et al [36], en la cual combina las operaciones de filtrado con una base de reglas; y, en tercer lugar el número de señales falsamente identificadas es descrito por:

- 1) Sondar el espectro de potencia de las señales individuales.
- 2) El cálculo de la distribución espacial de las señales.
- 3) Examinar la relación entre el tamaño la forma y el fondo del valor de cada pixel de microcalcificaciones.

En el artículo “*Detection of Malignancy in Digital Mammograms from Segmented Breast Region Using Morphological Techniques*”, realizado por E. Srinivasa y P. Madhuri, utilizaron las técnicas de segmentación siguientes [37]:

- a) Segmentación basada en umbralización (en inglés, *Threshold Based segmentation*)
- b) Segmentación basada en bordes (en inglés, *Edge based segmentation*)
- c) Técnicas de *Clustering*
- c) *Matching*

Las cuales sirvieron en su trabajo para mejorar la calidad de la imagen original. Otras técnicas usadas y que se mencionan en el mismo fueron la eliminación de ruido usando filtro Gaussiano. También se realiza un realce y la eliminación de información innecesaria en el fondo de la imagen, esto con operaciones morfológicas como Erosión y Sustracción [37].

En otro de los artículos titulado “*Microcalcifications Detection in Mammographic Images Using texture Coding*” manifiesta la utilización de un rango de codificación para reducir los niveles de gris. El objetivo de la primera fase es la reducción del número de niveles de gris de 256 a 9. La segunda parte del proceso que se menciona en el artículo es aumentar el número de niveles de gris del 9 al 16 para introducir la amplitud del píxel.

Todo el proceso aplicado que se menciona en el artículo es para resaltar la información requerida en la textura de las mamografías, codificando sus texturas y eliminando información superficial con el fin de disminuir la cantidad de datos innecesarios y poder analizar la imagen de una mejor manera [38].

4.3 Descripción del método propuesto

La ventaja de integrar el modelo en espiral y las etapas de un sistema de visión artificial, es que, podemos regresar a un proceso anterior, para mejorarlo o corregirlo, siendo así factible ir interactuando de acuerdo a como la imagen se vaya presentando al aplicar las diferentes técnicas de procesamiento. El usar el modelo en espiral se fundamenta principalmente en

múltiples repeticiones del proceso de ciclo de vida, los cuales se pueden dividir en mini proyectos, que se obtendrán en cada giro e interacción de la espiral, la ventaja de utilizar este modelo es que tendremos puntos de control en cada interacción. Cada vuelta en espiral representa una fase del proceso.

En la Fig. 4.3. Se presenta el modelo propuesto, que permite garantizar un mejor control en cada uno de los procesos que se va realizando con la imagen mamográfica, permitiéndonos de esta manera ir evaluando y realizando las pruebas respectivas para que el resultado final sea el óptimo, para la detección de las microcalcificaciones.

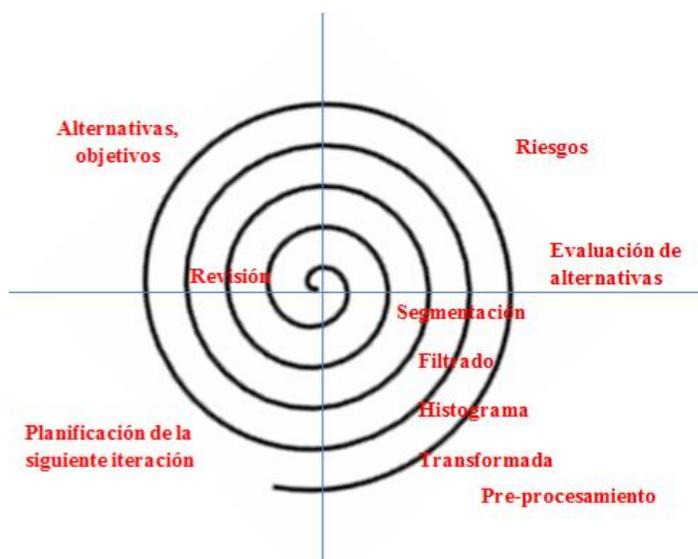


Fig. 4.3. Método Propuesto

Una vez definido las herramientas que se concretaron para desarrollar el trabajo de investigación propuesto, se va a proceder a explicar cómo se cumplieron cada uno de los objetivos planteados y qué actividades se realizaron en cada uno de ellos para alcanzar los resultados esperados. Para desarrollar cada objetivo y garantizar el éxito de los mismos se ha incorporado el ciclo de vida de desarrollo de software en espiral que se explicó en el apartado anterior.

4.3.1 Objetivo1: Analizar las imágenes de la base de datos utilizadas.

Las imágenes mamográficas utilizadas para experimentar los algoritmos propuestos en el desarrollo de este proyecto de tesis, se adquirieron de la base de Datos MI-AS [31]. MIAS es una base de imágenes de acceso libre y que cuenta con un total de 322 imágenes en formato .pgm que se pueden descargar directamente desde la página oficial [31]. Para su fácil almacenamiento el formato en el que se encuentran digitalizadas estas imágenes es el *Portable graymap format* (pgm), formato de archivo en escala de grises.

Utiliza 8 bits por píxel si el valor máximo de gris es de 255, utiliza 16 bits por píxel si el valor es mayor que 255 y menor que 65536. Como característica importante de estas imágenes es el tamaño, que en este caso, es de 1024x 1024 píxeles.

De este total de imágenes se utilizó una muestra de aproximadamente 30 imágenes para las pruebas experimentales. Para conocer cuál de las imágenes presentan malignidad y cuáles no, se tiene una lista detallada de la siguiente manera [31]:

Primera columna:

Número de referencia de base de datos MIAS.

Segunda columna:

Carácter del tejido de fondo:

F Grasos

T Fatty-glandular

D-glandular denso

Tercera columna:

Clase de anormalidad presente:

CALC calcificación

Masas CIRC Well-defined/circumscribed

Masas SPIC espiculada

MISC Otros, masas mal definidas

Distorsión de la arquitectura ARCH

ASYM Asimetría

NORMA normal

Cuarta columna:

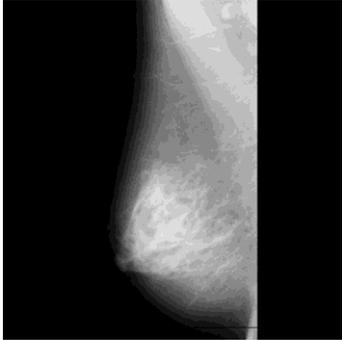
La gravedad de la anomalía;

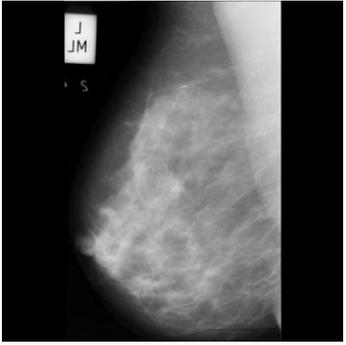
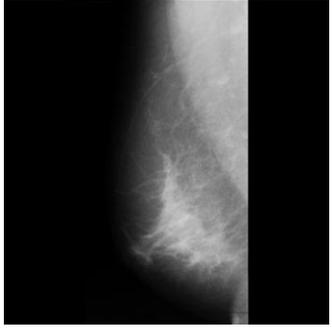
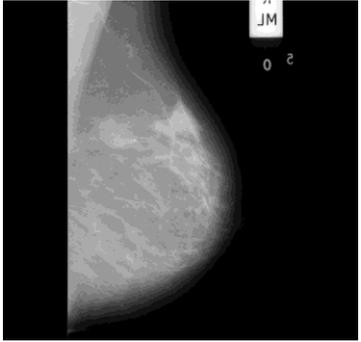
B benigna

M maligno

De esta información detallada, nos valemos para escoger y clasificar las imágenes con las cuales trabajamos para realizar las respectivas pruebas (Tabla 4.1).

Tabla 4.1. Imágenes de la base de datos MI-AS utilizadas en los experimentos

Identificación de la Imagen	Imágenes
<p data-bbox="565 1255 737 1287">Mdb211.pgm</p> <ul style="list-style-type: none"><li data-bbox="305 1310 894 1341">*Carácter de tejido de fondo Graso-Glandular.<li data-bbox="305 1365 894 1396">*Clase de anormalidad presente, calcificación.<li data-bbox="305 1419 846 1451">*La gravedad de la anormalidad, Maligno.	

<p>Mdb219.pgm</p> <ul style="list-style-type: none"> *Carácter de tejido de fondo, Graso – Glandular. *Clase de anomalía presente, calcificación. *La gravedad de la anomalía, Benigna. 	
<p>Mdb233.pgm</p> <ul style="list-style-type: none"> *Carácter de tejido de fondo Graso – Glandular. *Clase de anomalía presente, calcificación. *La gravedad de la anomalía, Maligno. 	
<p>Mdb248.pgm</p> <ul style="list-style-type: none"> *Imagen con Tejido de fondo graso. *La anomalía presente es la calcificación. *Gravedad de la anomalía Benigna. 	

4.3.2 Objetivo2: Definir los algoritmos para identificación de las características a detectar en las imágenes.

El primer paso después de la adquisición de imágenes consiste en realizar el preprocesamiento de la misma y elegir mediante experimentación las técnicas que permitan dar mejor solución al problema, que pueda presentar la imagen, con el fin de optimizarla para posteriores procesos o etapas del análisis. Las técnicas que se suelen aplicar corresponden con cada una de las siguientes etapas básica de un Sistema de Visión Artificial:

La etapa del pre-procesamiento

El pre-procesamiento no es más que el conjunto de técnicas que vamos a utilizar en las imágenes, con el fin de mejorar la calidad y facilitar la búsqueda de la información que necesitamos, en este caso poder detectar las microcalcificaciones. En esta etapa de procesamiento previo es donde, mediante filtros y transformaciones geométricas, se elimina la información que no forma parte de la imagen o se realzan partes que son de interés [25] de la misma. En la sección “Descripción de los pasos efectuados en esta tesis”, se desglosará los pasos del pre-procesamiento que fueron realizados en este trabajo.

La etapa de la segmentación

Una vez que una imagen ha sido tratada adecuadamente después de su adquisición, las siguientes etapas son la partición de la imagen en sus regiones de interés y su posterior evaluación y validación. El principal objetivo que se tiene al aplicar segmentación a una imagen, es dividirla en partes que tengan una fuerte correlación con objetos o áreas contenidas en la imagen, es decir se separan los elementos que interesan de una escena para comprenderla. El proceso de segmentación de una imagen depende del problema que se desee resolver [24]. Las técnicas clásicas son las siguientes:

- **Segmentación basada en umbralización.** La umbralización es un proceso que permite convertir una imagen de niveles de gris o de color en una imagen binaria, de

tal forma que los objetos de interés se etiqueten con un valor distinto al de los píxeles del fondo. Aunque, para rendimiento de los procesos se opta por convertir la imagen en niveles de gris, no obstante la extensión a color es inmediata si sólo se usa uno de las componentes RGB o alguna mezcla de las tres. Las técnicas de umbralización, puesto que se aplican sobre la información proporcionada por el histograma, está basada en la distribución de grises en la imagen y no la información espacial, los métodos de segmentación basados en la umbralización, como único medio de segmentación, resulten limitados en muchos problemas reales. Por ello, la mayoría de las veces se usan acompañados de otros métodos [24]. La elección de un valor de umbral correcto resulta decisivo para llevar a cabo la segmentación de una imagen de manera satisfactoria y es una tarea que el analista de las imágenes debe fijar como un criterio importante. Para ello, existen distintos tipos de umbralización, los más comunes son los basados en un umbral fijo, umbral global, umbral optimo, umbral adaptativo, entre otros muchos. En general, el umbral nos sirve para reducir a dos colores una imagen, el blanco y el negro, y de esta manera poder extraer los objetos con respecto de su fondo. En la sección “Descripción de los pasos efectuados en esta tesis” se explica la técnica aplicada en este trabajo.

