

**APLICACIÓN SOFTWARE PARA APOYO AL DIAGNÓSTICO DE LESIONES
DE MAMA MEDIANTE ELASTOGRAFÍA ULTRASÓNICA**



**Daniel Felipe Giraldo Ruiz
Angela Maria Muñoz Rengifo**

**Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Electrónica, Instrumentación y Control
Ingeniería en Automática Industrial
Popayán, Octubre de 2016**

**APLICACIÓN SOFTWARE PARA APOYO AL DIAGNÓSTICO DE LESIONES
DE MAMA MEDIANTE ELASTOGRAFÍA ULTRASÓNICA**

**Daniel Felipe Giraldo Ruiz
Angela Maria Muñoz Rengifo**

**Monografía presentada como requisito parcial para optar por el título de
Ingeniero en Automática Industrial**

Director: Dr. Carlos Alberto Gaviria López

**Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Electrónica, Instrumentación y Control
Ingeniería en Automática Industrial
Popayán, Octubre de 2016**

Nota de aceptación: _____

Director: _____

Carlos Alberto Gaviria López

Firma del Jurado

Firma del Jurado

Popayán, Octubre de 2016

AGRADECIMIENTOS

A Dios por todo lo bueno en nuestras vidas, por cada nuevo día de vida, por acompañarnos, darnos fortaleza y sabiduría para terminar este proyecto.

A nuestros padres por todo su apoyo, confianza, su amor, sus sacrificios y esfuerzos para que salgamos adelante, porque siempre podemos contar con ustedes en momentos difíciles y buenos, los amamos mucho.

A nuestras hermanas por su amistad, confianza y su apoyo.

Gracias al doctor Carlos Alberto Gaviria López por su ayuda, tiempo y paciencia en el desarrollo de este proyecto.

A nuestros profesores, compañeros y amigos de carrera por todo lo compartido y aprendizajes que contribuyeron a nuestro crecimiento profesional y personal.

CONTENIDO

1. INTRODUCCIÓN.	2
1.1. JUSTIFICACIÓN	2
1.2. ALCANCES Y OBJETIVOS	3
1.2.1. Objetivo General	3
1.2.2. Objetivos Específicos	3
1.3. ESTRUCTURA DE LA TESIS	4
2. MARCO TEORICO.	5
2.1. REVISIÓN DE LA DETECCIÓN DE LESIONES DE MAMA	5
2.1.1. Mamografía	5
2.1.2. Imágenes por resonancia magnética	6
2.1.3. Ecografía (Ultrasonido)	6
2.1.4. Elastografía	9
2.1.5. Biopsia y pruebas de laboratorio	11
2.2. DIAGNÓSTICO DE LESIONES DE MAMA EN IMÁGENES DE ULTRASONIDO	12
2.2.1. Caracterización de las lesiones de mama en imágenes ecográficas	12
2.2.1.1. Lesiones benignas	12
2.2.1.2. Lesiones malignas	16
2.3. MÉTODOS PARA LA CLASIFICACIÓN DE LESIONES DE MAMA	21
2.3.1. BI-RADS Ecográfico	21
2.3.2. Escala de elasticidad UENO	26
2.4. DIAGNOSTICO ASISTIDO POR COMPUTADOR (CAD)	27
2.4.1. Pre procesamiento de imagen	28
2.4.1.1. Técnicas de filtrado	29
2.4.1.2. Técnicas en el dominio wavelet	32
2.4.2. Segmentación de lesión	35
2.4.2.1. Técnicas basadas en umbrales	36
2.4.2.2. Técnicas basadas en detección de bordes	37
2.4.2.3. Técnicas basadas en regiones	38

2.4.3.	Extracción de Características	40
2.4.3.1.	Características de textura	40
2.4.3.2.	Características morfológicas o geométricas	40
2.4.3.3.	Características basadas en modelos	41
2.4.3.4.	Características del elastograma	41
2.4.4.	Clasificación	44
2.4.4.1.	Clasificadores lineales	45
2.4.4.2.	Redes neuronales artificiales	46
2.4.4.3.	Árbol de decisión	47
2.4.4.4.	Máquinas de soporte vectorial	47
2.4.4.5.	Lógica difusa	48
3.	IMPLEMENTACIÓN DEL PROCESO CAD.....	51
3.1.	HERRAMIENTAS SOFTWARE	51
3.1.1.	QT 5.4	51
3.1.1.1.	QT Designer	52
3.1.2.	OpenCv 2.4.6	52
3.1.3.	Microsoft Visual Studio 2012 Professional	53
3.1.4.	Matlab Compiler Runtime 2013	53
3.1.5.	CMake 3.3.2	53
3.2.	ETAPAS CAD.....	54
3.2.1.	Pre procesamiento	54
3.2.2.	Segmentación	57
3.2.3.	Extracción de características	57
3.2.3.1.	Características descriptivas	58
3.2.3.2.	Características morfológicas	61
3.2.3.3.	Características de textura	63
3.2.3.4.	Características del elastograma	66
3.2.4.	Clasificación	67
4.	DESARROLLO DE LA APLICACIÓN	72
5.	RESULTADOS.....	77
5.1.	FASE 1	77

5.1.1.	Kmeans	78
5.1.2.	Red neuronal.....	79
5.2.	FASE 2.....	81
5.2.1.	Silhouette	81
5.2.2.	Stepwise.....	84
5.3.	FASE 3.....	85
5.3.1.	Kmeans	85
5.3.2.	Red neuronal.....	87
5.4.	FASE 4.....	90
5.5.	FASE 5.....	93
5.6.	FASE 6.....	96
6.	CONCLUSIONES Y TRABAJOS FUTUROS.	102
6.1.	CONCLUSIONES.....	102
6.2.	TRABAJOS FUTUROS	104
	BIBLIOGRAFÍA.....	105

LISTA DE FIGURAS

Figura 1. Ecografía en modo B	7
Figura 2. Módulo de young	9
Figura 3. Elastograma.....	10
Figura 4. Deformación de un quiste durante la ecografía	13
Figura 5. Deformación de un fibroadenoma durante la ecografía	14
Figura 6. Deformación de un lipoma durante la ecografía	14
Figura 7. Deformación de un tumor phyllodes durante la ecografía.....	15
Figura 8. Deformación de un galactocele durante la ecografía.....	15
Figura 9. Deformación de un papiloma intraductal durante la ecografía	16
Figura 10. Deformación de cáncer del epitelio ductal durante la ecografía	17
Figura 11. Deformación de cáncer ductal infiltrante durante la ecografía	18
Figura 12. Deformación de un carcinoma inflamatorio durante la ecografía.....	19
Figura 13. Deformación de un carcinoma lobular in situ durante la ecografía	19
Figura 14. Deformación de un tumor phyllodes durante la ecografía.....	20
Figura 15. Caracterización de las lesiones nodulares.....	23
Figura 16. Características ecogénicas de las lesiones nodulares.....	23
Figura 17. Escala UENO.....	26
Figura 18. Etapas de un sistema CAD Etapas de un sistema CAD	28
Figura 19. Histograma bimodal.....	37
Figura 20. Funcionamiento K-means	45
Figura 21. Topología de una RNAf	47
Figura 22. Resultado filtro wavelet con contracción garrote.	55
Figura 23. Pre procesamiento ROI	55
Figura 24. Escala de tonalidad de elastografía	56
Figura 25. Elastograma.....	56
Figura 26. Componente Hue del elastograma	56
Figura 27. Contorno de la lesión	57
Figura 28. Característica forma.....	59
Figura 29. Característica Orientación	60
Figura 30. Característica margen.....	60
Figura 31. Característica NSPD.....	62
Figura 32. Envoltente cóncava.....	62
Figura 33. Intersección de contornos.....	67
Figura 34. Ejemplo grafica Silhouette	68
Figura 35. Ejemplo grafica Stepwise.....	69
Figura 36. Diagrama de flujo del funcionamiento de la aplicación	72
Figura 37. Pantalla Inicial.....	73
Figura 38. Cálculo del elastograma	73
Figura 39. Selección de la ROI	74
Figura 40. Resultados clasificación.....	75
Figura 41. Visualización de características	76

Figura 42. Características caso similar	76
Figura 43. Comando Kmeans en Matlab	78
Figura 44. Entrenamiento de la red neuronal Matlab	79
Figura 45. Resultados optimización de características descriptivas y morfológicas	82
Figura 46. Resultados optimización de características de textura	82
Figura 47. Resultados optimización de características del elastograma.....	83
Figura 48. Resultados optimización de características silhouette.....	84
Figura 49. Comando Stepwise.....	84
Figura 50. Resultados optimización de características Stepwise.....	85
Figura 51. Entrenamiento árbol de decisión Matlab.....	90

LISTA DE TABLAS

Tabla 1. Características descriptivas de las lesiones nodulares	21
Tabla 2. Criterios de clasificación de masas	24
Tabla 3. Correlación entre categoría, lesión y probabilidad de cáncer	25
Tabla 4. Técnicas de pre procesamiento	34
Tabla 5. Técnicas de segmentación	39
Tabla 6. Características empleadas en la clasificación de lesiones de mama en imágenes de US	42
Tabla 7. Clasificadores comúnmente utilizados en el diagnóstico de lesiones de mama.....	48
Tabla 8. Resumen implementación proceso CAD	70
Tabla 9. Resultados entrenamiento Kmeans	78
Tabla 10. Clasificación RNA grupo 1	80
Tabla 11. Clasificación RNA grupo 2	80
Tabla 12. Resultado entrenamiento Kmeans con vector de características Silhouette.....	86
Tabla 13. Resultado entrenamiento Kmeans con vector de características Stepwise	86
Tabla 14. Clasificación RNA con vector de características Silhouette grupo 1	87
Tabla 15. Clasificación RNA con vector de características Silhouette grupo 2	88
Tabla 16. Clasificación RNA con vector de características Stepwise grupo 1	89
Tabla 17. Clasificación RNA con vector de características Stepwise grupo 1	89
Tabla 18. Clasificación árbol de decisión con vector de características Stepwise grupo 1.....	91
Tabla 19. Clasificación árbol de decisión con vector de características Stepwise grupo 2.....	91
Tabla 20. Matriz de confusión de las RNAs clasificación I.....	92
Tabla 21. Matriz de confusión de los árboles de decisión clasificación I	92
Tabla 22. Clasificación II RNA grupo 1	93
Tabla 23. Clasificación II RNA grupo 2	94
Tabla 24. Clasificación II Árbol de decisión grupo 1	94
Tabla 25. Clasificación II Árbol de decisión grupo 2	95
Tabla 26. Matriz de confusión de las RNAs clasificación II.....	95
Tabla 27. Matriz de confusión de los árboles de decisión clasificación II	96
Tabla 28. Clasificación I RNA sin características de elastografía grupo 1	96
Tabla 29. Clasificación I RNA sin características de elastografía grupo 2	97

Tabla 30. Matriz de confusión de las RNAs sin características de elastografía clasificación I.....98

Tabla 31. Clasificación II RNA sin características de elastografía grupo 198

Tabla 32. Clasificación II RNA sin características de elastografía grupo 299

Tabla 33. Matriz de confusión de las RNAs sin características de elastografía clasificación II.....99

Tabla 34. Clasificación UENO..... 100

LISTA DE ANEXOS

Anexo A. Técnicas de elastografía e imagenología.....	111
Anexo B. Manual de instalación.....	115
Anexo C. Construcción librería compartida en Matlab.....	124
Anexo D. Algoritmo filtro wavelet.....	126
Anexo E. Algoritmos para la extracción de características.....	129
Anexo F. Matriz de confusión.....	134
Anexo G. Manual de usuario.....	136
Anexo H. aporte de las características elastográficas tomando el total de características extraídas.....	147

GLOSARIO

Biopsia: Examen microscópico de un trozo de tejido o una parte de líquido orgánico que se extrae de un ser vivo.

Falsos positivos: Error por el cual al realizar una exploración física o una prueba complementaria su resultado es normal o no detecta la alteración, cuando en realidad hay una enfermedad en el paciente.

Nódulo: Masa de células redonda, abultada y dura de tejido que puede ser normal o patológica.

Técnica de cribado: Es una estrategia aplicada sobre una población para detectar una enfermedad en individuos sin signos o síntomas de esa enfermedad.

Mastectomía: Operación quirúrgica que consiste en la extirpación de la glándula mamaria o de una parte de ella.

Tisular: De los tejidos del organismo o relacionado con ellos.

Anecogénico: Que no refleja los ultrasonidos y no origina ecos, por lo cual no puede ser visualizado por medio de técnicas de ecografía o ultrasonografía.

Plano cutáneo: Referente a la piel.

Hipoecoico: Estructura que es menos reflexiva que la estructura circundante que resulta en una imagen más oscura que aparece en comparación a las estructuras circundantes.

Parénquima: Tejido esencial de determinados órganos, como los pulmones, el hígado o los riñones.

Lobulillo: Subdivisión de un lóbulo en el seno.

Epitelio: Tejido constituido por células íntimamente unidas, planas o prismáticas, que recubre la superficie externa del cuerpo y de ciertos órganos interiores.

Ecogénico: Que refleja las ondas de ultrasonido.

Calcificaciones: El calcio es un constituyente del cuerpo y tiene un papel importante en su funcionamiento. Está presente en los huesos y en los dientes. Cuando se encuentra en los tejidos u órganos en los que normalmente no está presente, se le llama calcificación.

Fibrótico: Desarrollo en exceso de tejido conectivo fibroso en un órgano o tejido como consecuencia de un proceso reparativo o reactivo.

Cribiforme: Que tiene numerosos agujeros.

Neoplasia: Formación anormal en alguna parte del cuerpo de un tejido nuevo de carácter tumoral, benigno o maligno.

Espiculado: Tumor mamario que presenta unos contornos de forma espicular o estrellada.

Opacidad: Cualidad de opaco.

RESUMEN

Para diagnosticar un cáncer una de las técnicas más utilizadas es la imagenología. La imagenología comprende técnicas tales como la mamografía, la ecografía, la resonancia magnética, entre otras. Con estos métodos se obtiene un diagnóstico aproximado de la lesión sospechosa, el cual se debe confirmar mediante la realización de una biopsia. La principal desventaja de estas técnicas es su difícil interpretación, por lo que los profesionales del diagnóstico toman mucho tiempo en el análisis y no logran dar diagnósticos precisos generando falso positivos y falsos negativos a pesar de su experiencia.

Una nueva técnica para mejorar el diagnóstico es la elastografía ultrasónica, esta evalúa objetivamente las consistencias o durezas de los tejidos y añade información a las características derivadas de la ecografía permitiendo mejorar los resultados. Los sistemas de diagnóstico asistido por computador (CAD) son una herramienta utilizada para el apoyo al profesional del diagnóstico, un sistema CAD analiza las imágenes médicas y brinda información de las áreas sospechosas.

En este trabajo se presentan y evalúan diferentes técnicas del procesamiento y clasificación de imágenes de ecografía y elastografía, con el objetivo de construir una aplicación software para el apoyo al diagnóstico de cáncer de mama. Se han construido diferentes algoritmos que permiten la extracción de características descriptivas, morfológicas, de textura y de elastografía, en la etapa de clasificación se han evaluado diversos clasificadores lineales, no lineales, supervisados y no supervisados.

A pesar de algunas limitaciones se han obtenido resultados que demuestran que es posible la diferenciación de lesiones malignas de benignas y de ciertas patologías a partir de imágenes de ecografía y elastografía. La aplicación desarrollada realiza distintas clasificaciones y da a conocer información de lesiones similares a la diagnosticada.

1. INTRODUCCIÓN.

1.1. JUSTIFICACIÓN

Dentro de las enfermedades más comunes y mortales se encuentra el cáncer, en Colombia en el periodo 2007-2011 se estimaron 29.734 casos nuevos de cáncer por año en hombres y 33.084 en mujeres. En las mujeres, principalmente se localizaron en mama, cuello del útero, colon-recto, estómago y tiroides. La tasa de mortalidad anual en mujeres fue de 16.572, una de las principales causas fue el cáncer de mama [1].

Debido a esto, los estudios se están enfocando en la detección precoz del cáncer, actividad tediosa y subjetiva que depende de la experiencia y pericia del radiólogo. Una de las herramientas de apoyo más utilizada en el análisis de imágenes de mama es el diagnóstico asistido por computador (CAD), la mayoría de los estudios realizados utilizan imágenes mamográficas. La mamografía es insuficiente en algunos casos, por este motivo se requiere de técnicas complementarias como la ecografía.

En la Universidad del Cauca se vienen desarrollando trabajos alrededor de la detección de cáncer empleando imágenes de ultrasonido, donde hasta el momento se cuenta con algoritmos para obtención de elastogramas. Partiendo de la investigación realizada, se planeó el desarrollo de una aplicación software para el apoyo al diagnóstico de cáncer de mama con el fin de complementar los estudios mamográficos, mitigar el uso de técnicas invasivas que además de aumentar los costos, generan cargas psicológicas en la paciente, reduciendo el número de falsos positivos en el diagnóstico de cáncer de mama.

1.2. ALCANCES Y OBJETIVOS

El interés en el diagnóstico de lesiones de mama ha generado la investigación y desarrollo de nuevas técnicas, una de estas es la elastografía ultrasónica que permite la extracción de características para la clasificación de las diferentes lesiones que se pueden presentar en la mama. Con el fin de desarrollar una aplicación software para el diagnóstico se utilizan las características y virtudes que ofrecen las imágenes de elastografía y ecografía. Se plantean los siguientes objetivos generales y específicos a seguir durante la ejecución del proyecto.

1.2.1. Objetivo General

Proponer una herramienta software para apoyo al diagnóstico de lesiones de mama utilizando elastografía ultrasónica.

1.2.2. Objetivos Específicos

- Diseñar algoritmos computacionales basados en técnicas de procesamiento de imágenes digitales, que permitan clasificar lesiones de mama a partir de las imágenes de ecografía y de elastografía mamarias.
- Construir una aplicación software para apoyo al diagnóstico de lesiones de mama, que permita la comparación entre casos similares ya diagnosticados y el caso actual a diagnosticar.
- Verificar la concordancia entre los resultados diagnósticos que entrega la aplicación desarrollada, y los resultados diagnósticos reales de pacientes; mediante el uso de bases de datos disponibles para investigación en elastografía.

1.3. ESTRUCTURA DE LA TESIS

A continuación, se realiza una breve descripción de los capítulos que conforman la tesis: (1) Introducción; (2) Marco teórico; (3) Implementación del proceso CAD; (4) Desarrollo de la aplicación software; (5) Resultados; (6) Conclusiones y trabajos futuros.

Capítulo 1. Este capítulo contiene la justificación y los objetivos por los cuales se desarrolló el proyecto.