- **Segmentación basada en contornos.** La segmentación basada en detección de contornos agrupa un gran número de técnicas que usan la información proporcionada por las fronteras o bordes de los objetos que aparecen en una imagen. Una de las más populares es la detección de bordes basada en el operador *Canny*, detección de contornos con filtros de gradiente, [24]. La detección de bordes nos permite detectar los cambio de intensidad de luz, y de esta manera poder determinar la profundidad, tamaño, orientación y propiedades dentro de la superficie de la imagen.
- **Segmentación basada en regiones.** En este tipo de métodos, por lo general, los algoritmos buscan regiones conectadas con píxeles con alguna característica de similitud, tal como el brillo, la textura, la velocidad de campo, etc. Algunos de los

métodos más populares de este grupo son: umbralización, crecimiento de región y clasificación. Sin embargo aplicando solo estos métodos no es suficiente para resolver los problemas de la detección y segmentación o delineación de los tejidos y estructuras satisfactoriamente. En estos últimos años, muchos investigadores han obtenido avances mediante mejoras de estos métodos basados en región. Estas se pueden clasificar en los siguientes métodos: basados en clasificación, basados en *clustering*, basados en morfología, basados en textura, basados en la extracción de características, basados en fusión, basados en redes neuronales, basados en lógica borrosa y métodos basados en fractales (Clarke et al. [39]; Haralick y Shapiro [40], [25], entre otros. En la sección “Descripción de los pasos efectuados en esta tesis” se describirá el método efectuado en este trabajo.

- **Segmentación basada en crecimiento de regiones.** Estas técnicas determinan zonas dentro de una imagen basándose en criterios de similitud y proximidad entre los píxeles de la misma. En estas técnicas la homogeneidad (o falta de homogeneidad) entre regiones adyacentes es el criterio utilizado para unir (o dividir) regiones de la imagen. Dicha homogeneidad se puede definir a partir de criterios como: el nivel de gris medio, el color, la forma, etc. El resultado de la segmentación es una partición de la imagen en regiones homogéneas.
- **Segmentación basada en morfología matemática.** La morfología matemática se refiere a una rama del procesamiento de imágenes no lineal y análisis que se concentra en la estructura geométrica de objetos o regiones dentro de las imágenes. El concepto básico es buscar una imagen con un elemento estructurante y cuantificar la manera en la cual el elemento estructurante quepa dentro de la imagen. [25]. Las operaciones morfológicas (que son filtros no lineales) simplifican las imágenes y las formas de los objetos y tienen muchas aplicaciones: adelgazar bordes de regiones, separar regiones unidas, unir regiones separadas tras la segmentación, contar el número de regiones en una imagen, etc. Otra son: extraer componentes de la imagen (contorno, esqueleto, cierre convexo...); Filtrado morfológico: adelgazamiento (erosión), dilatación, apertura y cierre de objetos.

De manera general las técnicas clásicas de segmentación que mencionábamos en la anterior sección no son capaces de explotar toda la información proporcionada por la modalidad de imágenes RM. Es por ello que los avances en técnicas de análisis de imágenes han sido y aún siguen siendo un reto importante con el objetivo desarrollar algoritmos que permitan la extracción de los datos de las IRM de forma óptima, dando solución en mayor o menor medida a la problemática asociada con ellas

En este trabajo debido a la estructura de las imágenes utilizadas se ha procesado la mamografía con la combinación de distintas técnicas que se explicaran en la siguiente sección.

Descripción de los pasos efectuados en esta tesis

En primer, lugar una vez realizado el paso de adquisición de las imágenes se ha procedido a realizar discretizar la imagen para poder operar sobre ella, en el dominio espacial. Tal como se comentó anteriormente, una imagen digital 2D es una función bidimensional discreta $f(x,y)$ que asocia a cada punto o píxel discreto (muestreado) un valor discreto (cuantizado) de brillo o intensidad correspondiente [41].

- **Muestreo:** discretización espacial, pues x e y son finitas y discontinuas.
- **Cuantización:** discretización del nivel de gris $f(x,y)$, que toma un valor discreto $\in [0, L-1]$, donde $L = N^\circ$ niveles = 2^k (potencia de 2).

Así, una lectura de la imagen con la cual vamos a trabajar es el paso mínimo para poder realizar una operación sobre la misma. Para ello, se hará uso de la herramienta *Image processing Toolbox* de Matlab R2010b. Algunos comandos útiles para realizar estas tareas son el **imread**, el **uigetfile**. Este último, permitirá escoger una imagen para abrirla, posteriormente daremos lectura y mostraremos la imagen, para esto utilizamos el comando **imread** y de esta manera podemos mostrar la imagen para trabajar.

A continuación una vez discretizada la imagen en el dominio espacial procedemos a realizar un análisis detallado y dar solución a los problemas detectados mediante las

técnicas que se requieran en esta etapa llamada pre-procesamiento. Para hacer más fácil la tarea del experto médico, se ha construido una interfaz sencilla y usable para la visualización de los resultados de cada uno de estos procesos. Así se procede a explicar cada uno de ellos.

Después se detecta automáticamente el tipo de la imagen, que la aplicación infiera si es RGB o en niveles de gris. Si es RGB entonces, se convierte en imagen a escala de grises, para ello se utiliza el comando **rgb2gray**. El siguiente paso es realizar una binarización óptima automáticamente mediante el algoritmo de Otsu [41].

La binarización es un proceso de reducción de la información de una imagen a dos valores: 0 (negro) y 255 (blanco), el método de Otsu proporciona el umbral óptimo ("*threshold*") para la segmentación de la imagen, bajo el criterio de máxima varianza entre fondo ("*background*") y objeto ("*foreground*") [35].

Este método es uno de los mejores de umbralización global y uno de los avances en técnicas de umbralización. Consiste básicamente en seleccionar un umbral óptimo por un criterio discriminante, a saber, la separabilidad máxima de las clases resultantes de los niveles de gris. La prueba es dependiente sólo del nivel gris del punto. [42]

En este caso se ha obtenido el umbral óptimo mediante el comando *graythresh* que realiza este proceso de retornar un valor normalizado entre 0 y 1, para posteriormente aplicarlo a la imagen y lograr una buena binarización. Basado en el umbral óptimo obtenido, convertimos la imagen a imagen binaria, esto con el comando *im2bw*.; La función *im2bw* convertirá la imagen a binaria basada en un umbral; de esta manera usando estos dos comandos convertimos la imagen en binario, y elegimos el umbral para minimizar la varianza de los píxeles blancos y negros. (Fig. 4.4 (b))

Una vez realizada la binarización de umbral automático, procedemos a realizar una operación morfológica la cual nos permitirá extraer componentes de imágenes útiles para la representación y descripción de forma de regiones. Para ello se aplicará a las imágenes los siguientes procesos:

- Extracción de límites o bordes
- Eliminar picos y huecos
- Etiquetar los objetos
- Calcular el área y caja de objetos
- Tomar el área máxima
- Obtener la región de interés (ROI) (Fig. 4.4(c))

El comando para realizar la operación morfológica fue **bwmorph**, con su operación respectiva, que para el caso fue “*open*” para eliminar los picos y “*close*” para eliminar los huecos. Para obtener la región de Interés (ROI) con la cual trabajaremos, hacemos uso del comando **bwlabel** el cual transforma la imagen en una representación de etiquetas. A continuación calculamos el área que nos interesa de la mamografía usando el comando **regionprops**, el cual determina las propiedad de cada objeto conectado en la imagen binaria.

Por último, Una vez que se ha detectado la región de la imagen, simplemente mediante la orden “**imdilate**”, dilatamos los bordes detectados que corresponden con la imagen seleccionada mediante la sentencia “**strel**”, la cual crea una estructura de elementos morfológica.

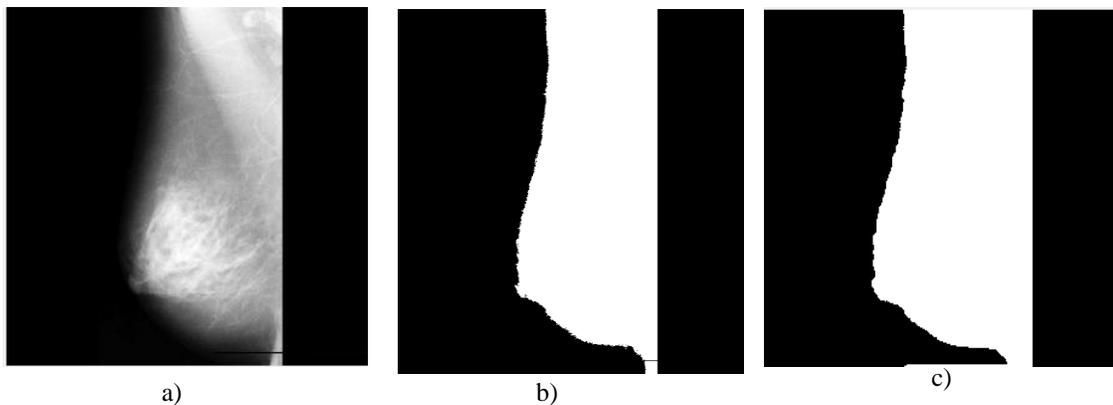


Fig. 4.4. a) Imagen Original; b) Imagen Binarizada (Otsu); c) Segmentación del área de interés, operación morfológica

Mejoramiento de Contraste

Para mejorar la calidad en la imagen se hace uso de una máscara *Unsharp*, enfatizando en su contenido de alta frecuencia, mejorando los bordes y la información de detalles. A veces es necesario corregir la nitidez de la imagen, ya que durante la digitalización puede haber cambios bruscos de la tonalidad de la misma.

En primer lugar, para mejorar el contraste se construyó una máscara de 3 x 3 mostrada en la Fig.4.5.

0	-1	0
-1	4	-1
0	-1	0

a)

-1	-1	-1
-1	8	-1
-1	-1	-1

b)

1	-2	1
-2	4	-2
1	-2	1

c)

Fig. 4.5. Mascaras empleadas para la máscara Unsharp

En la mayoría de las imágenes fue necesario emplear las mascara de la Fig.4.5 (b). La máscara *unsharp* aplicada está basada en:

$$y(n,m) = x(n,m) + \lambda z(n,m) \quad (6)$$

Donde $z(n,m)$ es la imagen de salida de un filtro lineal paso alto, $x(n,m)$ es la imagen de entrada, λ es el factor de mejoramiento que controla el nivel de mejoramiento de contraste de salida, $y(n,m)$ es la imagen de salida con un contraste mejorado. Un filtro paso alto comúnmente empleado para este método es el filtro Laplaciano el cual crea una máscara que realza los pixeles con respecto a su vecindad a través del aumento de su nivel de gris.

Una vez elegida la máscara con la cual trabajar, procedemos a aplicar el filtro respectivo con el comando **imfilter** y los parámetros necesarios. A continuación obtenemos el perímetro de los objetos de la imagen binaria con el comando **bwperim**, posteriormente lo dilataremos como en el caso anterior de segmentación y encontramos las coordenadas de los puntos blancos. El filtrado laplaciano lo usamos porque destaca las regiones donde hay

cambios bruscos de intensidad, permitiéndonos de esta manera dar un realce de intensidad en la zona que nos interesa de las microcalcificaciones. (Ver Fig. 4.6.).