Capítulo 2. En el desarrollo de este capítulo se muestran conceptos del diagnóstico y la clasificación de lesiones de mama desde un enfoque médico.

Capítulo 3. En este segmento de la tesis se plasma el proceso de implementación de los algoritmos utilizados en el proceso de diagnóstico asistido por computador (CAD), además de todas las herramientas software y las diferentes librerías que se utilizaron a lo largo del proyecto.

Capítulo 4. En este capítulo se presenta el desarrollo de la aplicación software teniendo en cuenta los algoritmos del capítulo 3. Se usa una base de datos de imágenes de ultrasonido.

Capítulo 5. En este capítulo se presentan y comparan los resultados de los algoritmos utilizados para la clasificación de lesiones, con los resultados reales obtenidos de la base de datos.

Capítulo 6. En este capítulo se plasman las conclusiones y se proponen posibles trabajos a futuro.

2. MARCO TEORICO.

2.1. REVISIÓN DE LA DETECCIÓN DE LESIONES DE MAMA

Dentro del diagnóstico de lesiones de mama se encuentran técnicas como autoexploración manual, examen clínico [2], biopsia, pruebas de laboratorio e imagenología. La imagenología hace uso de diferentes métodos para crear imágenes del interior del cuerpo [3] con fines diagnósticos o de tratamiento de enfermedades. Dentro de la imagenología existen diferentes técnicas, algunas de ellas se exponen a continuación:

2.1.1. Mamografía

La mamografía es una herramienta relativamente eficaz para la detección precoz de cáncer de mama. En un mamograma se puede encontrar cambios en los senos que podrían ser causados por un cáncer mucho antes de que se presenten síntomas físicos, esto convierte a la mamografía en la principal técnica inicial para el diagnóstico de lesiones de mama en mujeres mayores de 30 años. El procedimiento produce una imagen en blanco y negro que es interpretada por un radiólogo [4]. Los mamogramas tienen algunas falencias [5], como: (i) la detección de lesiones en mamas densas [6], (ii) los rayos x que no pueden ser aplicados a mujeres embarazadas o en estado de lactancia debido a la radiación ionizante, (iii) el proceso de mamografía es incómodo y doloroso ya que se ejerce presión sobre los senos lo cual resulta una desventaja para ser realizado en pacientes con implantes mamarios [7].

2.1.2. Imágenes por resonancia magnética

En esta técnica se generan imágenes del interior del cuerpo por radiofrecuencia [8] en lugar de rayos x. La MRI (imagen por resonancia magnética) se utiliza principalmente en dos casos: para medir el tamaño del tumor y ver si hay otros tumores en mujeres que han sido diagnosticadas con cáncer de seno, además en mujeres que presenten alto riesgo de padecer cáncer de mama.

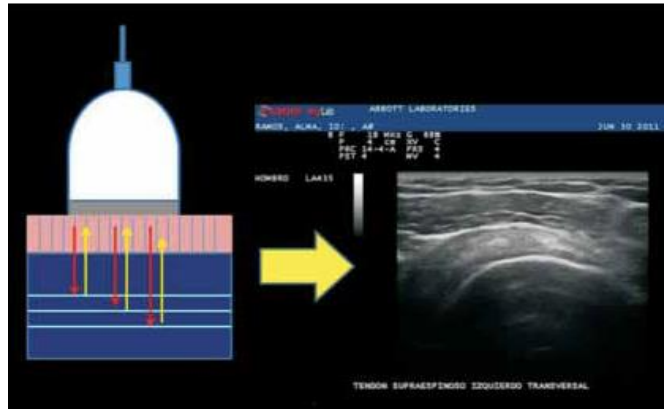
Las imágenes por resonancia magnética no suelen ser utilizadas por sí solas como una forma de detección de cáncer de mama, ya que son más propensas a generar falsos positivos, generando estudios adicionales como las intervenciones invasivas [4].

2.1.3. Ecografía (Ultrasonido)

La ecografía es una técnica que hace uso de las ondas sonoras para generar imágenes del interior del cuerpo [9], para esto se hace uso de un instrumento denominado transductor, el cual con ayuda de un gel que se aplica en la piel de la mama, muestra la estructura del tejido subyacente. Para la construcción de las imágenes el transductor emite las ondas sonoras y recoge los ecos que rebotan de los tejidos corporales [4], [10] y [11].

Existen tres tipos de ecografías: Modo A o de amplitud, usado principalmente en la diferenciación entre estructuras quísticas y sólidas, Modo M se emplea para las estructuras en movimiento y el Modo B es la representación pictórica de la suma de los ecos en diferentes direcciones (axial y lateral), ver figura 1, el modo ultrasonográfico más utilizado es el Modo B dinámico donde se obtienen varias imágenes por segundo [9].

Figura 1. Ecografía en modo B [9]



Comúnmente la ecografía es usada para el análisis de posibles lesiones en la mama que se han detectado anteriormente durante un examen físico o mamograma [4], a diferencia de un mamograma se puede indicar si el nódulo es un quiste o un tumor [3], con esto se demuestra que es un método complementario de uso esencial [12], [8]. Esta herramienta es cada día más valiosa debido a su amplia disponibilidad, su bajo costo y su facilidad de ejecución, sin embargo la ecografía presenta dificultad en la detección de microcalcificaciones, lo que es un factor importante en la detección temprana de cáncer [12], no reconoce algunos cánceres o puede presentar dificultad en el diagnóstico debido a la semejanza morfológica entre tumores benignos y malignos [13], generando problemas de falsos positivos como todas las técnicas de imagenología [4], esta técnica depende del nivel de destreza del operador del transductor y de la tecnología disponible [9], [13], al igual que la pericia del radiólogo en la lectura de las imágenes y requiere de mucho tiempo de exploración [12], por estas razones el uso de dos o más técnicas en la detección de cáncer de mama se considera necesario [4], por ejemplo la unión de la ecografía, la mamografía y la exploración clínica [12].

El uso de ultrasonido como técnica de cribado tiene ventajas sobre la mamografía en los siguientes casos:

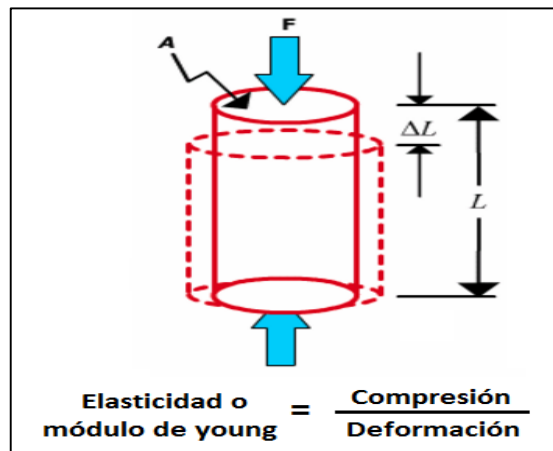
1. En pacientes jóvenes menores de 30 años que presentan síntomas, ya que los senos suelen ser muy densos y los tumores benignos se pueden reconocer fácilmente.
2. En mujeres gestantes y lactantes se presentan inconvenientes al realizar un mamograma, generando molestias en la paciente, debido a que las mamas aumentan su densidad y sensibilidad.
3. En los casos en los que se presenta inflamación en la mama, la toma de una ecografía causa menos dolor y la identificación de las lesiones se facilita más que en la mamografía.
4. En pacientes mastectomizadas, para identificar la reaparición de tumores malignos o posibles complicaciones de la cirugía [12].

La ecografía es considerada una técnica complementaria fundamental de los estudios realizados en una mamografía [14], [15], por ejemplo en pacientes con implantes mamarios es muy útil en el estudio del parénquima mamario, además permite la identificación de componentes introductoros o sólidos asociados a las microcalcificaciones, el estudio de mujeres con secreción por el pezón, mujeres con mamografía negativa y lesión palpable, mujeres con enfermedad metastásica conocida, mujeres con factores de riesgo elevado y mamas densas, mujeres con alta sospecha de malignidad o carcinoma de mama confirmado, seguimiento evolutivo de pacientes sometidos a diferentes tratamientos y en la guía de procedimientos intervencionistas [12].

2.1.4. Elastografía

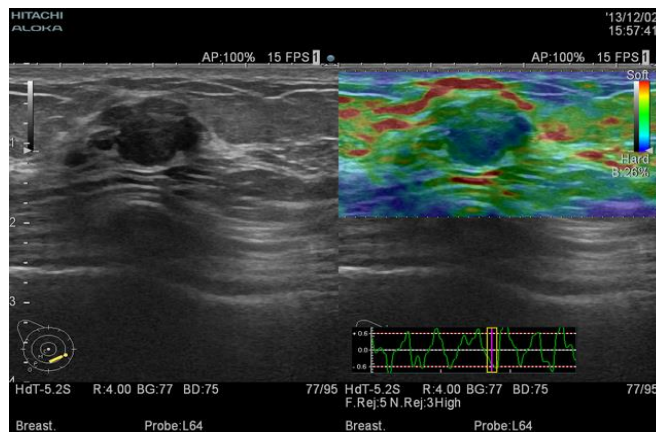
La ecografía de mama ha tenido una continua evolución partiendo del doppler, hasta llegar al empleo del contraste endovenoso y la elastografía. La elastografía es una técnica en la que se miden las características elásticas de los tejidos mamarios (“strain”), después de aplicar una compresión externa utilizando equipos de ultrasonidos convencionales la principal información obtenida de elasticidad tisular es el desplazamiento relativo del tejido el cual indica la dureza del tejido. La elasticidad puede caracterizarse mediante “el cociente entre la compresión realizada contra un tejido y la deformación que se consigue con ella”, conocido como el módulo de Young (Figura 2). Cuando el tejido evaluado es blando, es probable que se trate de una lesión benigna, de lo contrario si el tejido es duro es probable que la lesión sea maligna. Para tener un estudio fiable, con el objetivo de contar con un diagnóstico más preciso, se hace uso de la elastografía, junto con el estudio ecográfico mamario convencional [13], [16], [17]. Gracias a las bondades de la elastografía y la ecografía mamaria el trabajo realizado se fundamentó en estas técnicas, por lo que se ampliarán los conceptos de funcionamiento.

Figura 2. Módulo de young



Para realizar un elastograma normalmente se evalúa la lesión en Modo B, luego se realiza el estudio elastográfico, los datos elastosonográficos se traducen a una escala de 256 colores que oscilan entre rojo y azul; representando los tejidos blandos y duros respectivamente sobre la imagen en Modo B, a cada pixel de la imagen elastográfica se le asigna uno de estos 256 colores (Figura 3). El tamaño de la región de interés (ROI), donde se encuentra la lesión debe ser tal que la lesión no ocupe más de un tercio del área total de la ROI [17].

Figura 3. Elastograma [18].



Existen diversas modalidades de elastografía: la semicuantitativa, la cuantitativa y la cualitativa. La elastografía semicuantitativa inicialmente toma los datos correspondientes a la anatomía tisular antes de la compresión, después se aplica una presión mediante el compresor o transductor ecográfico y se obtienen otros datos de la anatomía tisular (deformación), ya obtenidos los dos mapas anatómicos se obtiene por medio de comparación el desplazamiento del tejido deformado que se refleja en un mapa de colores. La tensión tisular mecánica se calcula con el gradiente del desplazamiento.

Por otra parte, la elastografía cuantitativa mide el desplazamiento independientemente de la presión aplicada enviando micro pulsos acústicos con niveles bajos de energía a los tejidos, creando un mapa del desplazamiento del

tejido circundante. Existen tres subgrupos de imágenes dentro de la elastografía cuantitativa: la cualitativa genera un mapa de colores de la lesión con respecto al tejido adyacente, la cuantitativa, que proporciona valores sobre la mayor o menor deformidad en una área de interés seleccionada y por último la combinación de las dos anteriores [17], [19]. En el Anexo A se mencionan algunas de las técnicas de elastografía cuantitativa y cualitativa además de técnicas de imagenología.

Gracias a los avances tecnológicos día a día surgen nuevas ideas para el diagnóstico y detección de cáncer a partir de la evaluación de imágenes médicas, algunas de ellas se fundamentan en técnicas ya establecidas como las expuestas anteriormente. Dentro de los estudios experimentales más relevantes para la mama encontramos los estudios ópticos por imagen en donde se envía luz hacia el seno y posteriormente se mide la cantidad de luz que regresa o pasa a través del tejido y variaciones de la mamografía en donde se utilizan azúcares con partículas radiactivas para la detección de células cancerosas (tomografía por emisión de positrones) [4]. La unión de técnicas ha generado nuevos métodos híbridos en los cuales se demuestra la forma (radiología e imagen), con lo que se generan imágenes que demuestran el funcionamiento de las células (medicina nuclear), lo cual tiene como principal ventaja la unión de la anatomía con los procesos fisiológicos, siendo ahora conocido con el nombre de radiología e imagen molecular [8].

2.1.5. Biopsia y pruebas de laboratorio

Si el resultado de cualquiera de las técnicas por imagenología es sospechoso se procede a realizar una biopsia que es la única manera de comprobar si se trata o no de cáncer. Durante esta prueba, se realiza la extracción de células del área sospechosa para ser estudiadas en un laboratorio. Uno de los tipos más comunes de biopsia es la biopsia con aguja, en esta se usa una aguja guiada por el médico el cual utiliza palpación o imágenes (como un mamografía o ecografía) para

guiarse y extraer una porción del área anormal, si esta muestra extraída no genera una respuesta concreta, se procede a realizar cirugía en donde se retira todo o parte de la protuberancia sospechosa [3].

2.2. DIAGNÓSTICO DE LESIONES DE MAMA EN IMÁGENES DE ULTRASONIDO

Teniendo en cuenta lo mencionado anteriormente se procede a escoger el ultrasonido como método de imagenología para realizar una herramienta software para apoyar el diagnóstico de cáncer de mama, sabiendo que este es la base para el desarrollo de elastogramas y recordando que ayuda a distinguir lesiones malignas de benignas, a la detección temprana y es no invasivo.

A continuación, se realiza una descripción de las características ecográficas que presentan las lesiones de mama más comunes.

2.2.1. Caracterización de las lesiones de mama en imágenes ecográficas

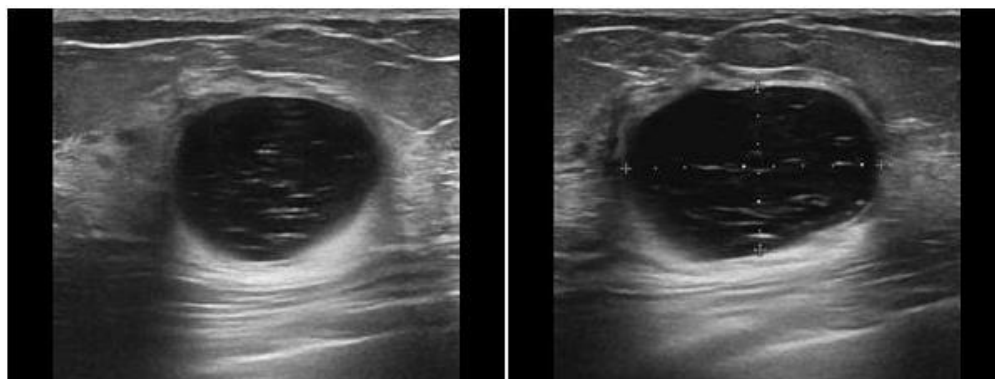
En la lectura de imágenes de ultrasonido (US) tanto radiólogos como técnicos especialistas tienen en cuenta ciertas características que ayudan a determinar la malignidad o benignidad de las lesiones encontradas permitiendo realizar su clasificación, dicha clasificación es subjetiva por lo que en algunos casos se generan falsos positivos o falsos negativos. Las lesiones de mama se dividen en benignas y malignas.

2.2.1.1. Lesiones benignas

- **Quistes:** el US es el método de elección para la diferenciación entre una lesión quística y una masa sólida, las lesiones quísticas se pueden categorizar como quistes simples y quistes complicados o complejos, los quistes complejos

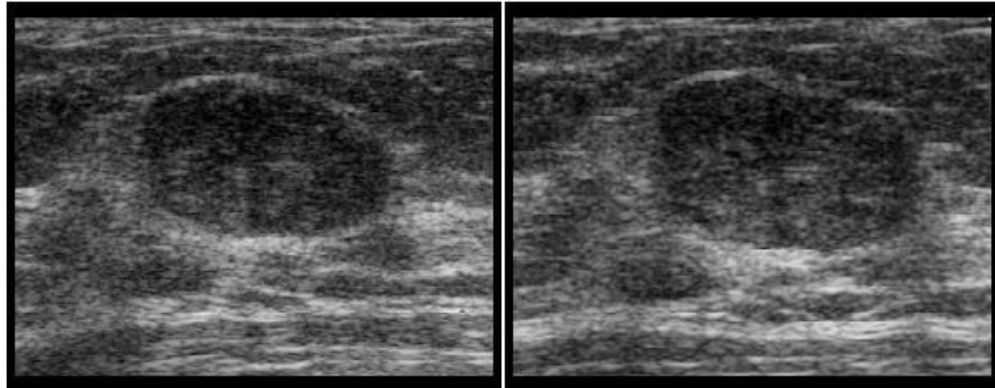
poseen un componente líquido y uno sólido, por esta razón pueden ser signos de lesiones malignas. Una lesión quística debe ser anecogénica, redonda u oval, los quistes simples son observados con margen nítido, regular, realce acústico posterior y orientación paralela (Figura 4), por otra parte los quistes complejos presentan características contrarias [7] , [20].

Figura 4. Deformación de un quiste durante la ecografía [18]



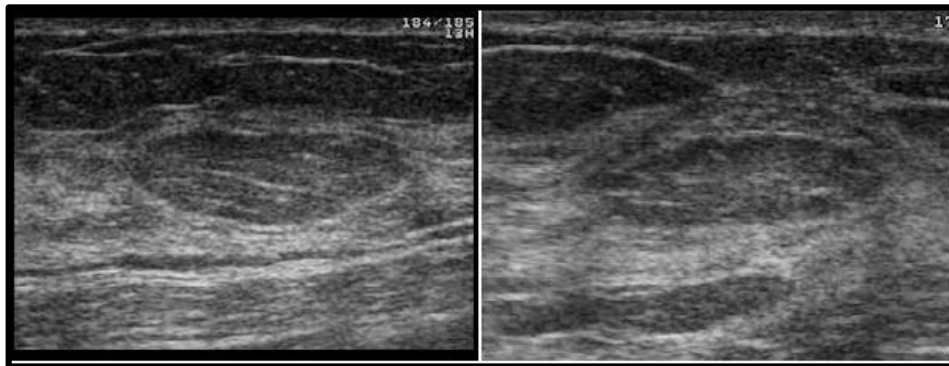
- **Fibroadenoma:** lesión más común en adolescentes y jóvenes, los fibroadenomas clásicos y los fibroadenomas celulares son aquellos que son benignos. Inician con la presencia de un nódulo liso redondo, de márgenes circunscritos, orientación paralela al plano cutáneo y móvil que puede percibirse firme o de consistencia elástica (Figura 5), en general los fibroadenomas son hipoecoicos en relación con el parénquima fibroglandular e isoecoicos con los lobulillos grasos de la mama, con ecoestructura interna heterogénea, la masa es ovalada y bien definida, de crecimiento rápido, con compresión del tejido mamario circundante, muestra sombra acústica posterior que no se debe a calcificaciones [7], [21], [22], [23].

Figura 5. Deformación de un fibroadenoma durante la ecografía [18]



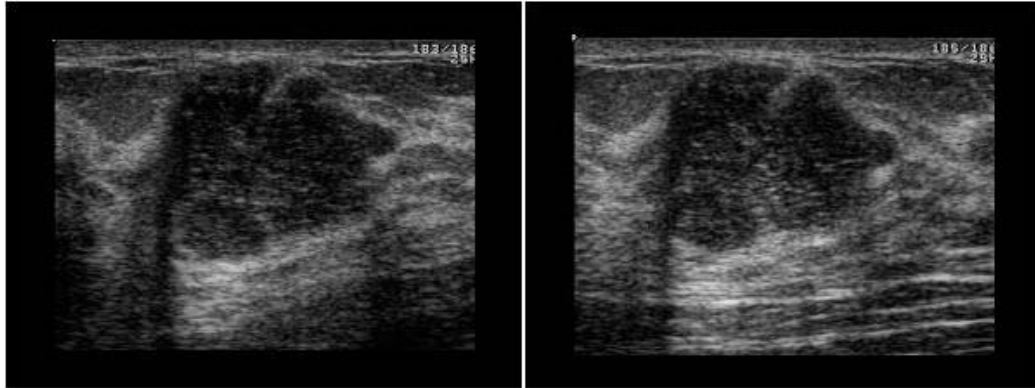
- **Fibroadenolipoma y lipoma:** masa poco frecuente formada de tejido adiposo epitelial y fibroso, se presenta sin síntomas en personas mayores, se distinguen como lesiones heterogéneas con componentes ecogénicos por la grasa (Figura 6), pueden presentar sombra acústica posterior, los lipomas presentan cápsula fina y ecos moderados internos [7].

Figura 6. Deformación de un lipoma durante la ecografía [18]



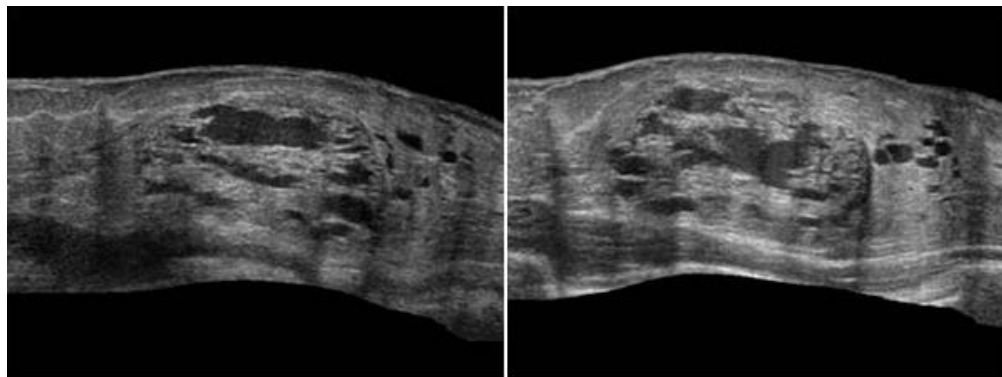
- **Tumor phyllodes:** lesión rara, tiende a ser redondeada y lobulada, de consistencia elástica, de bordes bien definidos y puede confundirse con un fibroadenoma gigante (Figura 7), tiene apariencia ecográfica inespecífica, se visualiza con ecogenicidad interna heterogénea y realce acústico superior y áreas hipoecoicas. [7], [24]

Figura 7. Deformación de un tumor phyllodes durante la ecografía [18]



- **Galactocele:** A menudo se presenta en pacientes durante la lactancia o posterior a ella [7]. En sus primeras etapas de desarrollo el galactocele está compuesto por leche en su mayoría, se manifiesta como una masa bien definida, con reforzamiento acústico posterior y material ecogénico (Figura 8). Por otra parte en sus etapas más avanzadas aparece como masa heterogénea, con mezcla de baja y alta ecogenicidad interna [25].

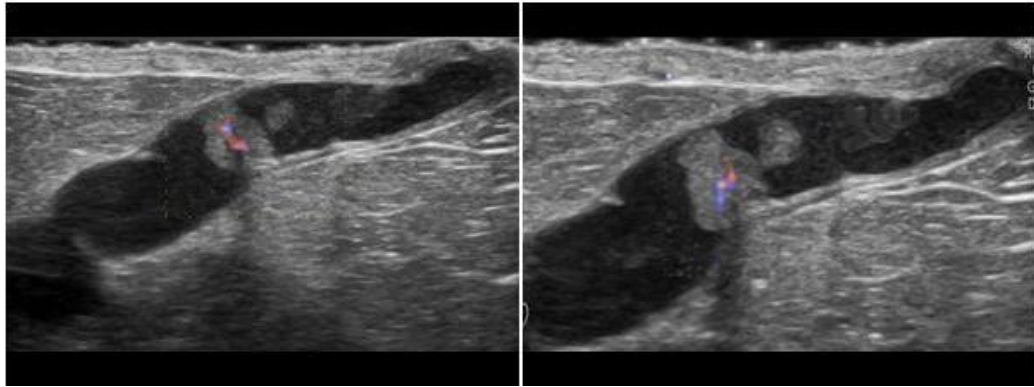
Figura 8. Deformación de un galactocele durante la ecografía [18]



- **Papilomatosis intraductal:** neoplasia resultante de una proliferación epitelial dentro de un conducto galactóforo, se evidencian como lesiones sólidas bien circunscritas (Figura 9), con áreas hipoeecogénicas redondeadas u ovaladas,

comúnmente bien delimitadas o masas quísticas con componentes sólidos [7], [26], [23].

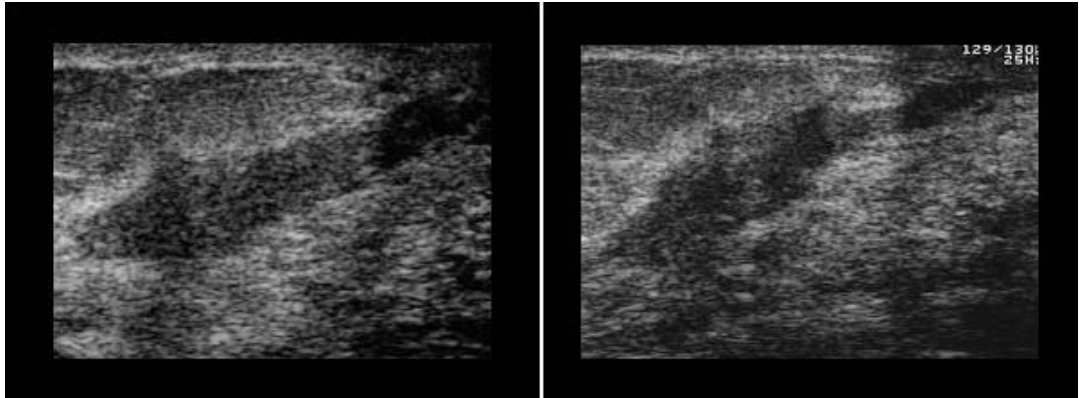
Figura 9. Deformación de un papiloma intraductal durante la ecografía [18]



2.2.1.2. Lesiones malignas

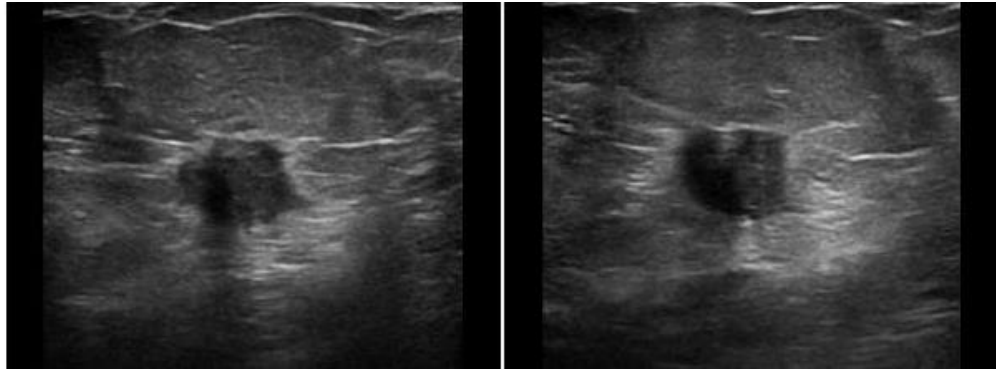
- **Cáncer del epitelio ductal:** Los pacientes generalmente se presentan con una masa palpable, con descarga a través del pezón. El cáncer del epitelio ductal es identificado como una masa microlobulada, con presencia de pequeñas imágenes ecogénicas sin transmisión de sombra acústica, hipoecoica con extensión ductal y calcificaciones punteadas (Figura 10), cuando no hay presencia de calcificaciones puede ser mal diagnosticada pues se visualiza como hipoecoica, redondeada y de bordes bien definidos, características propias de los quistes [7].

Figura 10. Deformación de cáncer del epitelio ductal durante la ecografía [18]



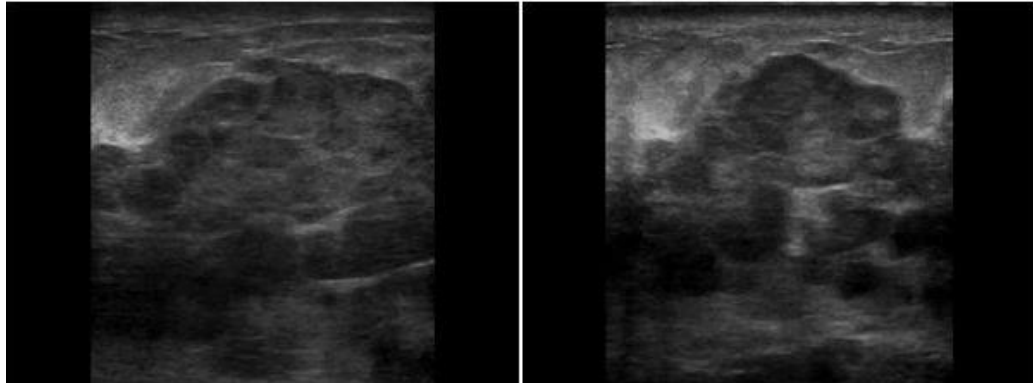
- **Comedocarcinoma:** masa palpable considerada la forma más común de cáncer del epitelio ductal, se identifica como tejido irregular hipoeico inespecífico, puede presentar sombra acústica si las calcificaciones son lo bastante densas o existen componentes fibróticos [7].
- **No comedocarcinomas:** son la segunda forma en frecuencia de cáncer del epitelio ductal, presentan células más uniformes y de menor tamaño, diferenciadas con espacios cribiformes [7].
- **Carcinoma ductal infiltrante o invasivo:** este tipo de cáncer genera imágenes ecográficas muy variables. La descripción más frecuente es la de una estructura hipoeicoica de forma irregular, que con frecuencia tiene margen anterior triangular; atenúa y dispersa el sonido (Figura 11), produciendo una sombra retro tumoral [7].

Figura 11. Deformación de cáncer ductal infiltrante durante la ecografía [18]



- **Carcinoma tubular:** forma bien diferenciada del carcinoma ductal invasivo, por lo general son tumores palpables, esta neoplasia puede ser pequeña, con distintos grados de irregularidad de sus márgenes [7].
- **Carcinoma papilar:** lesión intraquística, aunque quizá el quiste es solo una porción de un conducto dilatado quísticamente más que un lóbulo dilatado, presenta una forma homogénea bien delimitada y probablemente isoecoica o con refuerzo acústico posterior [7].
- **Carcinoma inflamatorio:** se manifiesta con calor, eritema y la piel engrosada (piel de naranja), dolor y borde palpable en el margen de la induración. La ecografía no presenta gran utilidad salvo que se sospeche una masa profunda o un absceso (Figura 12), el engrosamiento de la piel es demostrado por la separación de dos líneas ecogénicas normales por la presencia de una banda hipoeoica, también se reporta la presencia de áreas focales de transmisión de área acústica sin evidencia de tumoración [7].

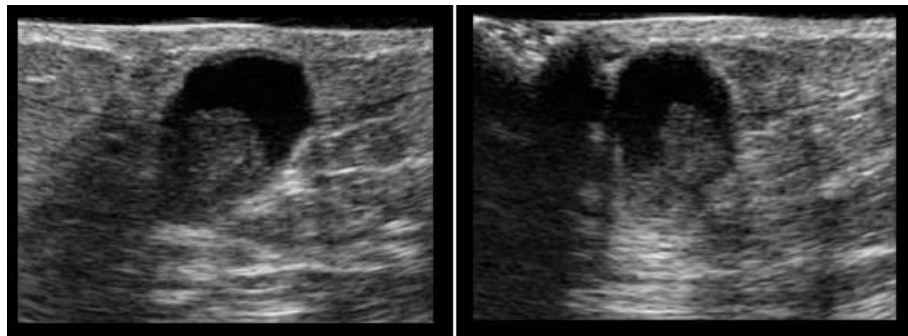
Figura 12. Deformación de un carcinoma inflamatorio durante la ecografía [18]



- **Carcinoma del lóbulo:**

- **Carcinoma lobular in situ:** rara vez forma masas palpables, los hallazgos ecográficos no están bien caracterizados (Figura 13), algunos han descrito un tejido hipoeicoico con sombra acústica posterior asociada [7].

Figura 13. Deformación de un carcinoma lobular in situ durante la ecografía [18]

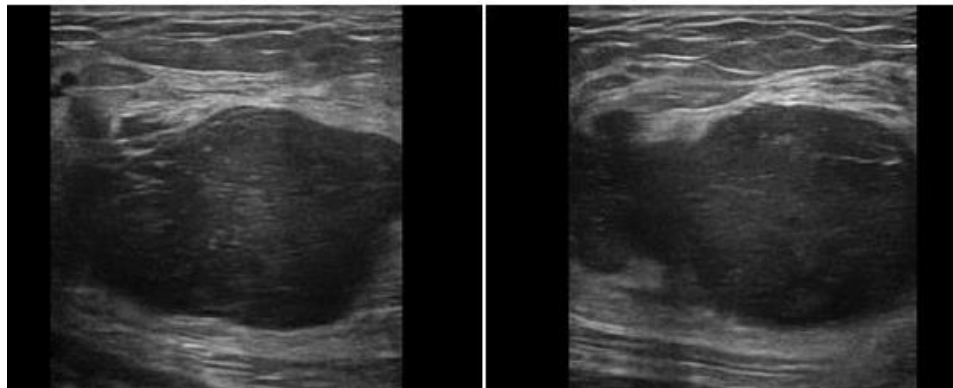


- **Carcinoma lobular infiltrante o invasivo:** no se puede diferenciar por ecografía del carcinoma ductal infiltrante, se identifica como un área de parénquima hipoeicoico con grados variables de sombra acústica posterior [7].

- **Tumores malignos del estroma**

- **Tumor phyllodes:** se presentan como una masa de crecimiento rápido, normalmente tienen contornos lisos y lobulados con cierta movilidad incluso en tumores de gran tamaño, los tumores mayores a 3 centímetros a menudo son malignos (Figura 14). La ecografía de los tumores phyllodes es idéntica a la de los fibroadenomas, por lo general son lesiones bien delimitadas que se distinguen sólo por su mayor tamaño y la presencia de ecos internos [7].

Figura 14. Deformación de un tumor phyllodes durante la ecografía [18]



- **Lesiones metastásicas de mama:** tiene características ecográficas que reflejan su forma y composición, suelen ser redondas u ovaladas, con cierta lobulación, ecos internos variables y reflexiones heterogéneas bastante prominentes [7].
- **Traumatismos:** pueden resultar en necrosis grasa como una masa indolora, dura que puede encontrarse adherida a la piel y producir retracción de esta, en la ecografía pueden presentarse como tumoraciones hipoecoicas, con septaciones o heterogéneas [7].

- **Cicatriz radial o lesión esclerosante compleja:** no es palpable, en la ecografía es identificada como lesión ecogénica espiculada que no evidencia tumoración central pudiendo presentar calcificaciones con sombra acústica posterior [7].

2.3. MÉTODOS PARA LA CLASIFICACIÓN DE LESIONES DE MAMA

2.3.1. BI-RADS Ecográfico

Es un método para la clasificación de lesiones de cáncer de mama desarrollado en 1991 por el Colegio Americano de Radiología (ACR), considerado el lenguaje universal del diagnóstico de la patología mamaria. Las características en las que se basa son forma, orientación, margen, patrón ecogénico y características posteriores [23], [27], [28]. En la tabla 1 se realiza una breve descripción de las características descriptivas y en la tabla 2 se presentan los criterios de clasificación de lesiones a partir de las características descriptivas.

Tabla 1. Características descriptivas de las lesiones nodulares [23]

CARACTERÍSTICAS	DESCRIPCIÓN
Forma	Redonda: cuando la masa tiene su diámetro anteroposterior igual al transversal (A)
	Oval: incluye las formas lobuladas, considerándose las lesiones con máximo tres lobulaciones (B)
	Irregular: cuando no corresponde a ninguna de las formas anteriores (C)
Orientación	Paralela: nódulo más ancho que alto según el eje mayor de la lesión con respecto al plano cutáneo. (D)
	Anti paralela: nódulo más alto que ancho según el eje mayor de la lesión con respecto al plano cutáneo (E)
Margen	Circunscrito: cuando está bien definido con una transición abrupta entre la lesión y el tejido que la rodea (F).