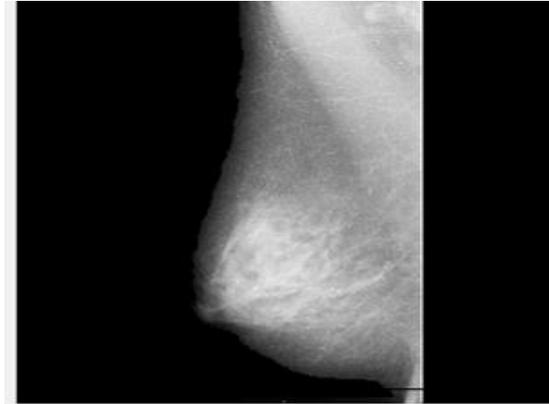


Fig.4.6. Resultado de aplicar máscara unsharp (filtro laplaciano)

Mejoramiento a través de una modificación de histograma

El histograma de una imagen representa la frecuencia relativa de ocurrencia de los niveles de gris dentro de una imagen. Las diferentes técnicas que se usan en el modelamiento del histograma, nos permitirán modificar la imagen, para mejorar la intensidad de la mamografía.

Una vez aplicado la máscara *unsharp* podemos tener una mejor visualización de las microcalcificaciones en la imagen, procediendo a la modificación de los niveles de gris. Aquí necesitaremos seleccionar un umbral, el cual permitirá separar los valores de intensidad de los pixeles correspondientes a microcalcificaciones y el resto de la imagen; para efecto se ha probado con dos funciones matemáticas: modificación lineal (Ecuación 7) y modificación exponencial (Ecuación 8).

$$P(i,j) = \begin{cases} S(i,j) - \bar{D} & , \text{ si } S(i,j) \geq T \\ S(i,j) - \bar{D} & , \text{ si } S(i,j) < T \end{cases} \quad (7)$$

$$P(i,j) = \begin{cases} e^{S(i,j)/k} - \bar{D} & , \text{ si } S(i,j) \geq T \\ e^{S(i,j)/k} - \bar{D} & , \text{ si } S(i,j) < T \end{cases} \quad (8)$$

Donde $P(i, j)$ es el pixel de salida, $S(i, j)$ es el pixel de entrada, T es el umbral, D es un entero positivo y k está dada por la Ecuación 9:

$$k = \frac{S(i, j)_{\max}}{\ln S(i, j)_{\max}} \quad (9)$$

Donde $S(i, j)_{\max}$ es el nivel máximo de gris de la imagen. Los niveles de gris están codificados entre 0 que representa el color negro y 255 representa el color blanco. Los Valores T y D fueron ajustados por prueba y error, de acuerdo a la observación en las imágenes, obteniendo así finalmente los valores de $T= 210$ y $D=50$ [43]. En la Fig. 4.9(a) se puede observar la imagen resultante al aplicar la ecuación 7 correspondiente a la modificación de histograma lineal, mientras tanto en la Fig. 4.7 (b) observamos la aplicación de la ecuación 8 que corresponde a la modificación de histograma exponencial. Luego de realizadas las pruebas respectivas, haciendo uso de la ecuación 3 se intenta dar un poco más de luminosidad a las calcificaciones en comparación con el resto de la imagen, podemos observar que los valores entre los componentes de bajo y alto contraste puedan tener un rango considerable entre sus niveles de intensidad, resaltando de esta manera las microcalcificaciones en la imagen mamográfica.

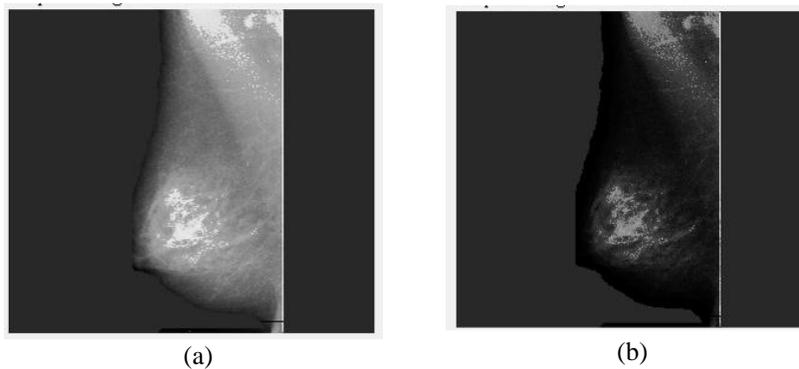


Fig. 4.7. (a) Resultado de aplicar Histograma Lineal, ecuación 2; (b) Resultado de aplicar histograma exponencial, ecuación 3.

Detección de Microcalcificaciones

Análisis y Aplicación de la Transformada Wavelet

EL análisis wavelet no es más que la descomposición de una señal arbitraria en versiones escaladas y trasladadas de la wavelet original. El análisis de wavelets puede muchas veces comprimir o eliminar ruido sin degradación apreciable. El análisis Wavelet empleado fue en dos dimensiones, las sub-bandas de aproximación son suprimidas, debido a que estas contienen los componentes de baja frecuencia de la imagen, por tanto todos los elementos de la matriz son llevados a cero. Finalmente, después de aplicada la transformada wavelet en dos dimensiones, se regresa al espacio de la imagen, transformando inversamente los coeficientes corregidos con la ayuda de la función **idwt2**.

EL algoritmo de descomposición y reconstrucción se lo observa en la Fig. 4.8.

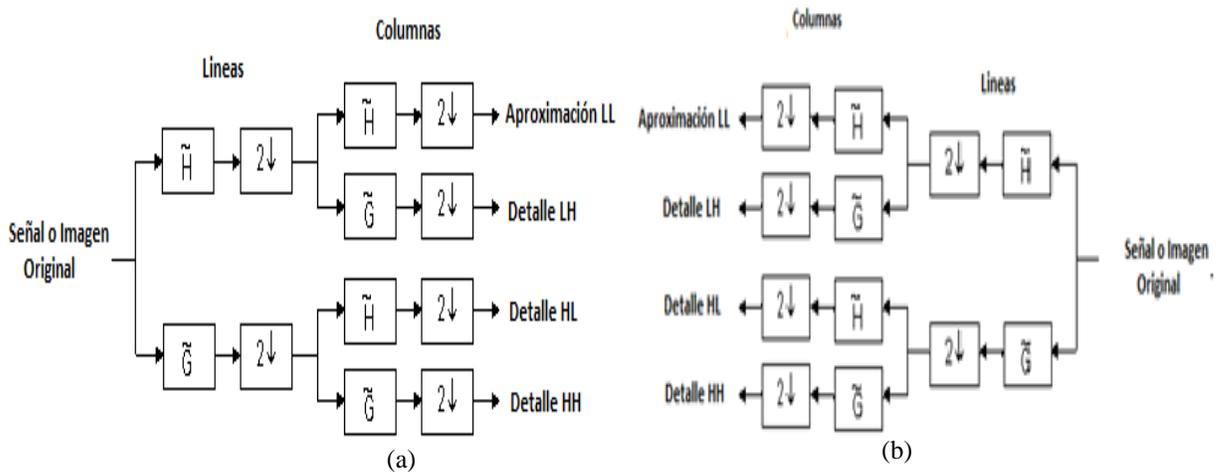


Fig. 4.8. (a) Algoritmo Descomposición de la Transformada Discreta wavelet Bidimensional; (b) Algoritmo Reconstrucción de la Transformada Discreta wavelet Bidimensional Inversa

En la Fig. 4.9 se muestra el proceso al aplicar la transformada Wavelet 2D a la imagen Original. Por último se procede a umbralizar la imagen resultante para eliminar algunas partes reconstruidas no correspondientes a microcalcificaciones.

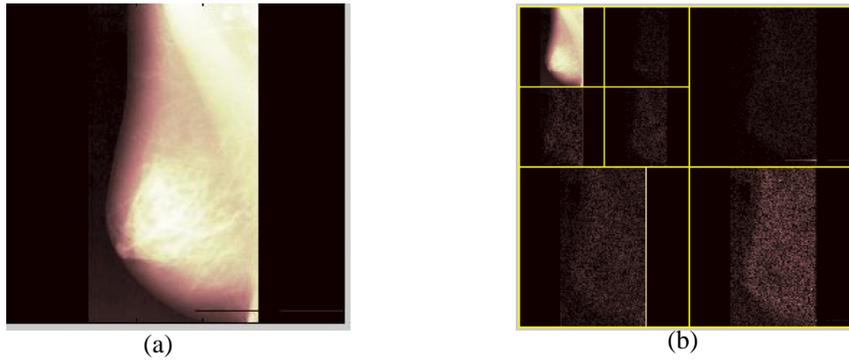


Fig. 4.9. (a) Imagen Original; (b) Descomposición Nivel 2

En la Fig. 4.10., se muestran diferentes resultados obtenidos con los experimentos realizados con diferentes familias de wavelets madre: *Coiflets*, *Daubechies*, *Symlets* y su respectivo Orden.

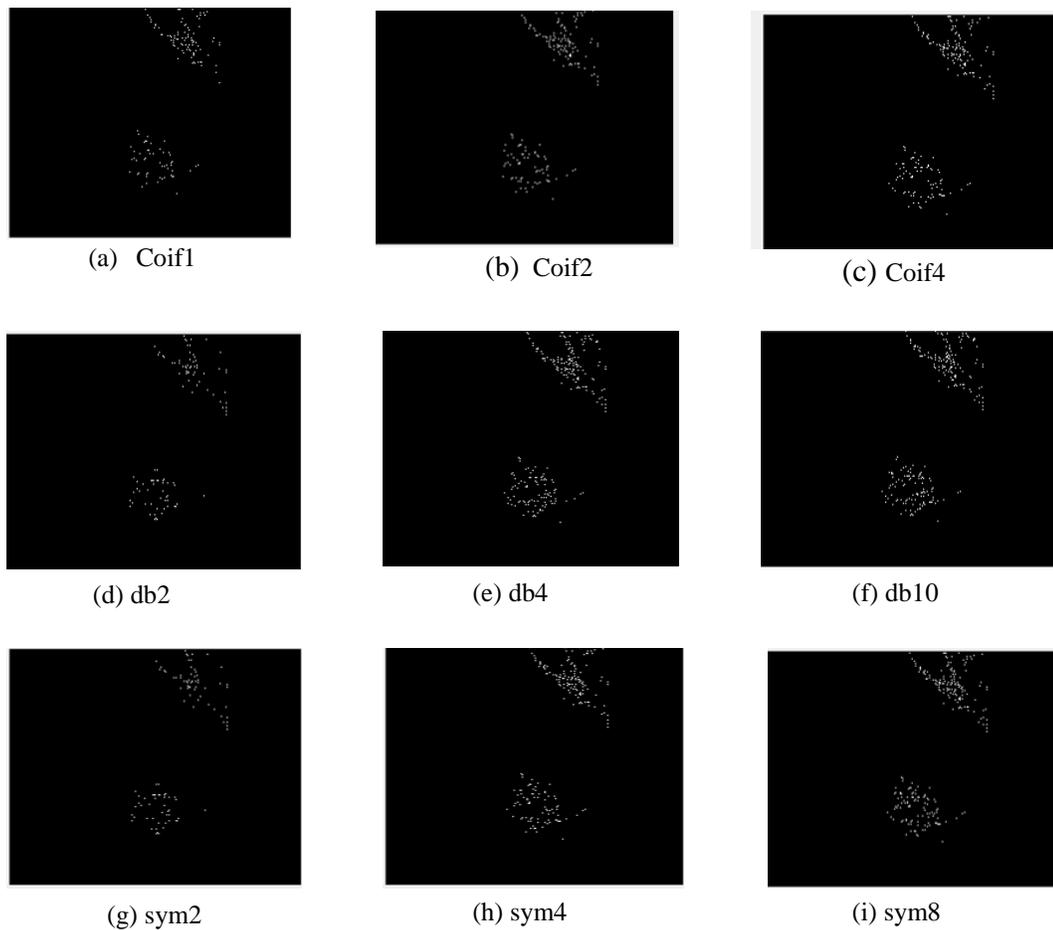


Fig.4.10. Imágenes Resultantes de aplicar las Familias Wavelet

Con dichas pruebas realizadas a las imágenes utilizadas, y comparando los resultados obtenidos respecto a las demás wavelets (Fig. 4.10) se ha podido comprobar que la wavelet madre *coiflets* es la que mejor resultado arroja, tal como se puede ver en la Fig.4.11.

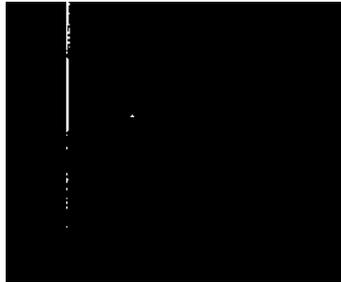


Fig.4.11. Resultado de Aplicar Transformada Discreta de Wavelet. Coif2

4.3.3 Objetivo3: Cuantificar las propiedades de las características identificadas

Con la etapa de reconocimiento o clasificación se pretende distinguir y describir los objetos, gracias al análisis de ciertas características que se establecen previamente para diferenciarlos. Por ejemplo el color, forma (área, perímetro,...). Cada objeto representado mediante un vector con los valores de sus características, es clasificado a partir de las mismas. En el caso de esta tesis, la clasificación posible esperada es identificar si hay microcalcificación o no. Dicha clasificación depende de la característica del píxel, de la forma de la microcalcificación y la cantidad de píxeles que forman la patología (racimo, si tiene más de cuatro calcificaciones,...). Para su interpretación, se contó con el conocimiento del experto médico, concluyendo el tipo de calcificación.

Así, la respuesta de este tipo de proceso será la obtención de una clasificación robusta de algún tipo de anomalías presente en una imagen de mamografía. Las anomalías que se presentan en la mama se pueden descubrir mediante un programa de detección precoz, y por lo tanto poder tratarlas adecuadamente en una etapa previa al desarrollo de la anomalía.

Las patologías de la mama pueden ser de dos tipos: las opacidades infraclínicas y las microcalcificaciones. [44]:

1. **Las Opacidades infraclínicas**, o distorsiones de la estructura mamaria, y masas irregulares, se caracterizan por presentar dos formas: las estrelladas y las redondas.
 - a. Las estrelladas son aquellas que pueden presentarse con centro denso y que suponen el 20-30% de los cánceres infraclínicos. El otro tipo de estrelladas son las estrelladas sin centro denso (distorsiones localizadas) representan el 10-20% de los cánceres infraclínicos. El 60% de estas lesiones sometidas a biopsia son malignas.
 - b. **Redondeadas**. Estas representa el 10-20% de los cánceres infraclínicos, pero más del 50% de ellas, cuando se someten a biopsia resultan ser benignas

2. **Microcalcificaciones**. Las microcalcificaciones son minúsculos depósitos de calcio que aparecen como pequeños puntos blancos y brillantes en la mamografía (Fig. 4.12). Estas pueden presentarse aisladas o en grupos. Los conglomerados de microcalcificaciones se definen por contener dentro de un 1cm^2 de la región de la ROI al menos 5 microcalcificaciones. Las microcalcificaciones agrupadas en mamografías son un signo temprano de una lesión maligna en casi la mitad de los cánceres de mama. Su tamaño varía desde los 0.1 mm hasta 5 mm de diámetro, por tanto un radiólogo deberá analizar exhaustivamente la mamografía con lupa para poder localizar las microcalcificaciones, por tanto constituyen uno de los problemas de diagnósticos más difíciles del tratamiento de cáncer de mama [44].

La clasificación de estas microcalcificaciones se divide en los cinco tipos que se muestran en la Tabla 4.2

Tabla 4.2 Clasificación de Microcalcificaciones

Tipo	Forma
I	Microcalcificaciones anulares, redondeadas, de centro claro.(0% Malignidad).
II	Microcalcificaciones puntiformes regulares, redondeadas, (10% Malignidad).
III	Microcalcificaciones "en polvo", muy finas, sin poder precisar su forma ni su número (19% Malignidad).
IV	Microcalcificaciones puntiformes irregulares, son poliédricas, en grano de sal (29% Malignidad).
V	Microcalcificaciones vermiculares, son alargadas, como un árbol sin hojas (72% Malignidad).

Además de la anterior clasificación (Tabla 4.2), se han considerado otros factores a la hora de valorar qué tipo de microcalcificaciones son más susceptibles de corresponder a carcinomas:

- Número de microcalcificaciones por cm^2 : mayores de 20 sugiere la presencia de patología maligna.
- Número total de microcalcificaciones: mayor de 30 microcalcificaciones en total es sospechoso de malignidad.
- Irregularidad de la densidad y del tamaño: Más común en los tumores malignos.
- Distribución lineal o en ramas: Típico de tumor maligno.

Y es aquí, donde precisamente el presente trabajo tiene aplicabilidad pues va a contribuir en su detección y ayuda al diagnóstico clínico de la entidad SOLCA, ya que en este trabajo se ha centrado en detectar esta patología.

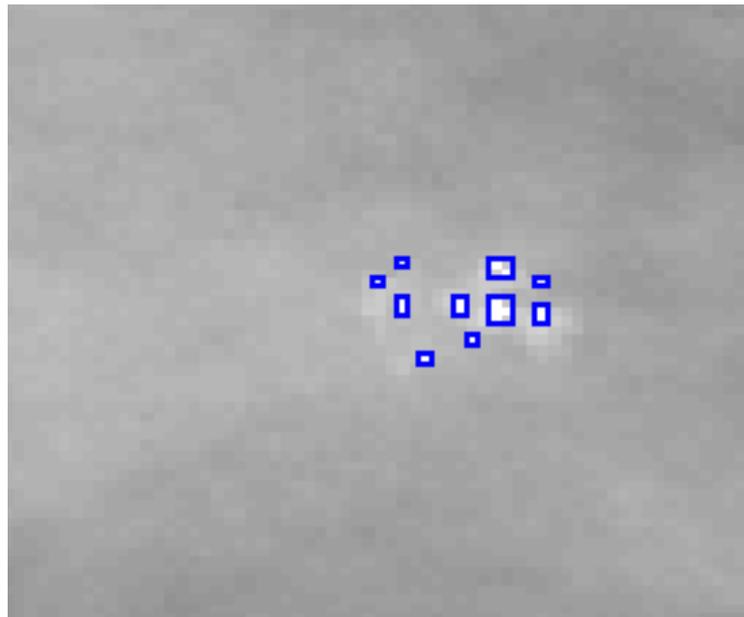


Fig. 4.12 . Corte ampliado de mamografía mdb211.pgm con múltiples microcalcificaciones (marcadas con recuadrados)

4.3.4 Objetivo4: Evaluar la calidad de la detección automática realizada

En este apartado se presentará la evaluación de los resultados obtenidos con la técnica de detección de microcalcificaciones propuesta, tomando como entrada las imágenes mostradas en la (Fig. 4.11) de las distintas pruebas correspondientes a la clasificación obtenida con cada familia de wavelets utilizada y la segmentación del *ground Truth* delineado por el experto. La validación de los resultados se ha realizado mediante dos tipos de técnicas: la validación cualitativa y la validación cuantitativa que se describen a continuación.

Validación cualitativa

Para evaluar la bondad de la clasificación se ha implementado varios procesos que demuestran la adecuada detección de las microcalcificaciones además de la observación visual del grupo de expertos médicos.

Para ello, se utilizó una muestra de 30 imágenes obtenidas de la base de datos MI-AS [31], de las cuales 22 presentan calcificaciones. Una vez que se detectan las microcalcificaciones (Fig. 4.11), procedemos a realizar una serie de medidas. Una forma consiste es hacer una superposición de la imagen resultante sobre la imagen original o preprocesada y obtener algunos parámetros basados en región, como por ejemplo, el etiquetado y la conectividad de los píxeles vecinos, el conteo de píxeles clasificados, etc.

En este caso, del etiquetado de los píxeles vecinos conectados, se puede considerar la vecindad a 4 respecto a un pixel p y se denota por $N_4(p)$. Existe también 4 vecinos diagonales de p y se les denota por $N_D(p)$. $N_4(p)$ y $N_D(p)$ juntos, forman la vecindad a 8 que se denota por $N_8(p)$ [25].

El concepto de conectividad es utilizado para establecer los límites de objetos en regiones dentro de una imagen. Para determinar si dos pixeles están conectados se determina si son adyacentes en algún sentido ($N_D(p)$, $N_4(p)$ por ejemplo) y si sus niveles de gris satisfacen un criterio de similaridad (si son iguales, por ejemplo).

Para realizar esta tarea se ha utilizado los comandos *bwlabel* del *Image Processing Toolbox de matlab* y el comando *regionprops* para calcular y determinan algunas propiedades de cada objeto conectado (Fig. 4.13).

En este caso se ha dibujado un rectángulo a través de los bordes de los puntos de microcalcificaciones detectados, mediante el comando *rectangle* para mostrar la localización de los mismos (Fig. 4.13).

No obstante, para verificar de forma visual la localización efectiva de la técnica propuesta de la región en donde se encuentra la malignidad y probar que el algoritmo propuesto es robusto se ha superpuesto el resultado sobre la imagen filtrada (Fig. 4.14) y las estrategias utilizadas en el anterior objetivo, tales como el etiquetado para clasificar los objetos según su conectividad. Esta técnica de la calidad visual mediante la aplicación del zoom nos permite poder ver objetivamente la localización y detección de la patología (Tabla 4.2) de los resultados obtenidos al aplicar las técnicas propuestas. Además para confrontar la validez de los mismos, en el uso médico de la misma, se aporta un certificado escrito emitido por expertos de la Unidad Oncológica de SOLCA que acreditan la bondad de este trabajo (Ver anexo 1).

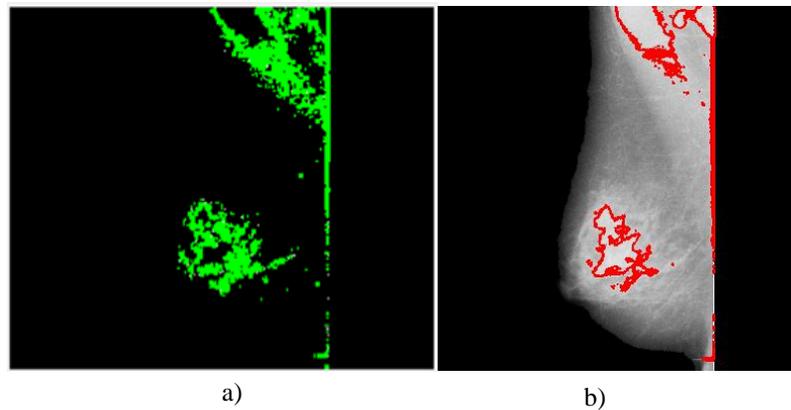


Fig.4.13. a) Resultado de Etiquetar las microcalcificaciones detectadas en la imagen mdb211.pgm, aplicando la wavelet madre *coif2* y b) la misma imagen superpuesta en la imagen original aplicando *coif2*

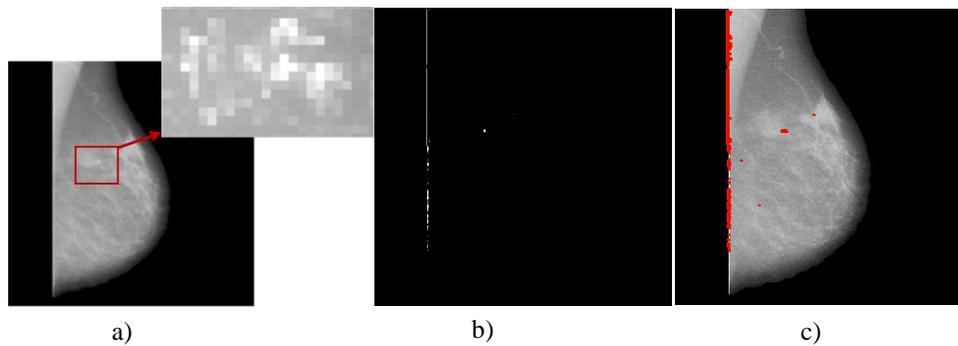


Fig. 4.14. Microcalcificaciones detectadas: a) La imagen mdb248.pgm original en zoom la localización de las calcificaciones, b) imagen de las microcalcificaciones obtenidas con el método propuesto etiquetadas en rojo y c) la imagen b) superpuesta en la imagen original filtrada a)

Validación cuantitativa

En este paso de la validación cuantitativa, debemos indicar que no se pudo contar con un “*ground truth*” para todas las imágenes con las que se ha experimentado, debido a que delinear manualmente la zona de la patología conlleva muchas horas del especialista, en todo caso el criterio del mismo validó y garantizó la adecuada clasificación de la patología estudiada con esta técnica.