	<p>No circunscrito: cuando no se definen con precisión, pueden ser:</p> <ul style="list-style-type: none"> • <i>Indistinto:</i> no existe una clara delimitación entre el nódulo y tejido circundante. (G) • <i>Angular:</i> Se observan bordes puntiagudos, formando ángulos agudos. (H) • <i>Micro lobulado:</i> cuando se observan pequeñas lobulaciones que dan un aspecto ondulado. (I) • <i>Espiculado:</i> cuando se observan finas líneas que se proyectan del nódulo. (J)
Ecogenicidad	Anecoico: sin ecos internos. (A)
	Hiperecoico: cuando existe aumento de la ecogenicidad.(B)
	Quiste complejo y sólido: si tiene componente tanto anecoico como ecogénico. (C)
	Hipoecoico: presenta ecos de bajo nivel. (D)
	Isoecoico: tienen la misma ecogenicidad de la grasa.(E)
	Heterogéneo: combinación de las anteriores.
Acústicas posteriores	Ausencia: si no se visualiza sombra ni refuerzo acústico posterior y la ecogenicidad del tejido se mantiene igual a la del tejido adyacente y en la misma profundidad. (F)
	Refuerzo acústico posterior: aparece como una columna ecogénica detrás del nódulo, dado que no existe obstrucción de la transmisión sónica a su paso a través del mismo. (G)(I)
	Sombra acústica: atenuación posterior de la transmisión acústica. (H)(I)

Figura 15. Caracterización de las lesiones nodulares [28]

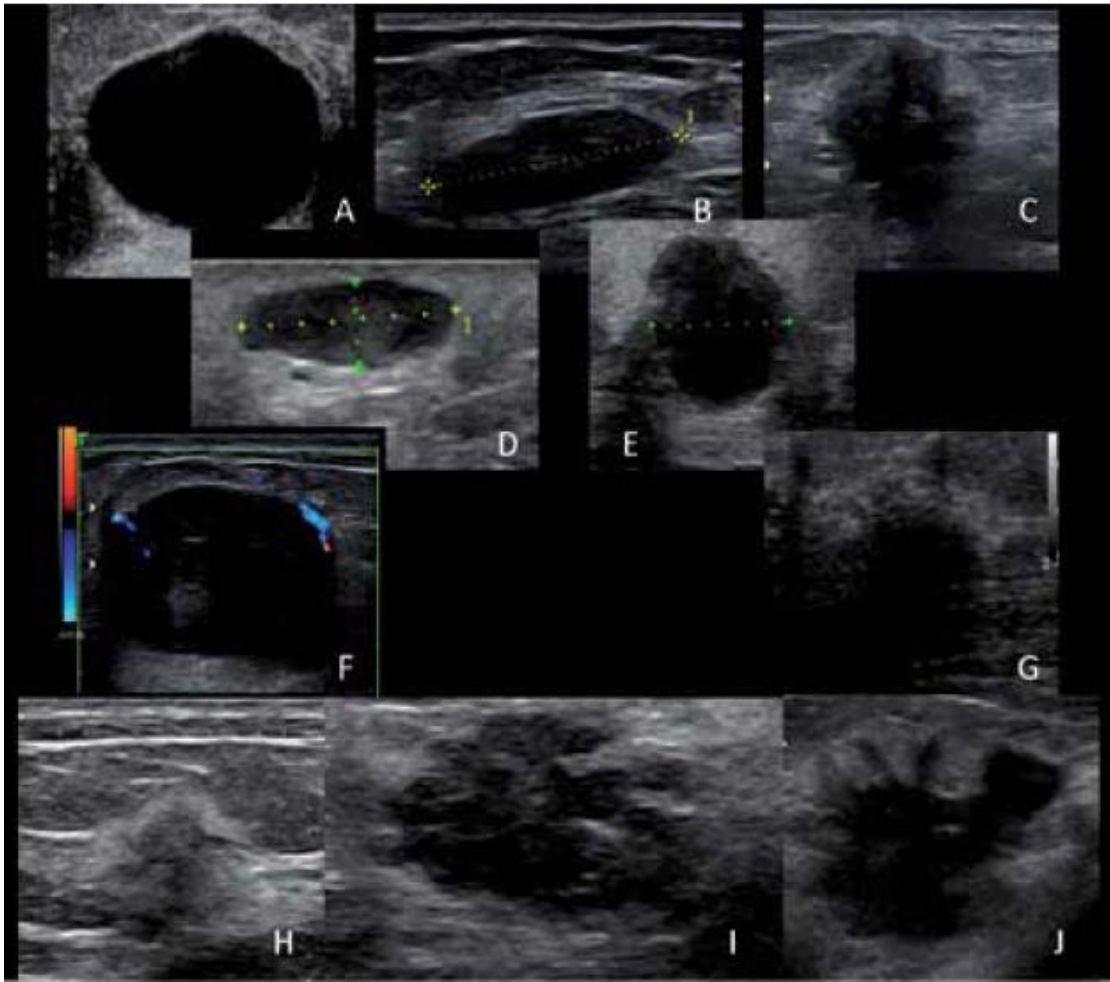


Figura 16. Características ecogénicas de las lesiones nodulares [28]

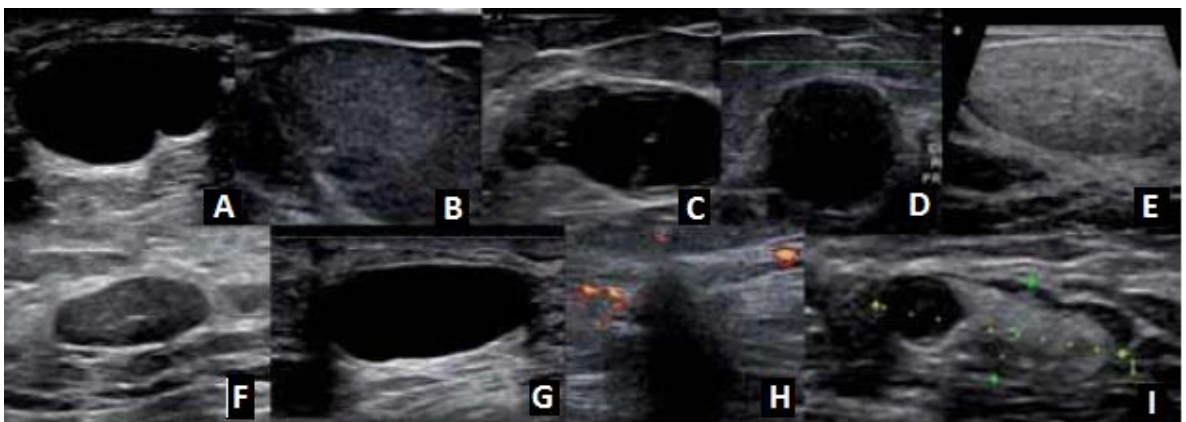


Tabla 2. Criterios de clasificación de masas

Descriptor	Características probablemente benigna	Características probablemente malignas	Características indeterminadas
Borde	Oval, redondeado	Irregular, lobulado	
Orientación	Paralelo a la piel	Perpendicular a la piel	
Márgenes	Circunscritos	Microlobulados, angulados o espiculados	
Ecogenicidad	Hiperecoico, anecoico	Complejo	Isoecoico, hipoecoico
Acústica posterior		Sombra, patrón heterogéneo	Reforzamiento acústico posterior

El método BI-RADS se divide en las siguientes categorías, en la tabla 3 se resume la probabilidad de cáncer de cada categoría.

- **Categoría 0:** se considera una categoría incompleta, para establecer una categoría necesita una evaluación adicional, bien sea mediante técnicas de imagen (proyecciones adicionales, ecografía) o comparación con mamografías anteriores.
- **Categoría 1:** normal, ningún hallazgo a destacar. Se recomienda seguimiento a intervalo normal.
- **Categoría 2:** normal, pero existen hallazgos benignos. Se recomienda seguimiento a intervalo normal. Las lesiones que se encuentran comúnmente en esta categoría son quistes simples, ganglios inflamatorios, hamartomas y lipomas.
- **Categoría 3:** hallazgos con una probabilidad de malignidad <2%. Se describen 3 hallazgos específicos: nódulo sólido circunscrito no calcificado, asimetría focal, micro calcificaciones puntiformes agrupadas y orientación paralela a la

superficie cutánea. [29] La mayoría de los fibroadenomas cuentan con características de categoría 3.

- **Categoría 4:** incluye aquellas lesiones que van a requerir intervencionismo, si bien tienen un rango de probabilidad de malignidad muy amplio (2-95%). Por ello, se sugiere una división en tres subcategorías: 4a: baja sospecha de malignidad donde se reconocen lesiones como fibroadenomas, papilomas, entre otros, 4b: riesgo intermedio de malignidad (requiere correlación radio-patológica), 4c: riesgo moderado de malignidad.
- **Categoría 5:** hallazgos típicamente malignos, con una probabilidad >95%. La actitud recomendada es tomar acciones apropiadas, las lesiones presentes en esta categoría son carcinomas invasivos, tubulares, necrosis grasa, cicatriz radiada y adenosis esclerosante.
- **Categoría 6:** lesiones con malignidad demostrada mediante biopsia, previa a terapias definitivas (cirugía, radioterapia o quimioterapia), y por lo tanto no se debe confirmar su malignidad. Se utiliza en casos de segundas opiniones o en la monitorización de la quimioterapia [6], [13], [30].

Tabla 3. Correlación entre categoría, lesión y probabilidad de cáncer [23]

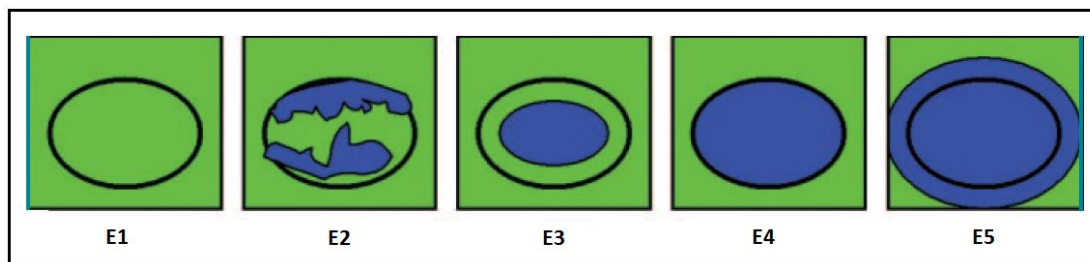
CATEGORÍA	LESIÓN	PROBABILIDAD DE CÁNCER
0	Imagen no diagnóstica.	
1	Hallazgos normales.	0 %
2	Quiste simple, quiste septo fino, fibroadenoma estable, quiste complicado no sintomático, nódulo ecogénico homogéneo y ectasia ductal simple.	0 %

3	Nódulo sólido, fibroadenoma no palpable, microquistes agrupados, quiste complicado aislado, necrosis grasa y ectasia ductal compleja.	> 0 % - ≤ 2 %
4	Nódulo sólido, fibroadenoma palpable, nódulo heterogéneo, nódulo sólido con flujo doppler, nódulo intraductal y lesión quística sólida.	4A: > 2 % ≤ 10 %
		4B: > 10 % ≤ 50 %
		4C: > 50 % ≤ 95 %
5	Nódulo microlobulado y nódulo con extensión ductal.	≥ 95 %
6		100 %

2.3.2. Escala de elasticidad UENO

Es la escala de elasticidad más utilizada, propuesta por Ithoh, *et al* que utiliza una puntuación de 1 a 5, donde una lesión con puntuación entre 1 y 3 es considerada benigna y con puntuación de 4 y 5 se considera maligna, ver figura 17. [31]

Figura 17. Escala UENO [31]



- **Puntaje 1 (E1):** indica una tensión o elasticidad uniforme en toda la lesión, que se muestra homogéneamente de color verde. Representa normalmente a un quiste o lipoma que puede corresponder a la categoría 1 o 2 de BI-RADS.

- **Puntaje 2 (E2):** indica una lesión fundamentalmente elástica con algunas zonas de ausencia de elasticidad, que se observa de manera heterogénea, lo cual indica un patrón de predominancia de tensión alta de la lesión correspondiente a una combinación de colores verde y azul. Características que identifican generalmente a los fibroadenomas, se pueden clasificar en la categoría 3 de BI-RADS.
- **Puntaje 3 (E3):** indica elasticidad en la periferia de la lesión y ausencia de elasticidad en el centro (periferia en verde y centro en azul), al igual que el puntaje 2 representa a los fibroadenomas.
- **Puntaje 4 (E4):** indica ausencia de elasticidad en toda la lesión, en el elastograma se observa una marca uniforme de color azul dentro de la margen visible de la lesión, de esta manera se reconoce un carcinoma lobular correspondiente a la categoría 5 de BI-RADS.
- **Puntaje 5 (E5):** indica ausencia de elasticidad en toda la lesión y en el área circundante, se visualiza una marca azul en todas partes de la lesión y en los tejidos adyacentes, lo cual describe un carcinoma invasivo, se puede clasificar en la categoría 5 de BI-RADS [17], [31], [32].

2.4. DIAGNOSTICO ASISTIDO POR COMPUTADOR (CAD)

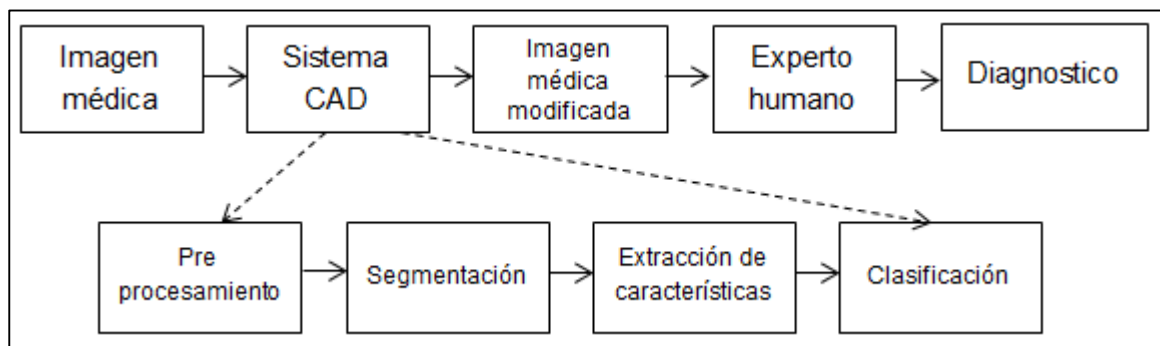
El diagnóstico asistido por computador es una herramienta software basada en algoritmos, aplicada sobre imágenes digitalizadas (RM, mamografía, US, radiografías convencionales, etc.) para el diagnóstico de lesiones, principalmente de mama, pulmón y colon, que integra el diagnóstico por imagen con la informática, el procesamiento de imágenes, reconocimiento de patrones y la inteligencia artificial [33]. La lectura de imágenes digitalizadas puede ser simple o doble, es decir, si la lectura la hace un radiólogo es simple y si es realizada por

dos radiólogos es doble, en este caso se estudia que CAD pueda sustituir uno de los radiólogos. [34]

CAD se utiliza para eliminar la variabilidad entre radiólogos y ayudarlos a realizar un diagnóstico preciso, su principal ventaja es que se pueden obtener características computacionales y estadísticas imposibles de lograr intuitivamente por los médicos, también se puede minimizar la dependencia natural del operador en las imágenes por ultrasonido [35].

La lectura de imágenes digitales por medio de sistema CAD precisa de las siguientes etapas (Figura 18):

Figura 18. Etapas de un sistema CAD Etapas de un sistema CAD



2.4.1. Pre procesamiento de imagen

Al tomar una imagen de ultrasonido se introduce ruido y se disminuye la calidad de la imagen. Por este motivo es necesario reducir el ruido característico de imágenes de ultrasonido conocido como ruido moteado o speckle, lo que permite obtener áreas más homogéneas que pueden ser delineadas con menor esfuerzo, eliminando información que no se considera relevante aunque se puede comprometer la detección de pequeños elementos importantes [36]. Existen diferentes enfoques para realizar el pre procesamiento de imágenes entre ellos se encuentra la modificación global del histograma, donde se realiza un cambio por

medio de aproximaciones en la distribución de los tonos de la imagen con el fin de mejorarla, el procesamiento local: aplica técnicas en ventanas locales de la imagen, el procesamiento multi escala, basado en el uso de pequeñas ondas que modifican los coeficientes para realzar las anomalías y posteriormente devuelven la imagen a su dominio original [37], técnicas basadas en lógica difusa [31] y operaciones morfológicas que se basan en la forma que permiten la eliminación del ruido, el hallazgo de protuberancias o agujeros y el aislamiento de elementos individuales. Los operadores morfológicos más utilizados son la dilatación y la erosión, la dilatación expande la frontera del objeto y la erosión la reduce [38] [39]. Algunas técnicas para reducir el speckle y mejorar las características de la imagen son las técnicas de filtrado y técnicas en el dominio wavelet.

2.4.1.1. Técnicas de filtrado

Los filtros más utilizados para la reducción de speckle son el Lee, Frost, Kuan basados en el filtro de media y Gamma MAP ya que estos además de eliminar el ruido preservan características importantes en el diagnóstico de lesiones como el gradiente, retención de bordes y líneas de la imagen.

- **Filtro de media:** El filtro de media reemplaza cada píxel por el valor promedio de las intensidades en su vecindad. Se puede reducir localmente la varianza y es fácil de implementar. Tiene el efecto de suavizar y difuminar la imagen, es óptimo para el ruido aditivo gaussiano en el sentido del error cuadrático medio. La imagen de ruido moteado es un modelo multiplicativo con ruido no gaussiano, por lo tanto, el filtro de media simple no es eficaz en este caso [40] [41] [42].
- **Filtro de mediana:** es un caso especial de filtro de orden estadístico que opera, cambiando el valor de intensidad de cada píxel analizado, con la finalidad de hacer que los puntos de intensidad se hagan muy parecidos a sus

vecinos, eliminando puntos de muy alta y muy baja intensidad reemplazando cada pixel de la imagen por el valor de la mediana, calculada a partir de los valores de la máscara que lo rodea. Este filtro conserva la nitidez de los bordes y produce un desenfoque menor que el filtro de media, especialmente es eficaz cuando la imagen se ve afectada por ruido impulsivo. [40] [41] [42]

- **Filtro Lee:** Se hace uso del criterio de error medio cuadrático mínimo (MMSE), de esta manera se realiza una aproximación del modelo multiplicativo a un modelo lineal. Se representa con la siguiente ecuación:

$$\hat{R}(t) = \bar{I}(t) + [I(t) - \bar{I}(t)]W(t)$$

Donde $\hat{R}(t)$ es la imagen libre de ruido, $\bar{I}(t)$ es la media local de las intensidades dentro de la ventana de filtro, $I(t)$ es la intensidad de la imagen con ruido y $W(t)$ es la función de coeficiente que se define como:

$$W(t) = 1 - \frac{C_u^2}{C_I^2(t)}$$

Donde

$$C_u^2 = \frac{\sigma_Z^2}{\bar{Z}^2}$$

$$C_I^2(t) = \frac{\sigma_I^2(t)}{\bar{I}^2(t)}$$

Donde $\sigma_I^2(t)$ es la varianza de las intensidades dentro de la ventana de filtro, σ_Z^2 y \bar{Z} son la varianza y la media de la intensidad en una área homogénea de la imagen respectivamente. [40] [41] [42]

- **Filtro Kuan:** Este reduce el Speckle en áreas homogéneas. El filtro Kuan es una generalización del filtro Lee donde el valor de $W(t)$ se define como:

$$W(t) = \frac{1 - \frac{C_u^2}{C_I^2(t)}}{1 + C_u^2}$$

En las regiones homogéneas el valor de $W(t)$ se aproxima a 0 actuando como un filtro de media. Por otro lado, en las áreas donde existen bordes el valor de $W(t)$ se aproxima a 1 presentando la imagen original ya que los filtros se comportan como un filtro pasa todo. [40] [41] [42]

- **Filtro Frost:** Este filtro elimina el ruido y la opacidad de la imagen, conserva la forma de los bordes a diferencia de otros, debido a que la implementación de este filtro es menos destacada en la zona de los bordes. Este algoritmo asume que el ruido es multiplicativo con estadísticas estacionarias. La expresión utilizada en el cálculo del filtro es:

$$\hat{R}(x, y) = \sum_{i*j} m(x + i, y + j) I(x + i, y + j)$$

Donde i y j son los índices de la ventana de filtro y m es la función de ponderación:

$$m(x + i, y + j) = K_0^{-KC_I^2(t)\sqrt{i^2+j^2}}$$

Donde $t = (x, y)$, K_0 es la constante de normalización y K es el factor de amortiguamiento.