Conjuntamente, con el apoyo y el conocimiento del experto médico se pudo delinear de forma manual las calcificaciones de la imagen de test mdb248.pgm, adquiridas de la base de datos [31], quien con ello, pudo verificar que la clasificación automática se efectuó de manera adecuada. El *ground truth* delineado manualmente por un experto en el área ha servido para validar cuantitativamente la calidad de la clasificación de las imágenes mamográficas mediante las técnicas de precisión y exactitud, la correlación y el error cuadrático medio (en inglés, *Mean Square Error, MSE*).

Como se puede observar en la imagen de la Fig. 4.14 a), las microcalcificaciones aparecen como pequeñas manchas de color blanco brillante. En ocasiones no representan riesgo alguno. Pero cuando se muestran bajo ciertos patrones es posible que sean motivo de preocupación. Por ejemplo, algunas veces crecen en racimos o en una línea y pueden significar la presencia de cáncer.

En la Tabla 4.3 se observa los resultados experimentados con las imágenes adquiridas, corroborando de esta manera y conjuntamente con el criterio del médico especialista, la validez y veracidad del algoritmo aplicado, puesto que se puede observar las microcalcificaciones presentes en la imagen mamográfica con la implementación de un zoom a la imagen.

- La Exactitud (EC) es la proporción del número total de predicciones que fueron correctos. La exactitud se obtiene usando la ecuación:

$$EC = \frac{TN + TP}{TN + TP + FN + FP} \quad (10)$$

Donde TN es el total de negativos, TP total Positivos, FN Falsos Negativos, FP Falsos Positivos.

- La precisión (P) es la proporción de las predicciones de casos positivos que fueron correctos y que fueron obtenidos usando la ecuación:

$$P = \frac{TP}{FP + TP} \quad (11)$$

Para este caso, se ha utilizado una función propuesta y desarrollada en la Tesis Doctoral [25]

La covarianza

La covarianza permite obtener la relación lineal entre dos variables. Mientras más alta sea la covarianza entre dos variables, sus valores seguirán más de cerca las mismas tendencias dentro de un rango de puntos de datos. Se dice que dos variables están variando conjuntamente, y en el mismo sentido, cuando al crecer los valores de una de las variables también aumentan los de la otra.

La covarianza entre dos variables aleatorias es (Ecuación 12)

$$cov(x_1, x_2) = E[(x_1 - \mu_1) * (x_2 - \mu_2)] \quad (12)$$

Donde E es la esperanza y μ_i es igual a Ex_i

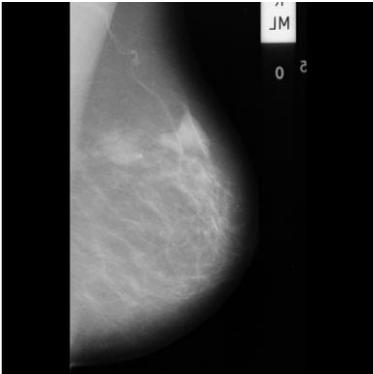
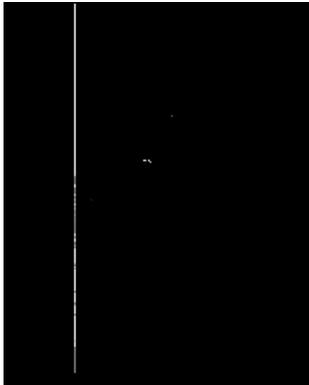
El error cuadrático medio

En el procesamiento de imágenes, se le suele utilizar para evaluar la calidad de los *denoising* o de la compresión de las imágenes, pero en este trabajo se ha aplicado para calcular el error cuadrático entre la segmentación del *ground truth* y la imagen segmentada automáticamente con el método propuesto (Ecuación 13).

$$mse = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n (iRef(i, j) - iS(i, j))^2 \quad (13)$$

Donde m es el número de filas y n el número de columnas en la imagen, *iRef(i,j)*, representa la imagen de referencia delineada por el experto médico (*ground truth*) e *iS(i, j)*, la imagen resultado de aplicar el método propuesto. Finalmente, los resultados después de aplicar estas técnicas se muestran en la Tabla 4.3

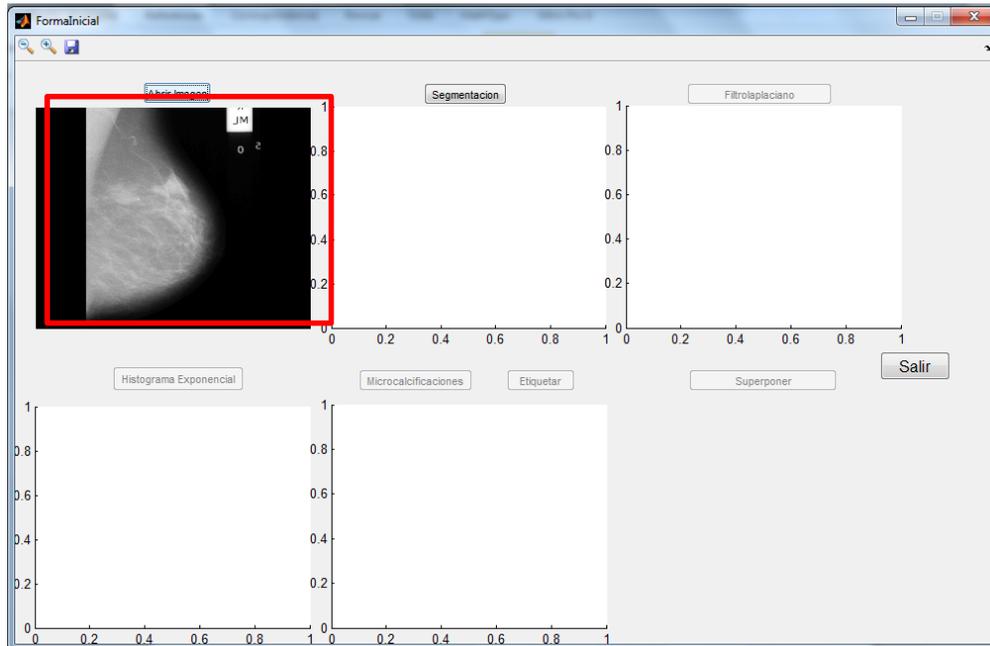
Tabla 4.3. Resultados de las métricas de validación aplicados a las imágenes de prueba

Imagen original	Imagen segmentada	<i>Ground truth</i>
 mdb248.pgm		
Métrica	Resultado clasificación	
Exactitud	91.98 %	
Precisión	93.98%	
Covarianza	0.0017 0.0001 0.0001 0.0007	
MSE	0.0022	

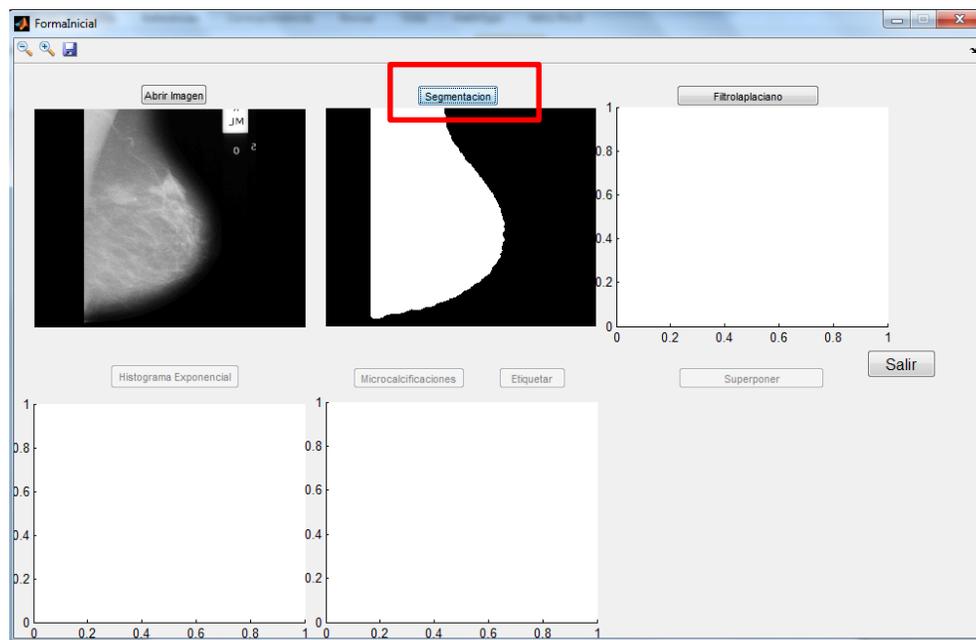
Para conseguir mejores resultados en la validación se realizó un post-procesamiento a la imagen resultado, mediante la aplicación de un filtro morfológico denominada erosión final o erosión extrema (***bwulterode***) para obtener la máxima región mediante la transformada de la distancia Euclidiana de la imagen complemento del resultado obtenido. La función utilizada para este fin realiza una operación conocida como apertura de un área.

Demostración del Algoritmo

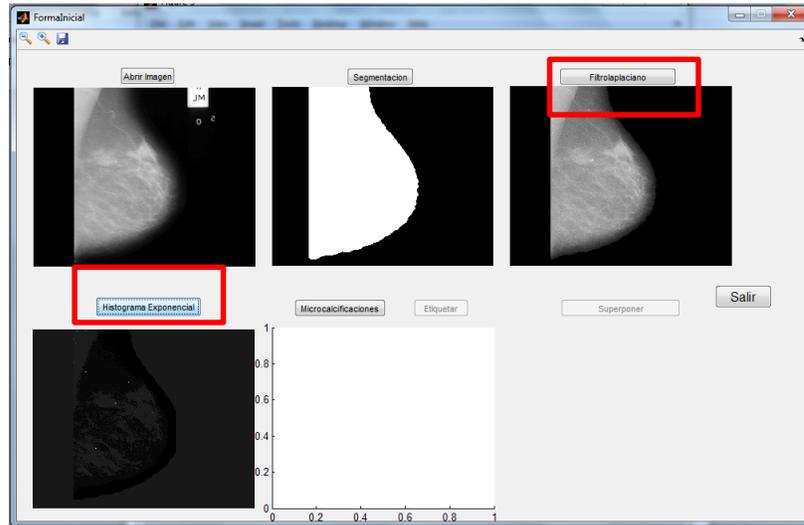
- En el primer recuadro se cargará la imagen para tratarla, analizarla y detectar las microcalcificaciones.



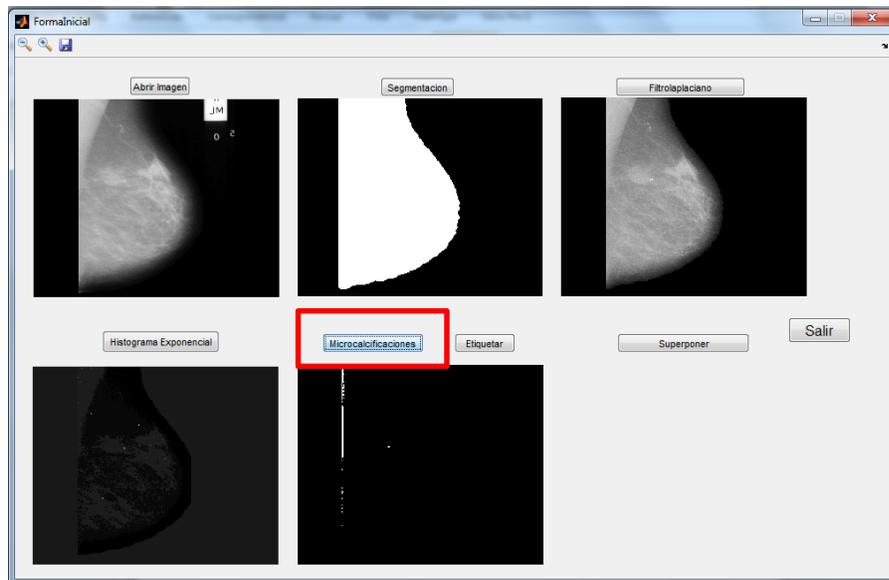
- Proceso de aplicar segmentación de la imagen.



- Filtro Laplaciano e Histograma Exponencial.

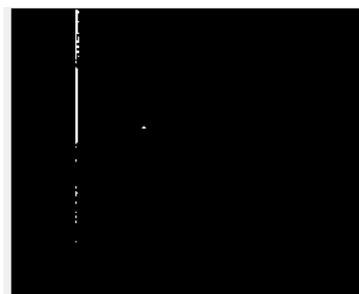
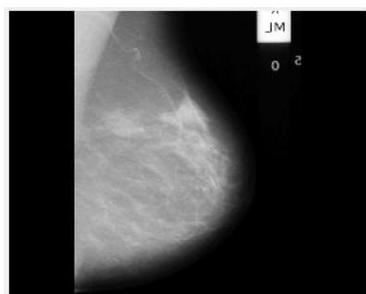


- Detección de Microcalcificaciones.



Imágen Original

Imágen Resultante



CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

- Los algoritmos desarrollados en las distintas etapas del análisis de este tipo de imágenes médicas puede detectar adecuadamente las pequeñas microcalcificaciones en imágenes mamográficas a través de las diferentes técnicas de realce, filtrado y el análisis en el dominio de la frecuencia con la aplicación de la transformada discreta wavelet 2D.
- Utilizando las técnicas propuestas en el presente trabajo se ha podido comprobar su rendimiento para detectar la región en donde se encuentra la lesión maligna y de esta forma ayudar al médico a definir con más precisión la región donde se encuentra la anomalía, en particular para el examen y tratamiento a futuro.
- Las métricas cuantitativas y cualitativas usadas como medida de calidad demuestran la capacidad del método propuesto para la supresión del ruido y los otros fenómenos presentes en las imágenes para realizar un adecuado procesamiento en cada una de las subsiguientes etapas del preprocesamiento.
- La desventaja de no contar con un “gold standard” para todas las imágenes ha restringido hacer validaciones de calidad cuantitativas más rigurosas que permitan emitir un dictamen más objetivo de la bondad de las técnicas desarrolladas, no

obstante, se contó con el criterio y conocimiento de los expertos médicos quienes pudieron afirmar la bondad de la técnica y la confianza para usarla en un futuro.

- El algoritmo desarrollado permite al médico especialista orientarse al lugar en donde se encuentran las pequeñas microcalcificaciones, y tomar una decisión concisa para su evaluación.
- A través de las pruebas experimentales el método propuesto permitió identificar y localizar adecuadamente las microcalcificaciones, especialmente aplicando la descomposición mediante las wavelets madre *coiflets* para las DWT2.
- Los resultados obtenidos con las métricas utilizadas para validar la calidad de la detección de las microcalcificaciones, a contexto de los resultados permiten afirmar que aplicando la descomposición con filtros ortogonales (familia wavelet madre *coiflets*) respecto a la db1 y db2, es la mejor en este caso, consiguiendo una exactitud del 91.98% y un error cuadrático medio de tan solo 0.0022.
- Se puede indicar que la incorporación del ciclo de vida del desarrollo de software en espiral en las etapas de un sistema de visión artificial permitió controlar cada uno de los procesos que se realizaron con la imagen mamográfica, permitiendo de esta manera evaluar y realizar las pruebas respectivas para que el resultado final sea óptimo, para la detección de las microcalcificaciones.

5.2 RECOMENDACIONES

- Se recomienda asignar un experto por parte del hospital SOLCA para que trace manualmente estas patologías para garantizar resultados más óptimos en la práctica médica y tener un “*gold standard*” de referencia para comparar la técnica automática.
- Para futuras aplicaciones, se recomienda crear un “*ground truth*” por parte de un experto médico del hospital SOLCA que grafique manualmente estas patologías para obtener resultados más confiables en la práctica médica y una valoración más objetiva de los métodos automáticos propuestos.

- Aplicar las técnicas propuestas en esta tesis sobre las mamografías propias de la Unidad Oncológica SOLCA.
- Como trabajos futuros se plantea extender estas técnicas a otras patologías relacionadas. Y que el sistema pueda inferir si hay presencia o ausencia de patologías en una imagen en cuestión.

REFERENCIAS

- [1] Microcalcificaciones mamarias: revisión de los descriptores y categorías BI-RADS, [online] Disponible en: http://www.scielo.cl/scielo.php?pid=S0717-75262013000500010&script=sci_arttext
- [2] Instituto Nacional de Câncer de Brasil (INCA), "Estimativa 2011: Incidência de Câncer no Brasil", 2011. [Online] Disponible en: <http://www.inca.gov.br/estimativa/2010/>
- [3] BO Anderson, R. Shyyan, A. Eniu, RA. Smith, CH. Yip, NS Bese, LW Chow, S. Masood, SD Ramsey, RW Carlson. Breast cancer in limited-resource countries: an overview of the Breast Health Global Initiative 2005 guidelines. *Breast J.* 2006, Suppl 1:S3-15. PMID: 16430397 [PubMed - indexed for MEDLINE]. [Online]. Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed/16430397>
- [4] MedlinePlus, cancer de mama. [Online]. Disponible en: <http://www.nlm.nih.gov/medlineplus/spanish/ency/article/000913.htm>.
- [5] Women's Health, [Online]. Disponible en: <http://www.webmd.com/women/picture-of-the-breasts>.
- [6] Câncer de mama [online] Disponible en: <http://www.hoy.com.ec/noticias-ecuador/cancer-de-mama-1-500-casos-al-ano-en-el-pais-509903.html>
- [7] Câncer de seno, tercero más letal entre ecuatorianas [online] Disponible en: http://www.telegrafo.com.ec/index.php?option=com_zoo&task=item&item_id=58264&Itemid=16
- [8] Segmentación y análisis morfométrico en mamografía [online] Disponible en: <http://www-rayos.medicina.uma.es/Rmf/RadiolDigital/TesisERB.htm>
- [9] Análisis de microcalcificaciones en mamografías [online] Disponible en: <http://riunet.upv.es/handle/10251/5643>
- [10] Mejora en la detección de microcalcificaciones en mamografías [online] Disponible en: <http://repositorio.bib.upct.es/dspace/handle/10317/201?mode=full>
- [11] Mamografía [online] Disponible en: <http://www.plannedparenthood.org/esp/temas-de-salud/mujeres-salud/mamografia-21195.htm>
- [12] Mamografía Selectiva de Detección y Diagnostico [online] Disponible en: <http://hostoben.blogspot.com/2008/09/mamografias-selectivas-de-deteccion.html>
- [13] ACR American College of Radiology [online]. Disponible en: <http://www.acr.org/Quality-Safety/Resources/BIRADS>.
- [14] Instituto Nacional de Câncer [online] Disponible en: <http://www.cancer.gov/espanol/recursos/hojas-informativas/deteccion-diagnostico/mamografias>
- [15] Sanitas [online] Disponible en: <http://www.sanitas.es/sanitas/seguros/es/particulares/biblioteca-de-salud/ginecologia/exploraciones-ginecologia/sin012124wr.html>.
- [16] Mamografía Tomosíntesis 3D [online]. Disponible en: <http://www.grupogamma.com/>

- procedimiento/tomosintesis-3d/
- [17] Mamografía [online] Disponible en: <http://www.mujsaludable.com/index.php/home/mejor-prevenir-que-lamentar/128-mamografia>
- [18] Kouro R. Samir, Musalem M. Rodrigo, (2002), Tutorial introductoria a la teoría de wavelet.
- [19] Pérez Enríquez Román (2007), El análisis Wavelet, Centro de Geociencias, UNAM [Online]. Disponible en: www.geociencias.unam.mx/geociencias/desarrollo/wavelets
- [20] Universidad de Oviedo, España. La transformada de Wavelet Discreta. [Online]. Disponible en: <http://coco.ccu.uniovi.es/immed/compresion/descripcion/spiht/discreta/discreta.htm>
- [21] Strang and T. Nguyen, "Wavelets and Filter Banks", Wellesley-Cambridge Press, 1996
- [22] R. Gonzalez C and R.E. Woods. "Digital Image Processing", 3° Ed. Prentice Hall, 2008.
- [23] Transformada Wavelet [Online] Disponible en: <http://www.dainf.ct.utfpr.edu.br/~ionildo/wavelet/cap3.htm>
- [24] J. F. Vélez, A. B. Moreno, A. Sánchez y J. L. Esteban, Visión por Computador, Editorial Dykinson y Serv. Publicaciones de la URJC, 2003. ISBN: 84-9772-069-5. Disponible online en: <http://www.visionporcomputador.es/libroVision/libro.html>
- [25] María G. Pérez. Métodos de Eliminación de Ruido y Segmentación de Tejidos Cerebrales en Imágenes médicas de Resonancia Magnética nuclear, Tesis Doctoral, 2012.
- [26] Paradigmas en el Desarrollo de Software [online]: Disponible en: <http://helisulbaransistemas.blogspot.com/2014/09/paradigmas-en-el-desarrollo-de-software.html>
- [27] Paradigma de Desarrollo de Software. [Online]. Disponible en: <http://proyectosinformaticoscht2009.wordpress.com/antes-de-empezar/marco-teorico/paradigmas-de-desarrollo-de-software/>
- [28] Procesamiento de Imágenes con MatLab [online]. Disponible en: http://asignatura.us.es/imagendigital/Matlab_PID_1314.pdf
- [29] Tutorial de Matlab [online]. Disponible en: http://pentagono.uniandes.edu.co/tutorial/Matlab/tutorial_matlab.pdf
- [30] Introducción a MatLab [online]. Disponible en: <http://www.mat.ucm.es/~jair/matlab/notas.htm#generalidades>
- [31] DataBase of Mammograms [online]. Disponible en: <http://peipa.essex.ac.uk/info/mias.html>
- [32] D.L. Pham, C. Xu, and J.L. Prince, "A survey current methods in segmentation. Annual Review of Biomedical Engineering", pp. 315-337, 2000. S. Basu, T. Fletcher, and R. Whitaker, "Rician noise removal in diffusion tensor RMI", Proc. MICCAI, Vol. 1, pp. 117-125, 2006.
- [33] S. Basu, T. Fletcher, and R. Whitaker, "Rician noise removal in diffusion tensor RMI", Proc. MICCAI, vol. 1, pp. 117-125, 2006.
- [34] K.J. Worsley and K.J. Friston, "Analysis of fMRI time-series revisited again", NeuroImage, Vol. 2,