El factor de amortiguamiento K se elige de manera que en regiones homogéneas, $KC_I^2(t)$ tiende a 0. Así, el valor de $m(x + i, y + j)$ se aproxima a 1, lo que hace que el filtro actúe como un filtro de media; en las áreas en las que existen bordes, $KC_I^2(t)$ se hace tan grande que el valor de $m(x + i, y + j)$ aproxima a 0 para los píxeles circundantes (x, y) , y sigue siendo 1 para el

píxel (x,y) . Esto hace que el filtro se comporte como un filtro pasa todo preservando la imagen observada inicialmente. [40] [41] [42]

- **Filtro Gamma MAP:** El filtro de máximo a posteriori (MAP) es un filtro no lineal a diferencia de los filtros Lee, Kuan y Frost que son lineales. El filtro MAP intenta estimar el nivel digital original del píxel, que se asume debe estar entre la media local y el nivel digital del píxel degradado (actual). La lógica del MAP maximiza la función de densidad de probabilidad a posteriori respecto a la imagen original. El algoritmo para la aplicación de este filtro se basa en la estimación de una señal observada x maximizando el teorema de Bayes:

$$f(x|z) = \frac{f(x|y)f(x)}{f(z)}$$

Donde es la función $f(x|z)$ de densidad de probabilidad a posteriori, $f(x)$ es el modelo de señal original sin ruido, $f(x|z)$ es el término de probabilidad máxima y $f(x)$ es el modelo de los datos observados. Para utilizarlo, se necesita el conocimiento a priori de la función de densidad de probabilidad (PDF) de la imagen. [40] [41] [42]

2.4.1.2. Técnicas en el dominio wavelet

El uso de transformadas wavelet hace una mejora suave en la imagen lo cual permite la eliminación de speckle, esta mejora suaviza las áreas homogéneas de la imagen y preserva los bordes. Existen varios métodos basados en modelos estadísticos para reducir el ruido.

La transformada wavelet discreta (DWT) proporciona una matriz de coeficientes, conocidos como coeficientes wavelet, estos componen un conjunto de sub bandas en diferentes orientaciones y escalas de resolución. A medida que la transformada

wavelet es buena en la compactación de la energía, los pequeños coeficientes probablemente representan el ruido, y los grandes coeficientes representan importantes características de la imagen. Los coeficientes que representan características tienden a persistir a través de las escalas y formar grupos conectados espacialmente dentro de cada sub-banda. Estas propiedades hacen atractivo a DWT para la eliminación de ruido. Se han desarrollado un número de técnicas de eliminación de puntos basado en wavelets.

El procedimiento general es:

1. Calcular la transformada wavelet discreta.
2. Eliminar el ruido cambiando los coeficientes de onda.
3. Aplicar la transformada wavelet inversa (IDWT). Para construir la imagen sin ruido.

La técnica más utilizada es la contracción wavelet, esta se basa en la umbralización de los coeficientes wavelet, suprimiendo los coeficientes que representan ruido y reteniendo los coeficientes que probablemente representan características de la imagen. Por lo general se realiza con uno de los dos esquemas dominantes de fijación de umbrales: umbralización dura y umbralización suave. La imagen en el dominio wavelet se define como:

$$o = s + n$$

Donde o es el coeficiente wavelet observado, s es el componente libre de ruido y n es el ruido aditivo. El estimador de contracción wavelet se representa como

$$\hat{s} = H * o$$

Donde H indica el factor de contracción. Para las normas wavelet de umbralización clásica se define un valor de umbral T y H se especifica de la siguiente manera. Para umbralización dura:

$$H = \begin{cases} 0 & \text{si } |o| < T \\ 1 & \text{si } |o| \geq T \end{cases}$$

Para umbralización suave:

$$H = \begin{cases} 0 & \text{si } |o| < T \\ 1 - \frac{T}{|o|} & \text{si } |o| \geq T \end{cases}$$

La umbralización suave proporciona suavidad, mientras que la umbralización dura conserva las características de la imagen. El inconveniente de los métodos de fijación de umbrales es que la elección del umbral generalmente se realiza de una manera especializada.

Por otra parte, existe la contracción garrote realizada por Gao donde se eliminan los coeficientes pequeños y se minimizan los coeficientes grandes por medio de una función no lineal. El factor de contracción garrote H , se define como:

$$H = \begin{cases} 0 & \text{si } |o| < T \\ o \left(1 - \left(\frac{T}{o} \right)^2 \right) & \text{si } |o| \geq T \end{cases}$$

Por medio de simulaciones Gao comprobó que la contracción garrote presenta un ECM menor que la contracción suave y mayor suavidad que la contracción dura [43] [44].

En la tabla 4 se resumen algunas de las técnicas que se pueden utilizar en la etapa de pre procesamiento.

Tabla 4. Técnicas de pre procesamiento

MÉTODO	DESCRIPCIÓN
Técnicas de filtrado.	Se hace uso de una ventana móvil para la convolución del filtro con la imagen para reducir el speckle.
	VENTAJAS

	Fácil y rápida.
	DESVENTAJAS
	<ul style="list-style-type: none"> • En la representación a única escala es difícil discriminar la señal del ruido. • Sensible al tamaño y la forma de la ventana de filtro.
Técnicas en el dominio wavelet.	DESCRIPCIÓN
	Transformar la imagen al dominio wavelet y eliminar el ruido mediante la modificación de los coeficientes wavelet.
	VENTAJAS
	<ul style="list-style-type: none"> • En el dominio wavelet, las estadísticas de las señales se simplifican. • El ruido y la señal se procesan en diferentes escalas y orientaciones.
	DESVENTAJAS
	DWT y IDWT cálculos de complejidad aumentan el tiempo.

2.4.2. Segmentación de lesión

En esta etapa se divide la imagen en regiones no superpuestas y se separa la lesión del fondo, se selecciona la región de interés (ROI) que será destinada a la extracción de características. Existen diferentes orientaciones que permiten realizar la segmentación de imágenes, las más comunes son las técnicas clásicas, dentro de estas se encuentra la umbralización global, local o adaptativa, aunque la más utilizada es la basada en emparejamiento de patrones que trata de segmentar las lesiones buscando patrones que se asemejen a prototipos obtenidos previamente de una muestra de entrenamiento. La substracción de imágenes, es otro enfoque que se aplica a pacientes a los que se les realiza seguimiento ya que

permite la comparación de imágenes médicas mediante el alineamiento elástico, identificando las ROI y las técnicas multiescala, en donde se aplican diferentes escalas para identificar posibles variaciones en el tamaño de las lesiones. [31], [37]

2.4.2.1. Técnicas basadas en umbrales

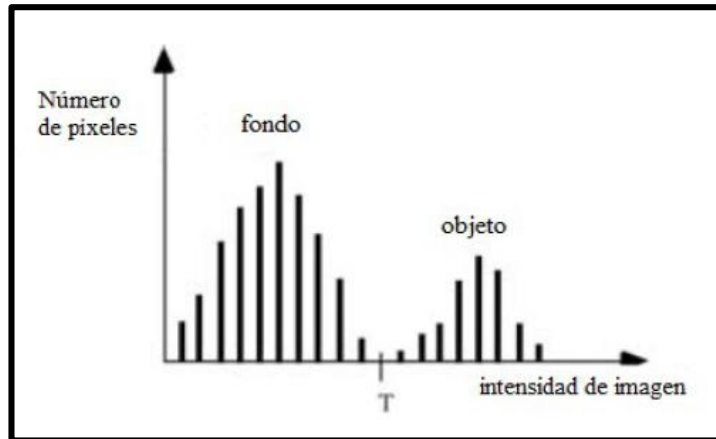
La umbralización es una técnica de segmentación simple y eficiente, el objetivo de esta técnica es separar los píxeles de la imagen en dos grupos por medio de la fijación de un umbral, de esta manera un grupo de píxeles tendrá valores inferiores al umbral mientras otro grupo tendrá valores superiores o iguales a este umbral. Estas técnicas se clasificadas en dos grupos:

- Umbral global: Para aplicar esta técnica la imagen debe tener un histograma bimodal (Figura 19), lo que permite separara los píxeles que componen el objeto de los píxeles del fondo, mediante un valor de umbral T . La imagen luego de la segmentación dada por el umbral se define como

$$B(x, y) = \begin{cases} t_0 & \text{si } I(x, y) < T \\ t_1 & \text{si } I(x, y) \geq T \end{cases}$$

Donde $B(x, y)$ es la imagen umbralizada, t_0 y t_1 son los valores que toma el pixel, si t_0 y t_1 toman los valores de 0 y 1 se dice que la imagen se ha binarizado e $I(x, y)$ es la imagen original.

Figura 19. Histograma bimodal [45]



- Umbral local o adaptativo: los umbrales locales son usados cuando la imagen no posee un histograma bimodal o no se obtiene una buena segmentación, por esto es necesario utilizar diferentes umbrales. Los métodos más comunes de umbralización local se basan en la división de la imagen y la evaluación de la intensidad de los píxeles, en el primer enfoque se obtienen histogramas de las sub imágenes, donde cada sub imagen contiene píxeles del objeto y el fondo, en el segundo enfoque se analizan las intensidades en la vecindad de cada pixel.

La desventaja de los umbrales locales frente a los umbrales globales es el costo computacional, sin embargo los umbrales locales funcionan mejor en la segmentación de imágenes con fondo variable [46], [45].

2.4.2.2. Técnicas basadas en detección de bordes

Estas técnicas se basan en la diferencia de intensidades entre la frontera del objeto y el fondo, la detección de bordes es de rápida ejecución y no requieren información anterior de la imagen, para aplicar esta técnica se debe realizar un

procesamiento a la imagen ya que por lo general los bordes no rodean completamente el objeto de interés lo cual causa un alto costo computacional.

El modelo de contorno activo más conocido como Snake, es un método de segmentación basado en la detección de bordes usado comúnmente en imágenes US. El modelo de snake modifica su forma y se aproxima al contorno deseado minimizando la energía asociada con el contorno actual como la suma de energías internas y externas. Durante el proceso de deformación, la fuerza se calcula a partir de la energía externa derivada de la energía característica de la imagen y de la energía interna derivada del modelo de contorno.

2.4.2.3. Técnicas basadas en regiones

La segmentación basada en regiones tiene en cuenta el criterio de similitud, este enfoque divide la imagen en regiones basándose en la similitud de la intensidad de los píxeles, suponiendo una intensidad uniforme. La más usada es la técnica de crecimiento de regiones la cual parte de un píxel o un grupo de píxeles ubicados en el área de interés. Luego, los píxeles en la vecindad son analizados uno a uno y agregados a la región en crecimiento cuando cumplen el criterio de homogeneidad, este puede ser la diferencia entre el valor de intensidad del píxel y la media del valor de intensidad dentro de la región, si el resultado de esta operación es menor a un valor predefinido el pixel se agrega a la región. El resultado de esta técnica es altamente dependiente del criterio de homogeneidad seleccionado, así como de los píxeles iniciales [39], [40], [45].

En la tabla 5 se resumen algunas de las técnicas que se pueden utilizar en la etapa de segmentación.

Tabla 5. Técnicas de segmentación

MÉTODO	DESCRIPCIÓN
Métodos de umbralización.	Se selecciona un valor de umbral para segmentar la imagen.
	VENTAJAS
	Simple y rápido.
	DESVENTAJAS
	No hay buenos resultados para las imágenes con histogramas no bimodales.
Modelo de contorno activo.	DESCRIPCIÓN
	Se utiliza el modo de deformación Snake.
	VENTAJAS
	Se puede extraer la lesión con diferente forma y mantener el límite correctamente.
	DESVENTAJAS
Velocidad de iteración lenta.	
Técnica de crecimiento de regiones.	DESCRIPCIÓN
	Se hace uso del criterio de homogeneidad para la clasificación de píxeles.
	VENTAJAS
	Preciso y exacto debido a la evaluación de todos los píxeles.
	DESVENTAJAS
Altamente dependiente del criterio de homogeneidad seleccionado y de los píxeles iniciales.	

2.4.3. Extracción de Características

En este paso se encuentra un conjunto de características de lesiones de mama ubicadas en la ROI para clasificar con precisión las imágenes y por lo tanto distinguir tejidos malignos y benignos. El rendimiento de CAD está asociado más a esta etapa que al proceso de segmentación [37], mejorando la detección de nódulos en el tejido estudiado. Las características de ecografía se pueden dividir en cuatro categorías [31]: características de textura, características lógicas o geométricas, características basadas en modelos y características descriptivas. La mayoría de las características descriptivas son usadas en el método de clasificación de lesiones de mama BI-RADS ecográfico descrito en la sección 2.3.1. En la tabla 6 se presentan diferentes características de cada una de las categorías.

2.4.3.1. Características de textura

Existen diferentes formas de obtener características de textura en imágenes digitalizadas, entre las más utilizadas se encuentran las matrices de concurrencia o también conocidas matrices de dependencia de niveles de gris [37] [40]. La textura en una imagen depende de variaciones de tonos, colores y forma de la superficie de la imagen, de esta manera la textura se representa en forma de sombras y regiones mejor iluminadas. Las características de textura ayudan en la identificación de objetos o regiones de interés. [47] [35]

2.4.3.2. Características morfológicas o geométricas

Se usa el área y el borde para calcular la forma de la lesión en la ROI. Ya que los tumores benignos generalmente tienen formas suaves, mientras que los malignos tienden a tener bordes irregulares [35], [48], [49].

2.4.3.3. Características basadas en modelos

Las características basadas en modelos se centran en el eco de retrodispersión del tejido mamario observado en las imágenes de US. Una vez el eco ha sido modelado y se encuentran los parámetros del modelo estos se utilizan como características para distinguir lesiones malignas y benignas. Los modelos más utilizados para el diagnóstico del cáncer de mama son la ley de potencia de ruido (PLSN), el modelo Nakagami, el modelo de distribución K y el modelo de espectro generalizado (GS).

Las ventajas de las características basadas en modelos son que no se ven influenciadas por la experiencia del radiólogo ni por las formas en que se obtienen las imágenes. La desventaja de las características basadas en modelos es que el fondo de los modelos es bastante complejo y la estimación de los parámetros es muy complicada [40].

2.4.3.4. Características del elastograma

Conjunto de características usadas en la categorización de las imágenes de elastografía a color para distinguir la elasticidad de los tejidos.

- Dureza media del contorno en modo-B: Esta es la dureza media calculada para el área encerrada por el contorno en modo-B superpuesto en la cara de la elastografía a color.
- Dureza media del contorno en el elastograma: Esta es la dureza media calculada para el área encerrada por el contorno de segmentación de lesión en el elastograma.
- Dureza media del área superpuesta: Esta es la dureza media del área de superposición tanto en la segmentación en modo-B y la segmentación en el elastograma

- Dureza media de la banda externa: esta es la dureza media del área en la banda externa de la lesión segmentada. La distancia usada para la banda exterior de la lesión tiene un límite de 5 píxeles.
- Dureza media de la banda interna: esta es la dureza media del área encerrada en la banda interna de la lesión segmentada. La banda usada tiene un límite de una distancia de 5 píxeles desde el borde de la lesión.
- Relación entre las diferentes durezas: se determinan umbrales para clasificar un píxel como duro, medio duro o blando en el elastograma. [31]

Tabla 6. Características empleadas en la clasificación de lesiones de mama en imágenes de US [40]

CATEGORÍA	CARACTERÍSTICA
Características descriptivas	Margen.
	Forma.
	Presencia de calcificaciones.
	Ecogenicidad.
	Acústica posterior.
	Extensión ductal.
	Orientación.
	Micro lobulaciones.
Características de textura	Coefficientes de auto covarianza.
	Diferencia de bloques de probabilidades inversas.
	Variación de bloques de los coeficientes de correlación locales.
	Varianza, auto correlación, o contraste promedio.
	Distorsión de distribución de coeficientes wavelet.
	Media y varianza de orden estadístico después de la

	descomposición wavelet.
	Contraste en los valores del nivel de gris.
	Correlación de la matriz de co-ocurrencia.
	Disimilitud.
	Frecuencia relativa de los elementos del borde.
	Comportamiento acústico posterior, diferencia mínima lateral o sombra acústica posterior.
	Homogeneidad de la lesión.
	Desviación estándar del valor de gris y su gradiente de la lesión.
	Características basadas en la matriz de dependencia espacial de tonos gris. (Correlación, energía, suma de entropía, diferencia de entropía, inercia y homogeneidad local).
	Características basadas en la matriz de dependencia de niveles de gris (contraste, media, entropía, diferencia inversa del momento y segundo momento angular).
	Dimensión fractal y características relacionadas.
Características morfológicas o geométricas	Espiculación.
	Relación entre la anchura y la profundidad.
	Patrón de ramificación.
	Numero de lobulaciones.
	Nitidez del margen.
	Ecogenicidad del margen.
	Variación angular en el margen.
	Numero de protuberancias y depresiones sustanciales.
	Índices de lobulaciones.
	Circunferencia elíptica normalizada.
	Esqueleto elíptico normalizado.

	Relación entre el eje mayor y el eje menor.
	Área de la lesión.
	Gradiente radial normalizado a lo largo del margen.
	Factor de forma.
	Convexidad.
	Circularidad del margen.
	Grado de interfaz abrupta tras el borde de la lesión.
	Característica angular.
Características basada en modelos	Características basadas en el modelo PLSM (ley de potencia de ruido de disparo).
	Características basadas en el modelo Nakagami.
	Características basadas en el modelo de espectro generalizado.
	Características basadas en el modelo de distribución k.
	Asimetría normalizada k.
	Fuerza del margen.
	Factor de speckle.
Características del elastograma.	Dureza media del contorno en modo B.
	Dureza media del contorno en elastografía.
	Dureza media del área superpuesta.
	Dureza media de la banda externa.
	Dureza media de la banda interna.
	Relación entre las diferentes durezas.

2.4.4. Clasificación

Después de obtener las características, estas se clasifican para determinar el grado de malignidad de la lesión o la patología de la misma. Existen varios tipos de clasificadores, se pueden usar individualmente o realizar una combinación que

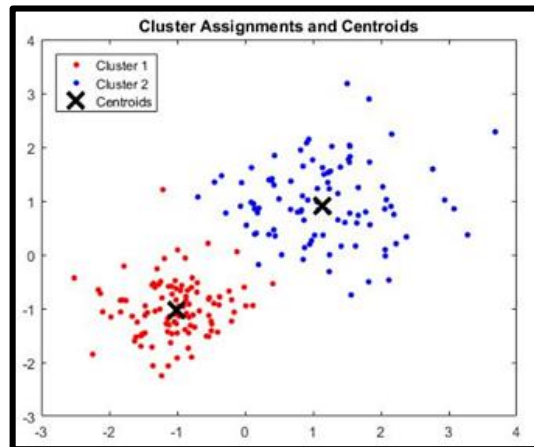
puede ser en paralelo, en cascada o de forma jerárquica para mejorar el rendimiento del proceso [31], [37], [40].

2.4.4.1. Clasificadores lineales

Los clasificadores lineales más utilizados para la detección y clasificación de lesiones de mama son el análisis discriminante lineal (LDA) y la regresión logística (REGLOG). La idea principal del LDA es encontrar la combinación lineal de las características que mejor separe dos o más clases de datos, en el LDA se asume que las variables independientes son continuas, con distribución normal y que la varianza en cada grupo es igual y REGLOG es un modelo que predice la probabilidad de un evento que ocurre como función de otros factores [40], [50].

Kmeans es un algoritmo de agrupamiento, en el que se crean conjuntos de datos similares entre sí, a cada conjunto de estos se le asigna un centroide, un punto se asigna a un grupo determinado si está más cerca al centroide de ese grupo que a cualquier otro. En la figura 20 se observa dos grupos de datos uno en rojo y otro en azul con su respectivos centroides [51].

Figura 20. Funcionamiento K-means [52]

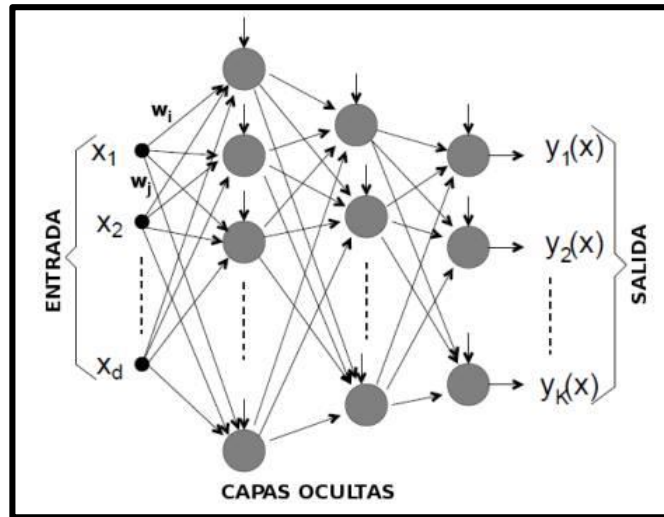


2.4.4.2. Redes neuronales artificiales

Las redes neuronales artificiales (RNA) imitan las propiedades del sistema nervioso biológico y las funciones de aprendizaje adaptativo biológico. Es un sistema de autoaprendizaje que cambia sus parámetros basados en la información externa o interna que fluye a través de la red durante la fase de aprendizaje. Una RNA se compone de una capa de entrada que es la capa que recibe la información proveniente de las fuentes externas, una capa de salida que transfiere la información obtenida y una o más capas ocultas que pueden estar interconectadas de distintas maneras, lo que determina, junto con su número, las distintas topologías de redes neuronales.

En la detección y clasificación del cáncer de mama, se utilizan con frecuencia tres tipos de redes neuronales artificiales: red neuronal Back-propagation, mapa de auto-organización (SOM) y redes neuronales jerárquicas [40] [47]. Sin embargo, la red neuronal más utilizada es la RNA feedforward que se caracteriza por ser un conjunto de neuronas que reciben información multivariable, la procesan y dan una respuesta que también puede ser multivariable. La topología y las interconexiones de la RNAf hacen fluir información de manera unidireccional como se puede observar en la figura 21. El funcionamiento de la red parte del ingreso de los datos a través de las neuronas de la capa de entrada, al pasar la información a la capa oculta, cada una de las neuronas receptoras recibe la suma ponderada de todas las entradas conectadas a ella, estas conexiones representa un peso [53].

Figura 21. Topología de una RNAf [53]



2.4.4.3. Árbol de decisión

Un árbol de decisión es un esquema de clasificación utilizado en el ámbito de la inteligencia artificial que está constituido por nodos internos que equivalen a un atributo o característica, un nodo de probabilidad que indica los posibles eventos que pueden ocurrir y las hojas que representan la decisión o la clase a la que pertenece. La manera como hace la clasificación un árbol de decisión es siguiendo las características más relevantes hasta llegar a una hoja que contiene la decisión final [40], [54], [55].

2.4.4.4. Máquinas de soporte vectorial

Las máquinas de soporte vectorial, son una técnica de aprendizaje automático, donde se busca obtener un hiperplano que permita separar dos o más clases de datos. Existen diferentes casos entre ellos se encuentran el caso linealmente separable donde existe un hiperplano que separa una clase de otra y el caso linealmente no separable donde se introduce una variable de holgura alrededor del hiperplano. También existen máquinas no lineales de soporte vectorial y máquina de soporte vectorial para la multclasificación [40], [56].

2.4.4.5. Lógica difusa

El uso de la lógica difusa en sistemas de reconocimiento de patrones y visión por computador ha aumentado ya que es bastante útil en problemas con alto grado de incertidumbre y donde se necesita usar el conocimiento de un experto que utiliza conceptos ambiguos o imprecisos [57]. La lógica difusa introduce una información que expresa el grado de pertenencia de una variable hacia un atributo. En los conjuntos difusos la función de pertenencia puede tomar valores entre 0 y 1, y la transición del valor entre 0 y 1 es gradual y no cambia de manera instantánea [38].

En la tabla 7 se resumen los clasificadores comúnmente utilizados para la clasificación de lesiones de mama.

Tabla 7. Clasificadores comúnmente utilizados en el diagnóstico de lesiones de mama

CLASIFICADOR	DESCRIPCIÓN
Clasificadores lineales.	Construyen límites de decisión mediante la optimización de criterios como LDA y LOGREG.
	VENTAJAS
	Sencillo y eficaz para datos separables linealmente.
	DESVENTAJAS
	Bajo rendimiento para datos no linealmente separables y adaptabilidad pobre para un problema complejo.
Redes neuronales artificiales (RNA).	DESCRIPCIÓN
	Construye funciones de mapeo no lineal: backpropagation, SOM y RNA jerárquica.
	VENTAJAS
	Robusta, no se necesita ninguna regla o expresión explícita, aprendizaje adaptativo y auto organización.
	DESVENTAJAS

	Largo tiempo de entrenamiento, complejidad de aprendizaje para grandes tareas, elevada cantidad de datos para el entrenamiento
Árbol de decisión	DESCRIPCIÓN
	Tiene una estructura de árbol con las reglas de clasificación en cada nodo.
	VENTAJAS
	Baja complejidad, trata problemas multi-clase y es de fácil interpretación.
	DESVENTAJAS
	La precisión depende totalmente del diseño del árbol y las características escogidas.
Máquina de soporte vectorial.	DESCRIPCIÓN
	Mapea los datos de entrada en un espacio de dimensión superior y busca un hiperplano óptimo para separar las muestras.
	VENTAJAS
	Proceso de entrenamiento más rápido que RNA, proceso de entrenamiento repetible y buen desempeño.
	DESVENTAJAS
	Aprendizaje supervisado (los datos de entrenamiento deben ser etiquetados), parámetros dependientes.
Lógica difusa	DESCRIPCIÓN
	Introduce una función que expresa el grado de pertenencia de una variable e entrada con relación a los conjuntos previamente definidos.
	VENTAJAS
	plantear el problema en los mismos términos en los

	que lo haría un experto humano, fácil implementación, buenos resultados en procesos no lineales
	DESVENTAJAS
	Múltiples definiciones de operadores y reglas de inferencia difusas y se debe tener un conocimiento previo para el planteamiento de reglas.

3. IMPLEMENTACIÓN DEL PROCESO CAD.

En este capítulo se presentan las herramientas software, así como las técnicas y procedimientos que se utilizaron para el desarrollo de la aplicación, teniendo en cuenta las etapas de un sistema CAD. En la etapa de pre procesamiento se presentarán los algoritmos utilizados para la mejora y eliminación de ruido en las imágenes de ecografía mamaria y de los elastogramas correspondientes. Posteriormente las técnicas de segmentación con las que se separa el fondo de la lesión permitiendo la extracción de diferentes características utilizadas en la clasificación de las lesiones.

Se cuenta con una base de datos de imágenes de ultrasonido de diferentes tipos de lesiones de mama descargada de [58], esta contiene un archivo tipo STRADWIN y STRADWIN DATA para cada imagen donde están contenidos los datos adquiridos en el US [59]. Estos archivos almacenan la imagen, la posición y el análisis de datos de una sola grabación hecha por el transductor, el archivo con extensión .sw contiene información textual y .sxi contiene datos binarios de la imagen.

3.1. HERRAMIENTAS SOFTWARE

3.1.1. QT 5.4

QT es un marco de desarrollo para escritorio y móviles, soporta plataformas como: Linux, OS X, Windows, VxWorks, QNX, Android, iOS, Blackberry, Sailfish OS entre otros.

QT es un marco escrito en C++ mas no un lenguaje de programación, su funcionamiento se basa en el procesador MOC (Meta Object Compiler), que se utiliza para extender el lenguaje C++ con funciones como señales y ranuras,

analiza los archivos de código fuente escrito en QT y genera un código fuente en C++ estándar compatible de esta manera con el propio marco, aplicaciones y bibliotecas utilizadas pueden ser compiladas por cualquier compilador de C++ estándar.

3.1.1.1. QT Designer

QT Designer permite el diseño y construcción de interfaces gráficas de usuario (GUI), utilizando la biblioteca de QT, permite desarrollar una interfaz de manera más sencilla y rápida sin tener que escribir en código fuente. El diseño de una interfaz en QT Designer es básico, pues este permite arrastrar los elementos gráficos a utilizar desde el menú a un formulario, tiene componentes como seleccionar, cortar, pegar, cambiar tamaños, entre otros. Este programa permite cambiar dinámicamente propiedades establecidas dentro del código y obtener un archivo con extensión (ui) con las características que contiene la interfaz [60].

3.1.2. OpenCv 2.4.6

OpenCv es una biblioteca abierta, bajo una licencia BSD es decir es libre para uso académico y comercial, tiene interfaces en C++, Python y Java, fue diseñado para la eficiencia computacional y para aplicaciones en tiempo real con el objetivo de suministrar un entorno de desarrollo fácil de utilizar y muy eficiente. Contiene múltiples funciones para áreas en el proceso de visión, reconocimiento de objetos, reconocimiento facial, visión estérea y visión robótica [61].

OpenCv permite realizar:

- Operaciones básicas.
- Procesado de imágenes y análisis.
- Análisis estructural.
- Análisis de movimiento.

- Reconocimiento de modelos.
- Reconstrucción 3D y calibración de cámaras.
- Interfaz gráfica.
- Entre otras.

3.1.3. Microsoft Visual Studio 2012 Professional

Microsoft Visual Studio es un entorno de desarrollo integrado, permite crear aplicaciones para Windows, Android e iOS. Soporta lenguajes tales como C++, Visual Basic, .NET y Java. Visual Studio permite crear aplicaciones cliente, aplicaciones web y servicios web en cualquier entorno soportado por la plataforma .NET. De esta manera se logra que las aplicaciones se intercomunicuen entre estaciones de trabajo, páginas web y dispositivos móviles [62].

3.1.4. Matlab Compiler Runtime 2013

MATLAB Runtime es un conjunto independiente de bibliotecas compartidas que permite la ejecución de aplicaciones o componentes de MATLAB compilados en ordenadores que no tienen instalado MATLAB. Cuando se utilizan conjuntamente, MATLAB, MATLAB Compiler y MATLAB Runtime le permiten crear y distribuir aplicaciones numéricas o componentes de software de forma rápida y segura [52].

3.1.5. CMake 3.3.2

CMake es un código abierto, diseñado para construir, probar y empaquetar software. Es utilizado para el proceso de compilación de software, así como los compiladores independientes de los archivos. CMake cubre la necesidad de un entorno de construcción de gran alcance en proyectos de código abierto.

Sus principales funcionalidades son: ficheros de configuración escritos en lenguaje de scripting específicos para CMake, análisis automáticos de dependencias para C, C++, Java entre otros, soporte para versiones de Microsoft Visual Studio, detección de cambios de ficheros, etc [63].

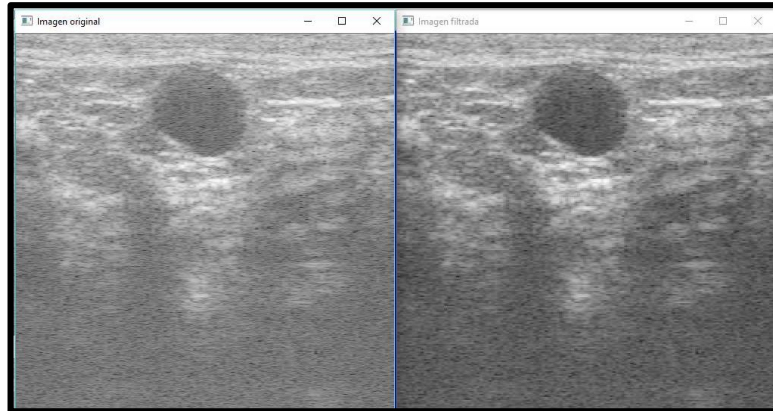
3.2. ETAPAS CAD

El desarrollo del proceso CAD se basa en las etapas descritas en [31], partiendo del software descargado desde [64]. A lo largo de las etapas CAD se hicieron modificaciones y se crearon nuevos algoritmos que permiten el cumplimiento de los objetivos, los métodos utilizados en cada etapa CAD se acogieron a partir de las ventajas y desventajas presentadas en las diferentes tablas de la sección 2.4. En el algoritmo de pre procesamiento se adicionó un filtro wavelet, a la segmentación no se le realizó ninguna mejora y en la extracción de características se crearon algoritmos para las características descriptivas, morfológicas y de textura, mientras que las características del elastograma se toman del software descargado, esto se desarrolló en OpenCV y los clasificadores en Matlab.

3.2.1. Pre procesamiento

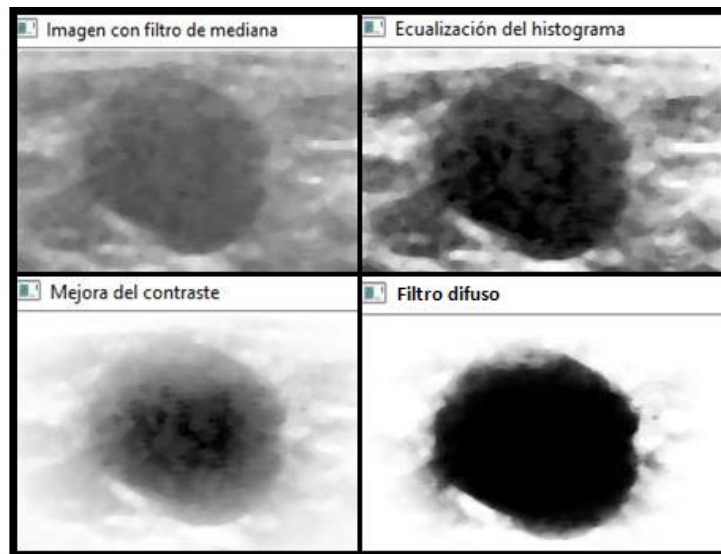
Una vez cargada la imagen de US a analizar se realiza un filtro wavelet a toda la imagen con el fin de eliminar el ruido multiplicativo, se utilizó un algoritmo con contracción garrote ya que de acuerdo a lo mencionado en la sección 2.4.1.2 ésta presenta un menor ECM y una mayor suavidad, el algoritmo utilizado se encuentra en el Anexo D. Los resultados obtenidos se pueden visualizar en la figura 22 donde en la parte izquierda se encuentra la imagen original y en la derecha la imagen filtrada, esta presenta mayor calidad lo cual es necesario para que el usuario realice una mejor selección de la ROI, donde se encuentra la lesión sospechosa.

Figura 22. Resultado filtro wavelet con contracción garrote.



La selección de la ROI se realiza de forma manual, esta puede hacerse en la imagen en modo B o en el elastograma con el fin de minimizar la carga computacional en las siguientes etapas de procesamiento. Para seguir mejorando la calidad de la ROI se aplica un filtro de mediana con un tamaño de 7×7 , se realiza la ecualización del histograma, la mejora del contraste y un filtro difuso siguiendo las tareas realizadas en la etapa de pre procesamiento descrita en [31], ver figura 23, código implementado en el software descargado.

Figura 23. Pre procesamiento ROI



En el elastograma (Figura 25) se aplica el mismo pre procesamiento desarrollado en la ROI de la imagen en modo B después de que esta se convierte del modelo de color RGB a HSV, del cual solo se toma el componente Hue (tono)(Figura 26), de donde se obtiene información de elasticidad por medio de la escala de tonalidad que se distribuye entre rojo, verde y azul representado tejidos blandos, moderados y duros respectivamente, ver figura 24.