- pp. 173-181, 1995.
- [35] R. M. Nishikawa, Y. Jiang, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, "Computer-aided detection of clustered microcalcifications," in Proc. IEEE Int. Conf.Syst., Man, Cybern., pp. 1375–1378, 1992.
 - [36] D. Zhao, M. Shridhar, and D. G. Daut, "Morphology on detection of calcifications in mammograms," in Proc. IEEE Int. Conf Acoustics, Speech, and Signal Processing, San Francisco, CA, pp. 129-132, 1992
 - [37] Dr. E.Sriniva, Dr. Madhuri P. "Detection of Malignancy in Digital Mammograms from Segmented Breast Region Using Morphological Techniques", vol. 5, 2013
 - [38] Fakhita Regragui, Microcalcifications Detection in Mammographic Images Using texture Coding, Vol.5, 2011, n° 8, 381-393
 - [39] L.P. Clarke, R.P. Velthuizen, M.A. Camacho, J.J. Heine, M. Vaidyanathan, L.O. Hall, and M.L. Silbiger, "MRI Segmentation and applications", Magnetic Resonance Imaging, Vol. 13, No. 3, pp. 343-368, 1995.
 - [40] RM, Haralick, LG Shapiro, "Image segmentation techniques", Comp. Vis. Graph image Proc., vol. 29, pp. 100-132, 1985.
 - [41] Métodos Otsu, [Online]. Disponible en: <http://es.scribd.com/doc/45875911/Metodo-Otsu>
 - [42] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics vol. 9, No.1, pp. 62–66, 1979..
 - [43] Nayid Triana, Alexander Cerquera, Computer-Aided Detection of Microcalcifications in Digital Mammograms to Support Early Diagnosis of Breast Cancer”
 - [44] Pedro José García La encina. "Mejora en la detección de microcalcificaciones en mamografías digitalizadas mediante la aplicación de arquitecturas neuronales, proyecto fin de carrera, 2004, Escuela Técnica Superior de Ingeniería de Telecomunicación Universidad Politécnica de Cartagena

ANEXOS

ANEXO 1

SOCIEDAD DE LUCHA CONTRA EL CANCER



UNIDAD ONCOLOGICA
SOLCA TUNGURAHUA

Izamba - Alfredo Coloma y Enrique Sánchez
Teléfonos: 2856098 - 2452750 - Fax: 2451995

Ambato, 08 de agosto de 2014

CERTIFICADO

A QUIEN INTERESE:

Por la presente certifico que revisé los resultados obtenidos a partir de imágenes de mamografías en el desarrollo de la tesis: *DETECCIÓN AUTOMÁTICA DE LESIONES EN MAMOGRAFÍAS USANDO TÉCNICAS DE ANÁLISIS DE IMÁGENES*, realizada por el estudiante Sr. CARLOS BOLÍVAR FIALLOS QUINTEROS y considero que reflejan adecuadamente la presencia de signos patológicos que requerimos desde el punto de vista médico para orientar la valoración de paciente con patología mamaria.

Es todo cuanto puedo certificar, pudiendo el interesado hacer uso del presente donde a bien tuviere

Atentamente:


U.O. SOLCA TUNGURAHUA
Dra. Aída Aguilar Salazar
CI 1801241389
Médica Tratante
SOLCA TUNGURAHUA

ANEXO 2

Código de Programación

- **Abrir una imagen.**

```
function AbrirForma_Callback(hObject, eventdata, handles)
% hObject    handle to AbrirForma (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
global imagen
% Abrimos la GUI para escoger la imagen
[nombre ruta]=uigetfile({'*.*'},'Seleccionar imagen');
if nombre ==0 % Si se presiona cancelar, no pasa nada
    return
end
```

- **Pre-procesamiento de la imagen. (Segmentación, Filtrado, Histograma)**

```
function Preprocesamiento_Callback(hObject, eventdata, handles)
% hObject    handle to Segmentacion (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% SEGMENTACIÓN
% Si la imagen es RGB, la pasamos a escala de grises
if size(imagen,3)==3 %3 planos: RGB
    img_gris=rgb2gray(imagen);
else
    img_gris=imagen;
end
% Binarización con umbral automático
bina=im2bw(img_gris,graythresh(img_gris)*0.7);
%Procesamiento morfológico
bina=bwmorph(bina,'open'); % Eliminar picos
bina=bwmorph(bina,'close'); % Eliminar huecos
img_gris= img_gris.*uint8(roi);

% Filtrado
% Obtener el contraste
Cont=str2double(get(handles.contraste,'String'));
% Histograma
% Definimos y Obtenemos el tipo de máscara
matrizA=[0,-1,0;-1,4,-1;0,-1,0];
matrizB=[-1,-1,-1;-1,8,-1;-1,-1,-1];
matrizC=[1,-2,1;-2,4,-2;1,-2,1];
% Filtramos en base a la máscara
B = imfilter(img_gris,matrizB','replicate');
% Cálculamos el valor de k
maximo=max(max(S));
k=maximo/log(double(maximo));
% Redondeo de los límites
```

```

lim=round(handles.limite);
% Asignar valores a T (Thereshold) y D
T=210;
D=50;
% Preasignación de memoria ceros
P=zeros(size(S));
% Realizamos el cálculo respectivo
% Visualizamos la imagen
imagesc(P, 'parent', handles.axes4), colormap('gray')

```

- **Detección de Microcalcificaciones**

```

function microcalcificaciones_Callback(hObject, eventdata, handles)
% hObject      handle to microcalcificaciones (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)
%Obtenemos la familia
% familia = 'coif';
% % obtenemos el orden
% orden = '2';
% % obtenemos el nivel
% nivel= '1';
% % Concatenamos la familia y el orden (Ejemplo: 'db4')
%wavel_orden=[familia,orden];
% Aplicamos la transformada wavelet
% [C,S] = wavedec2(double(P),str2double(nivel),wavel_orden);
% Get cA
[cA] = appcoef2(C,S,wavel_orden,1);
% % Get cH, cV and cD
% [cH,cV,cD] = detcoef2('all',C,S,1);
% % Eliminar el componente de frecuencia
% cA=zeros(size(cA));
%Familia coift
[LoD,HiD,LoR,HiR] = orthfilt('coif2');
imagesc(X, 'parent', handles.axes5);
colormap gray
axes(handles.axes5)
axis off
%Superposicion
calcificaciones=handles.microcal;
% llamamos a la imagen original
imagen_original=handles.imagen;
% Encontramos los puntos blancos;microcalcificaciones
[f c]=find(calcificaciones==1);
nm=[f c];
% Superponemos las microcalcificaciones en la imagen original
for ind=1:size(nm,1)
    imagen_original(nm(ind,1),nm(ind,2))=255;
end
% Mostramos la imagen original
imagesc(imagen_original, 'parent', handles.axes6);
colormap gray
axes(handles.axes6)

```

Manual de Usuario

Requerimientos:

Para sistemas de 32 bit que usan Windows como sistema operativo (Vista, XP SP1 y SP2, Server 2003 y 2000 SP3 y SP4) los requerimientos de hardware son los que se detallan a continuación:

Procesador: Intel Pentium (Pentium IV y posterior), Intel celeron, Intel Xeon, Intel Core, AMD Athlon, AMD Optaron

Disco duro: se requieren al menos 500MB de espacio disponible

Memoria RAM: 512 MB como mínimo.

Video: Tarjeta gráfica de 16, 24 o 32 bits con OpenGL, DirectX 9

Para sistemas de 64 bit que usan Windows como sistema operativo (Windows XP x64, Windows Server 2003, Windows Vista) los requerimientos de hardware son los que se detallan a continuación:

Procesador: Intel Pentium (Pentium IV y posterior), Intel celeron, Intel Xeon, Intel Core, AMD64

Disco duro: se requieren al menos 500 MB de espacio disponible

Memoria RAM: 512 MB como mínimo.

Video: Tarjeta gráfica de 16, 24 o 32 bits con OpenGL, DirectX 9

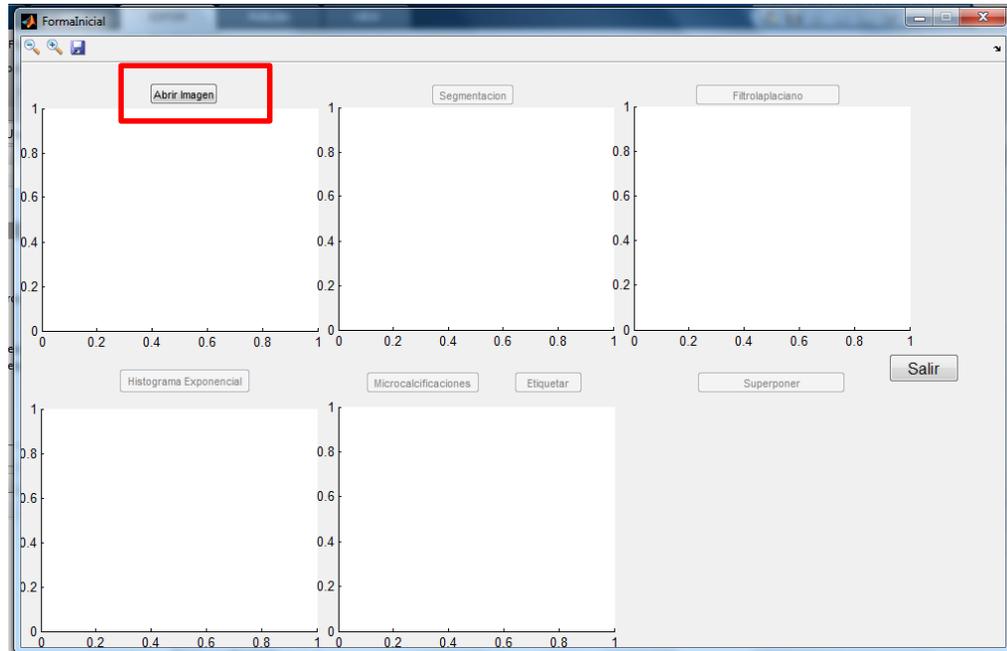
Ingreso a la Aplicación

Para Usar el algoritmo, damos clic en el icono del escritorio.

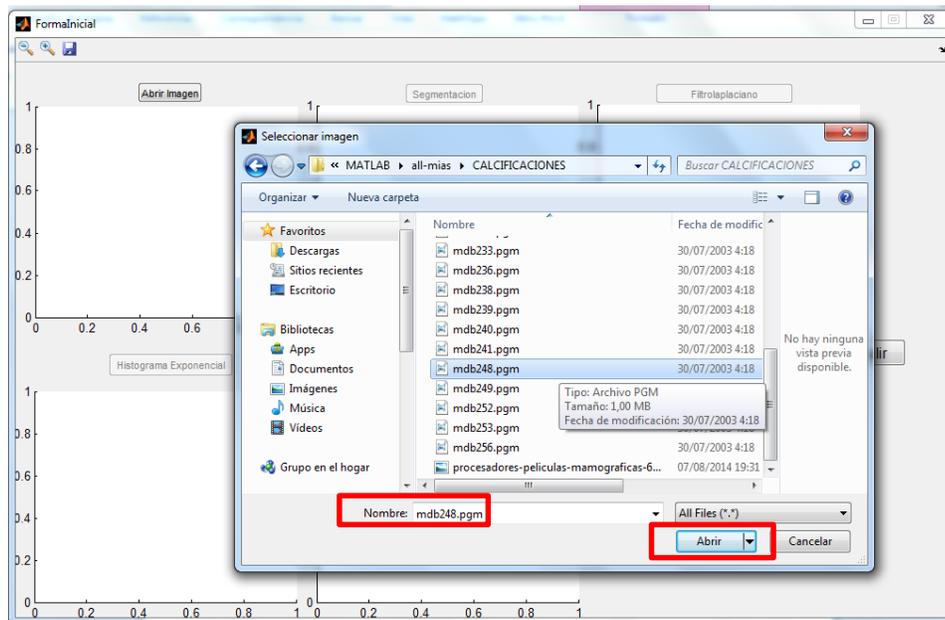


Se abrirá una ventana que trabajaremos con las distintas técnicas de aplicación.

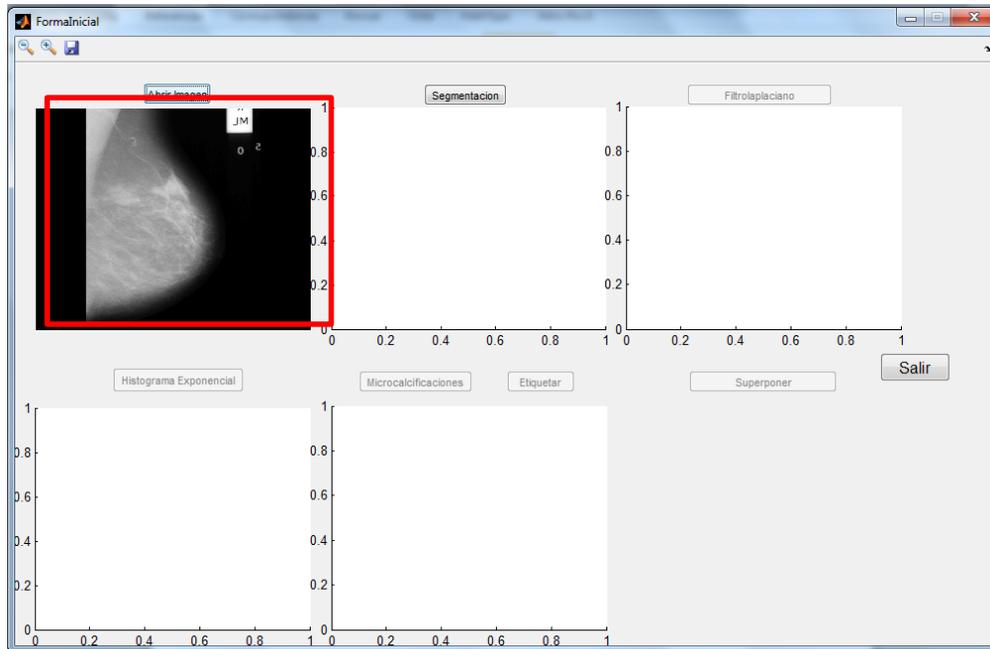
La interfaz de entrada para el médico especialista resultara de fácil manejo para el mismo. La ventana consta de 5 partes que se trabajarán para obtener una imagen de calidad y poder detectar las microcalcificaciones. Los botones se habilitarán conforme se realizan los pasos.



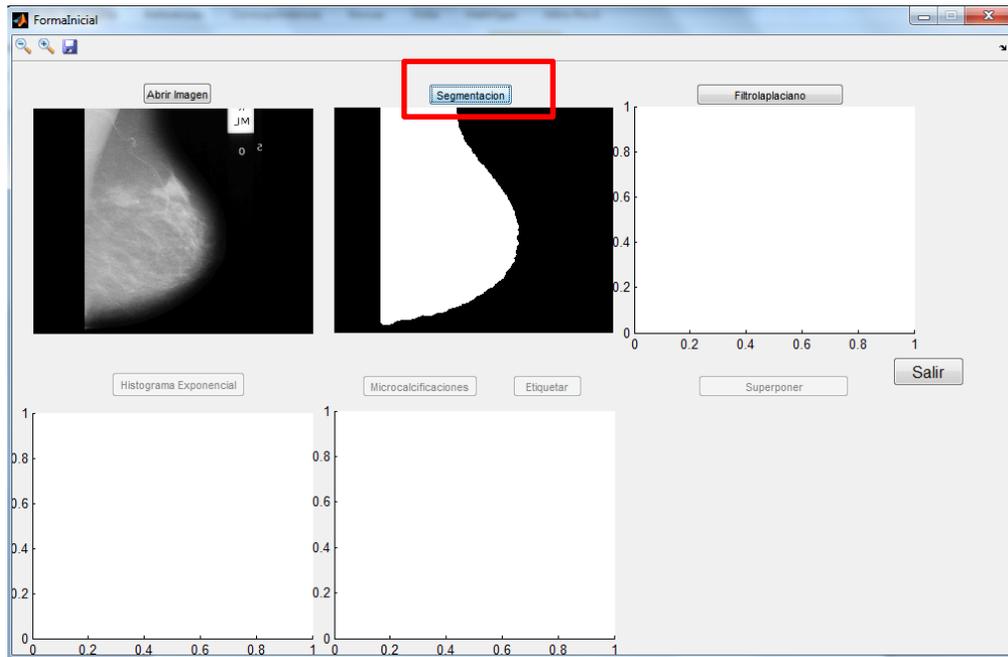
- Una vez, abierta la ventana, Escogemos y abrimos la imagen a trabajar, dando clic en el botón abrir.



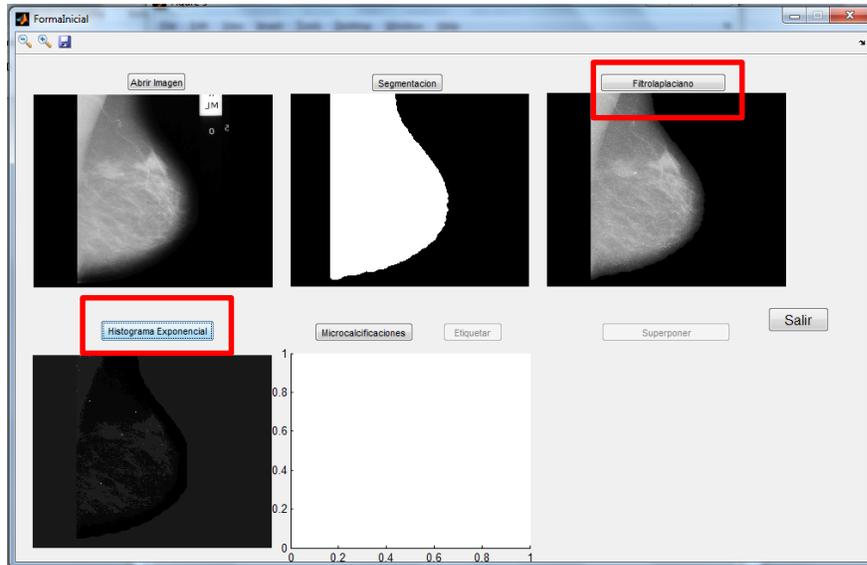
- Abierta la imagen, se nos cargara en el primer recuadro. Ahora podemos trabajar sobre la imagen y analizarla para la detección de microcalcificaciones.



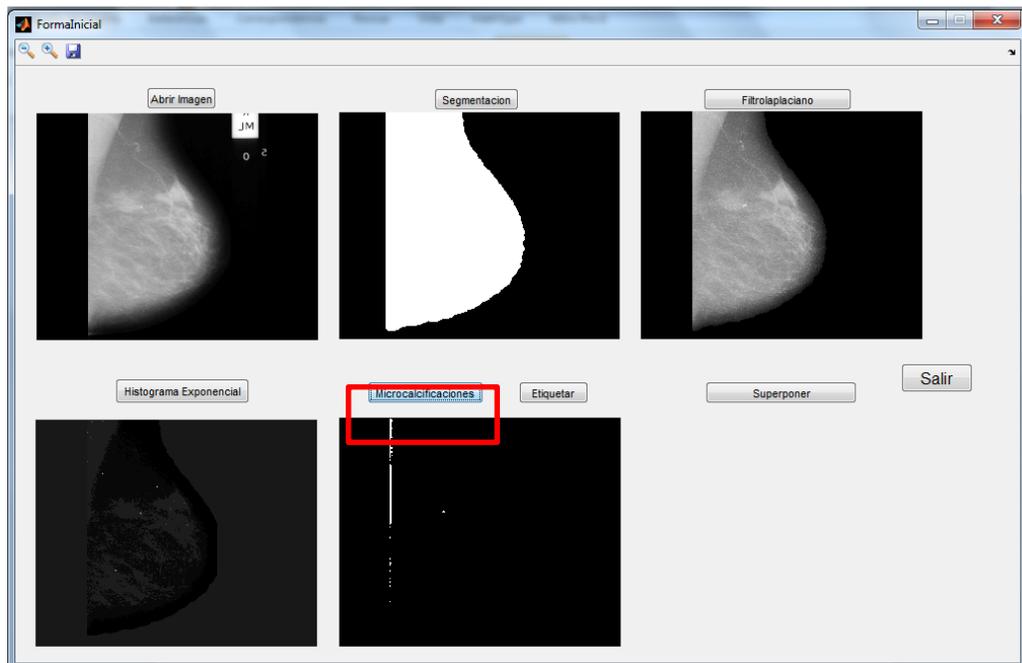
- El siguiente paso es segmentar la imagen, para eso damos clic sobre el botón segmentación.



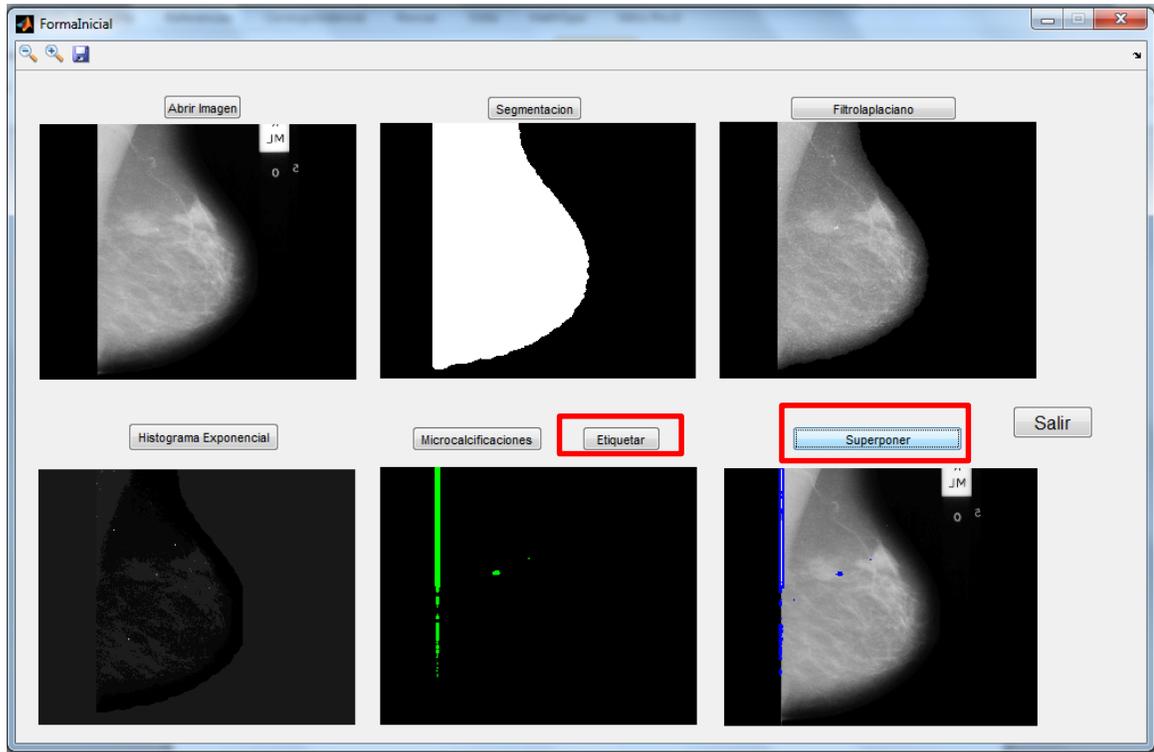
- El botón Filtro Lapaciano ha sido habilitado. Procedemos a dar clic sobre el mismo. A continuación daremos clic en el botón Histograma Exponencial.



- Realizado los pasos anteriores podemos encontrar las microcalcificaciones presentes en la imagen. Damos clic en el botón Microcalcificaciones.



- A continuación clic en el botón Etiquetar, finalmente clic en el botón superponer, de esta manera montaremos la imagen procesada sobre la imagen original con el hallazgo de las microcalcificaciones detectadas.



- Damos clic en el botón salir, culminado el proceso de detección de microcalcificaciones.