Figura 24. Escala de tonalidad de elastografía



Figura 25. Elastograma

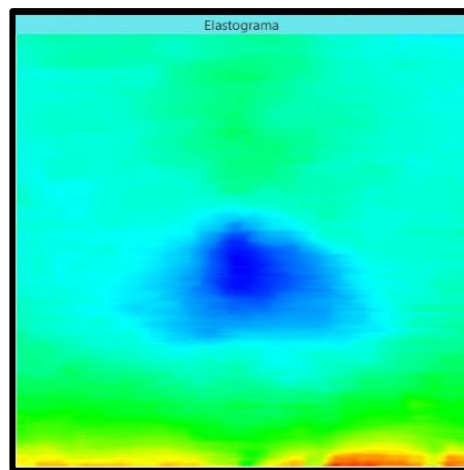
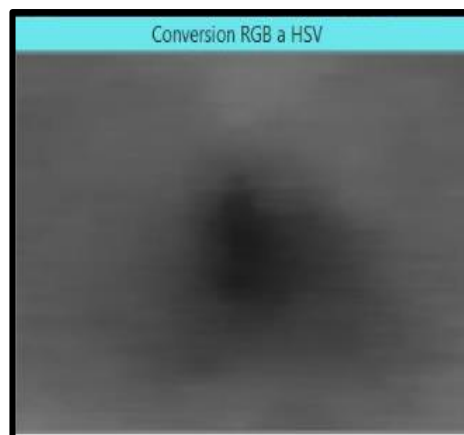


Figura 26. Componente Hue del elastograma

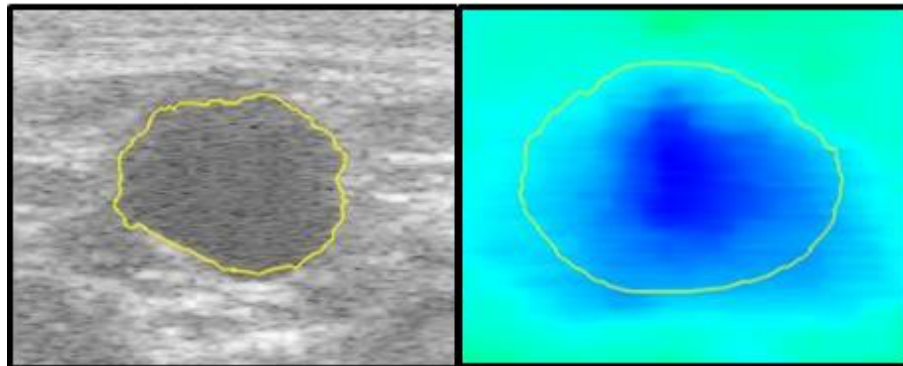


Después de la secuencia aplicada se logra destacar la lesión del fondo tanto en la imagen en modo B como en el elastograma, por lo que en la etapa de segmentación el objetivo es encontrar el contorno de la lesión.

3.2.2. Segmentación

Se hace uso de una técnica basada en umbralización para encontrar el contorno, esto es posible gracias a la etapa de pre procesamiento en la que se obtiene una imagen en donde se distingue la frontera entre el objeto y el fondo, para una mayor precisión se aplica un umbral de tal manera que la imagen quede binaria. Para encontrar el contorno de la lesión se utiliza el comando *findContours* de OpenCV con el que se obtiene el contorno de la Figura 27. Este proceso se tomó del software descargado.

Figura 27. Contorno de la lesión



3.2.3. Extracción de características

Una vez se cuenta con el contorno de la imagen se calculan las características que permiten hacer la clasificación de las lesiones, se escogieron algunas características descriptivas, morfológicas, de textura y del elastograma, descartando las características basadas en modelos debido a su complejidad en el cálculo.

3.2.3.1. Características descriptivas

Estas son una de las características más relevantes debido a que son las utilizadas en el método de clasificación BI-RADS, entre ellas se utilizan las características de forma, orientación y margen. Estas características se escogieron ya que son las más utilizadas en el diagnóstico de lesiones de mama.

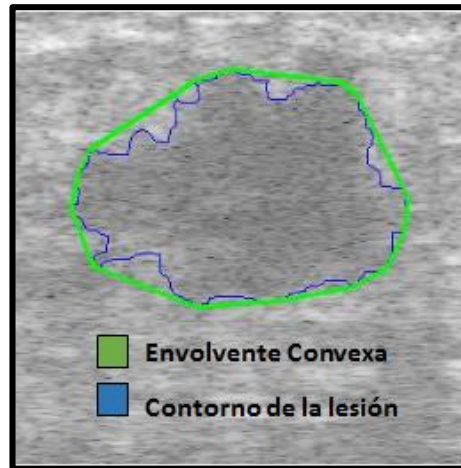
- Forma

Las lesiones benignas se caracterizan por tener una forma oval o redonda y las lesiones malignas tienen un borde irregular o lobulado. Para determinar la malignidad o benignidad de la lesión según la forma se encontró la envolvente convexa de la lesión por medio de la función *convexHull* de *OpenCV*. Una vez encontrada dicha envolvente, se calcula el área dentro del contorno de la lesión y dentro de la envolvente convexa (Anexo E, sección 1), se encuentra la relación Ra entre las áreas de la siguiente manera:

$$Ra = \frac{Al}{Ae}$$

Donde Al es el área dentro del contorno de la lesión y Ae es el área de la envolvente. En la figura 28 se observa el contorno de la lesión y su envolvente convexa. Si la lesión tiene forma redonda u ovalada las áreas tienden a ser iguales por lo que la relación toma un valor cercano a 1, de lo contrario si la lesión tiene bordes irregulares las áreas tienden a ser diferentes y el valor de la relación será cada vez menor.

Figura 28. Característica forma



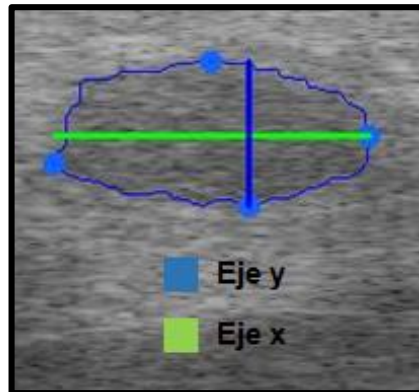
- Orientación

La orientación de la lesión se determina a partir de la relación de la longitud de sus ejes, mediante esta relación es posible saber si la lesión es paralela o anti paralela en relación al plano cutáneo (Anexo E, sección 2), las lesiones paralelas se consideran con un mayor grado de benignidad que las anti paralelas. La relación Re se calcula como:

$$Re = \frac{Ly}{Lx}$$

Donde Lx es la longitud del eje x de la lesión y Ly la longitud del eje y de la lesión, si el valor de Re es cercano a 0 la lesión es más ancha que alta, esto indica que la lesión es paralela, de lo contrario si el valor es mayor que 1 la lesión será anti paralela. En la figura 29 se puede observar los ejes x y y en una lesión.

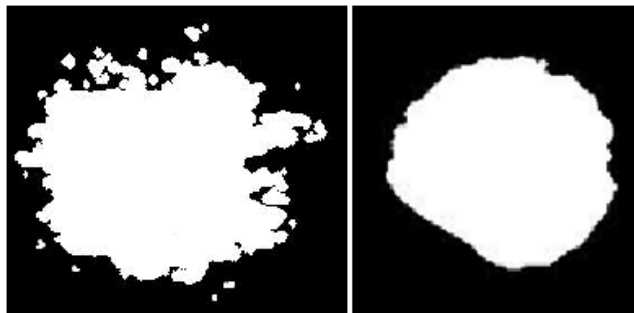
Figura 29. Característica Orientación



- Margen

La característica de margen describe la transición entre la lesión y el fondo de la ROI, entre mayor sea la distinción de esta frontera el nódulo tendrá márgenes más circunscritos siendo esta una característica de lesiones benignas. Para definir si una lesión tiene margen circunscrito o no circunscrito se hace uso de la ROI después de aplicarle el filtro difuso, esta imagen se convierte a blanco y negro mediante un umbral fijado, una vez realizada esta operación se hace un conteo de los objetos que se encuentran por fuera de la frontera de la lesión (Anexo E sección 3), en imágenes malignas con bordes poco definidos el número de objetos es mucho mayor que en imágenes benignas con bordes definidos. En la parte izquierda de la figura 30 se observa una lesión maligna y en la derecha una lesión benigna.

Figura 30. Característica margen



3.2.3.2. Características morfológicas

Estas características se basan en la forma de la lesión y son complementarias con las características descriptivas, lo cual permite realizar su cálculo de manera simultánea. El vector de características morfológicas está compuesto por: factor de forma, número de lobulaciones, área de la lesión, número sustancial de protuberancias y depresiones (NSPD), índice de lobulaciones y convexidad. El área de la lesión ya ha sido calculada en la característica de forma.

- Factor de forma

Esta característica es una relación entre el área y el perímetro de la lesión, sabiendo que las lesiones benignas tienden a ser redondeadas mientras las lesiones malignas tienen forma irregular. El factor de forma es calculado como:

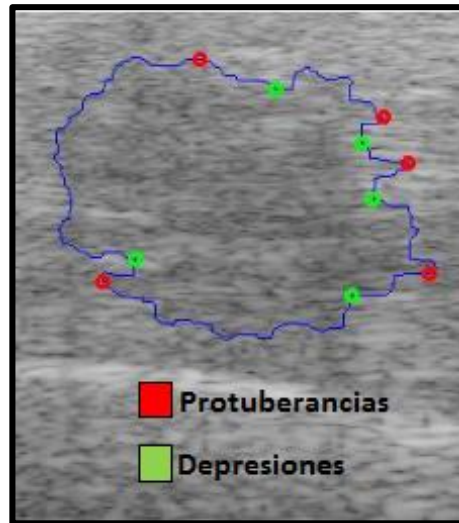
$$Factor\ de\ forma = \frac{4\pi * Area}{Perimetro^2}$$

Si la lesión fuera redonda el valor de *FF* sería 1, por lo que las lesiones con valores cercanos a este valor tienden a ser benignas [48]. Ver Anexo E sección 4.

- Número sustancial de protuberancias y depresiones (NSPD)

Haciendo uso de la envolvente convexa de la lesión encontrada, se procede a buscar los puntos en donde se presentan protuberancias y depresiones. Estos puntos se encuentran mediante la sentencia *convexityDefects* (Anexo E sección 5). NSPD es igual a la suma de las protuberancias y depresiones presentes en la lesión, entre mayor sea el valor de NSPD la lesión tendrá mayor grado de malignidad [40]. En la Figura 31 se muestra el número de protuberancias y depresiones encontradas en el contorno de una lesión.

Figura 31. Característica NSPD



- Número de lobulaciones

Las lobulaciones son salientes de la lesión, si el número de lobulaciones es mayor a 3 se considera la lesión como maligna. Para encontrar el número de lobulaciones se encontró la envolvente cóncava de la lesión, el número de lobulaciones es igual al número de objetos por fuera de dicha envolvente o al número de lados del polígono formado [40] (Anexo E, sección 6). En la Figura 32 se observa un polígono de seis lados que corresponde a la envolvente cóncava de una lesión.

Figura 32. Envolverte cóncava



- Índice de lobulaciones

Esta característica es complementaria de NSPD y del número de lobulaciones, el índice de lobulaciones IL se define como:

$$IL = \frac{AreaMax - Areamin}{\sum_1^n Ai}$$

Donde $AreaMax$ es el área de la lobulación más grande, $Areamin$ es el área de la lobulación más pequeña, n es el número total de lobulaciones y Ai es el área de cada lobulación [40].

- Convexidad

Una vez encontrada la envolvente convexa de la lesión se realiza una relación entre el perímetro de la envolvente y el perímetro de la lesión, esta relación será cercana a 1 para las lesiones benignas donde el perímetro de la lesión y el perímetro de la envolvente son semejantes [48].

$$Convexidad = \frac{Perimetro\ envolvente\ convexa}{Perimetro\ lesión}$$

3.2.3.3. Características de textura

Se realiza un análisis estadístico de cada una de las lesiones, para esto se encuentra la matriz de coocurrencia. La matriz de coocurrencia describe la frecuencia de un nivel de gris que aparece en una relación espacial específica con otro nivel de gris dentro del área de una ventada determinada, a partir de esta se extrae un conjunto de características propuestas por Haralick. Para el cálculo de la matriz de co-ocurrencia se sigue el procedimiento descrito en [65] - [66] teniendo en cuenta cuatro direcciones (0°, 45°, 90°, 135°) y una distancia entre píxeles de 5.

- Energía

La energía mide la uniformidad de la distribución de los niveles de gris en la imagen y el nivel de grosor de la textura, así como una idea de la suavidad de la textura. Si el valor es mayor, mayor es la uniformidad y menor es la variación en los niveles de gris.

$$Energía = \sum_{i=1}^N \sum_{j=1}^N p^2(i,j)$$

Donde $p(i,j)$ es el valor de la matriz de coocurrencia en la posición (i,j) .

- Entropía

La entropía mide la aleatoriedad de la distribución de los niveles de intensidad en la imagen.

$$Entropía = - \sum_{i=1}^N \sum_{j=1}^N p(i,j) \log[p(i,j)]$$

- Contraste

El contraste permite identificar las variaciones bruscas de niveles de gris en la imagen. Si la variación en los tonos de gris es mayor, mayor es el contraste y si el valor es 0 los niveles de gris son constantes en toda la imagen.

$$Contraste = \sum_{n=0}^{N-1} n^2 \left\{ \sum_{i=1}^N \sum_{j=1}^N p(i,j) \right\}$$

- Correlación

La correlación mide la probabilidad de relación entre las diferentes intensidades presentes en el entorno, de esta manera los pixeles más cercanos están más correlacionados que los pixeles lejanos. Si la correlación es 0 significa que no existe correlación entre los niveles de gris.

$$\text{Correlación} = \sum_{i=1}^N \sum_{j=1}^N p(i,j) \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]$$

Dónde:

$$\mu_i = \sum_{i=1}^N \sum_{j=1}^N i * p(i,j)$$

$$\mu_j = \sum_{i=1}^N \sum_{j=1}^N j * p(i,j)$$

$$\sigma_i = \sqrt{\sum_{i=1}^N \sum_{j=1}^N (i - \mu_i)^2 p(i,j)}$$

$$\sigma_j = \sqrt{\sum_{i=1}^N \sum_{j=1}^N (j - \mu_j)^2 p(i,j)}$$

- Homogeneidad

La homogeneidad es el cálculo de la regularidad de la textura en la imagen, si el contraste desciende la homogeneidad aumenta.

$$\text{Homogeneidad} = \sum_{i=1}^N \sum_{j=1}^N \frac{p(i,j)}{[1 + (i - j)^2]}$$

- Disimilaridad

Es una medida similar al contraste, su valor es alto si el contraste es alto.

$$\text{Disimilaridad} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i,j) |i - j|$$

- Varianza

Mide la variación de la intensidad en la textura alrededor de la media de la intensidad de la imagen procesada.

$$\text{Varianza} = \sum_{i=1}^N \sum_{j=1}^N (i - m)^2 p(i, j)$$

m Es la media de $p(i, j)$.

- Suma promedio

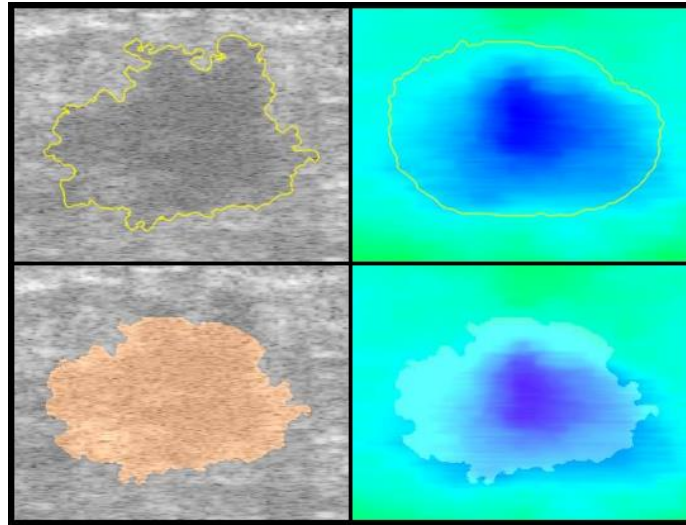
Es la medida promedio de niveles de gris, describe el brillo y la profundidad de sombra de la imagen. [67] [68]

$$\text{Suma promedio} = \sum_{k=2}^{2N} k \sum_{i=1}^N \sum_{j=1}^N p(i, j)$$

3.2.3.4. Características del elastograma

El cálculo de las características del elastograma se realiza en el software descargado, el objetivo es encontrar la dureza media de la lesión. El software descargado calcula las siguientes características: dureza media del contorno en modo B, dureza media del contorno en el elastograma, dureza media de la banda externa, dureza media de la banda interna y dureza media del área superpuesta, para encontrar el área superpuesta se realiza la intersección del contorno en modo b y el contorno en el elastograma como se observa en la Figura 33. La descripción de las características del elastograma se presenta en la sección 2.4.3.4.

Figura 33. Intersección de contornos



El cálculo de las características se hace por medio de la media de la dureza que se define como:

$$media = \frac{\sum_{i=0}^{n-1} hi}{n}$$

Donde hi es el valor del tono del pixel y n es el valor total de pixeles en cada una de las áreas.

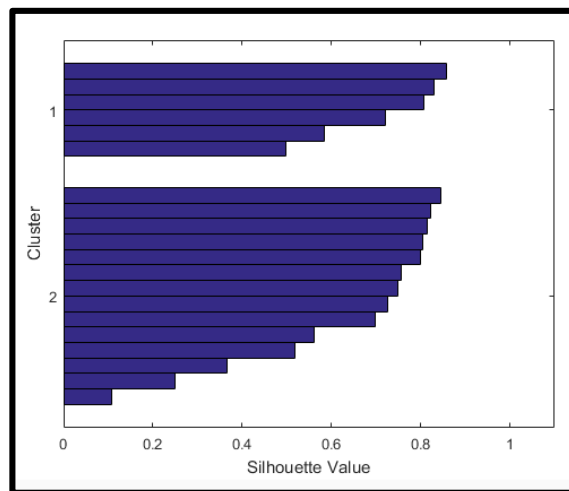
Finalmente se obtuvo un vector de 22 características el cual se subdivide en tres grupos, el primer grupo está conformado por las características descriptivas y morfológicas con un total de 9, el segundo grupo por las 5 características del elastograma y por ultimo 8 características de textura.

3.2.4. Clasificación

Después de calcular las características se procede a clasificar las lesiones, para esto se hace una optimización del vector de características con el fin de eliminar características redundantes e irrelevantes y de esta manera lograr un mejor diagnóstico.

Se hizo un procedimiento sistemático para una primera aproximación del vector óptimo de características, a partir del comando *silhouette* de Matlab el cual permite visualizar la separabilidad de las clases y la pertenencia de cada muestra evaluada a su respectiva clase, esta pertenencia tiene un valor comprendido entre -1 y 1 (Silhouette Value), un valor alto indica que la muestra está bien adaptada a su propia clase (Figura 34). De forma manual se realizaron combinaciones de las características pertenecientes a cada grupo hasta encontrar la combinación con mejor separabilidad y valores de pertenencia altos.

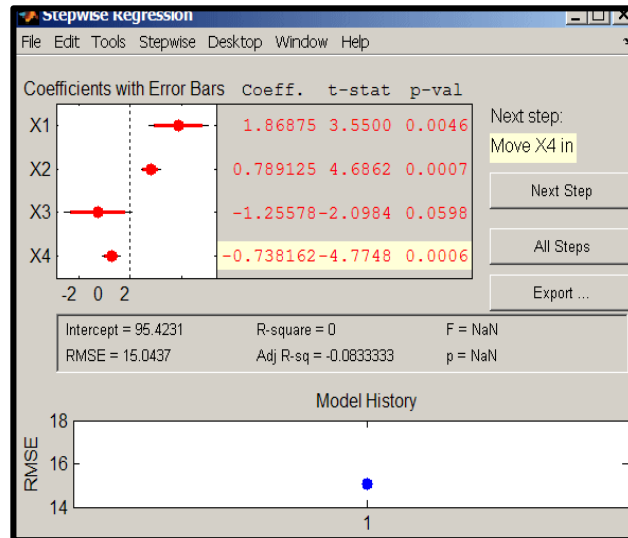
Figura 34. Ejemplo gráfica Silhouette [52]



En el proceso de lograr un mejor vector de características o rectificarlo se utilizó un modelo de regresión lineal, el cual estudia la relación entre variables, una variable dependiente y una o más independientes $Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n$. De acuerdo con la ecuación la variable dependiente Y se interpreta como una ecuación lineal de un conjunto de variables independientes X , cada una de las cuales va acompañada de un coeficiente B que indica el peso relativo de la variable en la ecuación, se calcula el error de la raíz cuadrada de la media (RMSE) entre valores observados y pronosticados. Para esto se utilizó el comando *stepwise* de Matlab el cual muestra la estimación de los parámetros que no

pertenece al modelo en rojo y los parámetros que pertenecen al modelo en un fondo a color y las barras de los coeficientes más cercanas a la línea punteada, en la parte inferior Model History se visualiza el RMSE que presenta cada una de las combinaciones (Figura 35).

Figura 35. Ejemplo gráfico Stepwise [52]



En la clasificación se utilizaron algoritmos de Matlab lineales como Kmeans. Kmeans es un clasificador no supervisado. Entre la clasificación no lineal se utilizó redes neuronales y árboles de decisión que son clasificadores supervisados. La red neuronal es de tipo feedforward, en Matlab se utiliza el comando *feedforwardnet*. El árbol de decisión se hizo con el comando *fitctree*, este utiliza una clasificación binaria que acepta una clasificación multiclases, se basa en las variables de entrada o atributos, para entregar una respuesta. Se escogieron clasificadores lineales y no lineales con el fin de realizar una comparación entre estos.

Adicionalmente se utiliza la clasificación UENO del software descargado, esta clasificación se realiza a partir de la característica de dureza media del contorno en modo B en la ROI.

En la tabla 8 se presenta un resumen de lo utilizado en las etapas del proceso CAD, en la columna Referencia se dan a conocer los artículos que presentan la teoría de cada uno de los algoritmos, en la columna Código se referencian las fuentes de donde se extrajeron los códigos, cuando se menciona que es de desarrollo propio es porque el código se construyó utilizando diferentes funciones de OpenCV o Matlab.

Tabla 8. Resumen implementación proceso CAD

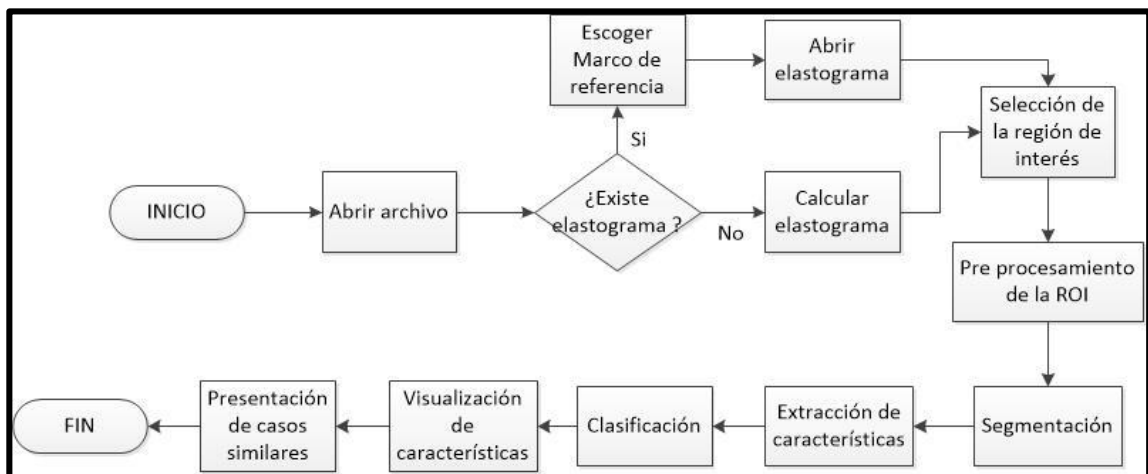
NOMBRE	REFERENCIA	CÓDIGO	HERRAMIENTA
ETAPA : PRE PROCESAMIENTO			
Filtrado Wavelet	Fuente [43] [44]	Anexo D	OpenCV
Filtro de mediana	Fuente [31]	Software [64]	OpenCV
Ecuilización del histograma	Fuente [31]	Software [64]	OpenCV
Mejora del contraste	Fuente [31]	Software [64]	OpenCV
Filtro difuso	Fuente [31]	Software [64]	OpenCV
ETAPA: SEGMENTACIÓN			
Técnica de umbralización	Fuente [31]	Software [64]	OpenCV
Contorno	Fuente [31]	Software [64]	OpenCV
ETAPA: EXTRACCIÓN DE CARACTERÍSTICAS			
Características descriptivas y morfológicas			
Forma	Fuente propia	Desarrollo propio	OpenCV
Orientación	Fuente propia	Desarrollo propio	OpenCV
Margen	Fuente propia	Desarrollo propio	OpenCV
Factor de forma	Fuente [48]	Desarrollo propio	OpenCV
NSPD	Fuente [40]	Desarrollo propio	OpenCV
Numero de lobulaciones	Fuente [40]	Desarrollo propio	OpenCV
Convexidad	Fuente [48]	Desarrollo propio	OpenCV
Índice de lobulaciones	Fuente [40]	Desarrollo propio	OpenCV
Área de la lesión	Fuente [40]	Desarrollo propio	OpenCV
Características de textura			
Energía	Fuente [67] [68]	Desarrollo propio	OpenCV
Entropía	Fuente [67] [68]	Desarrollo propio	OpenCV
Contraste	Fuente [67] [68]	Desarrollo propio	OpenCV
Correlación	Fuente [67] [68]	Desarrollo propio	OpenCV
Homogeneidad	Fuente [67] [68]	Desarrollo propio	OpenCV

Disimilaridad	Fuente [67] [68]	Desarrollo propio	OpenCV
Varianza	Fuente [67] [68]	Desarrollo propio	OpenCV
Suma promedio	Fuente [67] [68]	Desarrollo propio	OpenCV
Características del elastograma			
Dureza media del contorno en modo B	Fuente [31]	Software [64]	OpenCV
Dureza media del contorno en el elastograma	Fuente [31]	Software [64]	OpenCV
Dureza media del área superpuesta	Fuente [31]	Software [64]	OpenCV
Dureza media de la banda interna	Fuente [31]	Software [64]	OpenCV
Dureza media de la banda externa	Fuente [31]	Software [64]	OpenCV
ETAPA: CLASIFICACIÓN			
Silhouette	Fuente [52]	Desarrollo propio	Matlab
Stepwise	Fuente [52]	Desarrollo propio	Matlab
K means	Fuente [52]	Desarrollo propio	Matlab
Redes neuronales	Fuente [52]	Desarrollo propio	Matlab
Arboles de decisión	Fuente [52]	Desarrollo propio	Matlab

4. DESARROLLO DE LA APLICACIÓN

El principal objetivo de la aplicación software es el apoyo al diagnóstico de cáncer de mama permitiendo la comparación de casos similares con el caso a diagnosticar, para dar cumplimiento a este objetivo se elaboró un diagrama de flujo (Figura 36) que representa el funcionamiento de la aplicación.

Figura 36. Diagrama de flujo del funcionamiento de la aplicación



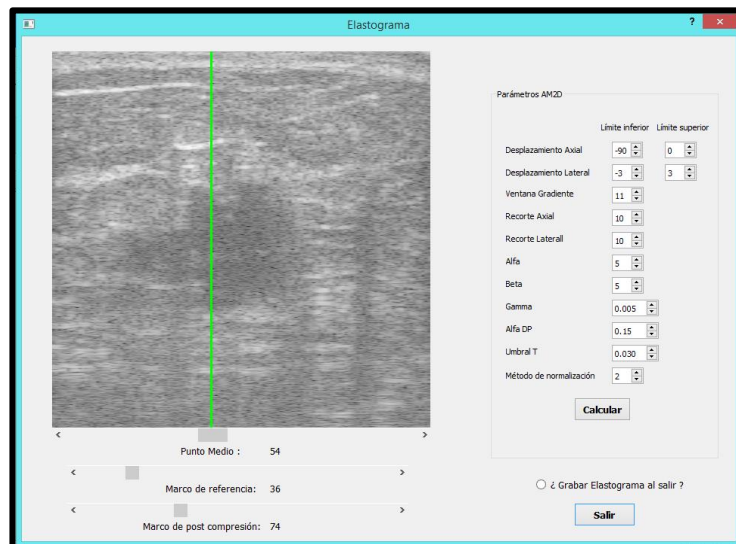
El primer paso es cargar los archivos .sw y .sxi de la lesión a diagnosticar, para esto se hace uso de una dll generada en Matlab la cual permite visualizar la ecografía en una escala de grises. La dll obtiene un marco de datos de radio frecuencia identificado por un número, el archivo sw contiene más de 100 marcos correspondientes a ecos obtenidos mientras se hace presión manual con la sonda en la superficie de la lesión. Para realizar este proceso se selecciona el botón **Abrir archivo** ubicado en la pantalla inicial ver Figura 37.

Figura 37. Pantalla Inicial



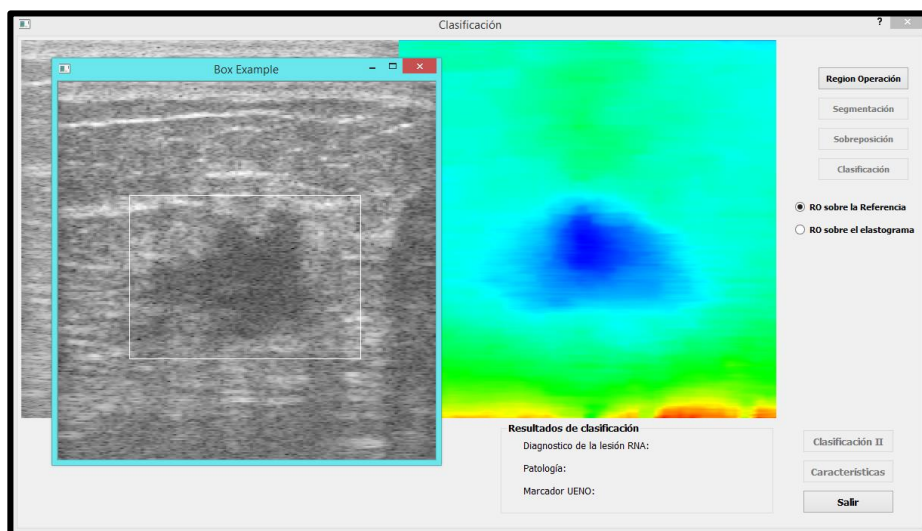
Si se cuenta con el elastograma de la lesión este se carga oprimiendo el botón **Abrir elastograma**, de lo contrario se debe calcular el elastograma (ver Figura 38) para lo cual se hace uso de la dll generada en Matlab, esta hace un filtro de mediana sobre la envolvente de los ecos de pre y pos compresión y luego usa el algoritmo AM2D para el cálculo del desplazamiento y el gradiente espacial. Para calcular el elastograma se utiliza el método AM2D utilizado en [69] y descargado de [70].

Figura 38. Cálculo del elastograma



Una vez obtenido el elastograma se procede a seleccionar la ROI, como se mencionó anteriormente esto se puede realizar en la imagen en modo B o en el elastograma, para esto se accede a la pantalla **Clasificación** por medio del botón **Calcular Segmentación**. La selección de la ROI (ver Figura 39) se hace fijando el apuntador sobre un punto y desplazándose hasta cubrir el área de interés que contiene la lesión.

Figura 39. Selección de la ROI



Después de seleccionada la ROI se hace el pre procesamiento y la segmentación de la lesión en la ecografía, para realizar este mismo procedimiento en el elastograma se pulsa el botón **Segmentación**. El pre procesamiento y la segmentación se realizan según lo presentado en la sección 3.2.

El botón **Sobreposición** permite sobreponer los dos contornos encontrados en la segmentación para realizar la extracción de características, el cálculo del vector de características y su posterior clasificación se desarrolla al oprimir el botón **Clasificación**. Cabe aclarar que se realizan dos clasificaciones esto debido a la información suministrada en la base de datos utilizada, en esta base de datos se realizaron dos diagnósticos para cada caso de estudio el primero divide las

lesiones en tres categorías: benignas, sospechosas y malignas y el segundo de acuerdo a la patología de la lesión encontrada en la biopsia: quistes, fibroadenomas y carcinomas invasivos. La base de datos consta de 63 imágenes de estas 24 son benignas, 14 sospechosas y 25 malignas de acuerdo a la primera clasificación, para la segunda clasificación se utilizaron solo 28 muestras divididas en 3 grupos: fibroadenomas, quistes y carcinomas invasivos ya que estas patologías son las que cuentan con mayor número de casos, las patologías sobrantes cuentan con máximo 4 casos lo cual resulta insuficiente para el entrenamiento y prueba de los diferentes clasificadores.

El primer diagnóstico se realiza al pulsar el botón **Clasificación** y el segundo al pulsar el botón **Clasificación II**, los resultados se visualizan en el recuadro Resultados de clasificación, además en este recuadro se puede visualizar la aproximación de la clasificación UENO (ver Figura 40).

Figura 40. Resultados clasificación

Resultados de clasificación	
Diagnostico de la lesión RNA:	Maligno
Patología:	C. Invasivo
Marcador UENO:	4

Para permitir al usuario comparar el caso diagnosticado con casos similares se despliegan las características obtenidas desde el botón **Características**, en la pantalla desplegada se encuentran los botones **Casos Similares I** y **Casos Similares II** (Figura 41), estos corresponden a los dos tipos de clasificaciones. Al oprimir cualquiera de los dos botones se muestran las carpetas que contiene la información de casos con el mismo diagnóstico, al elegir una de estas carpetas se puede observar la imagen de ecografía y las características, con lo que el usuario

puede realizar una comparación visual entre distintos casos y el caso diagnosticado, ver Figura 42.

Figura 41. Visualización de características

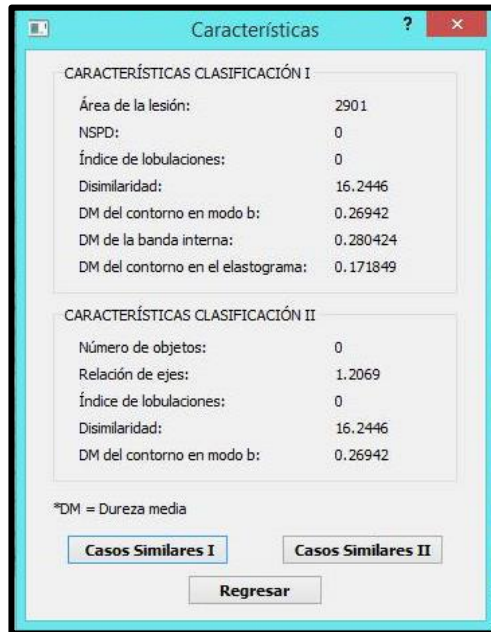
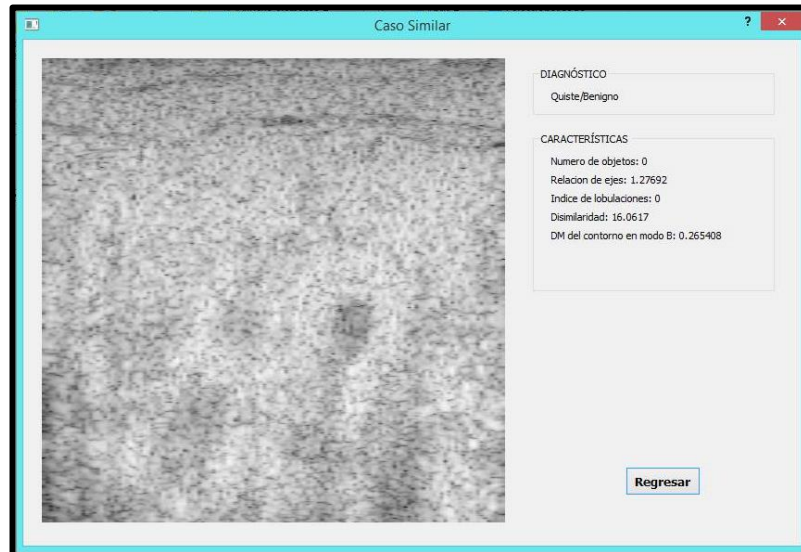


Figura 42. Características caso similar



En el Anexo G se presenta el manual de usuario de la aplicación el cual permite un mayor entendimiento del funcionamiento de la misma.

5. RESULTADOS.

Para la obtención de los resultados se encontraron las 22 características de las 63 muestras que conforman la base de datos y se planteó el siguiente plan de pruebas desarrollado en Matlab y comprobado en la aplicación:

- FASE 1: Prueba de los clasificadores Kmeans y red neuronal utilizando el total de las características extraídas en la clasificación I (lesiones benignas, sospechosas y malignas).
- FASE 2: Optimización del vector de características mediante la herramienta *silhouette* y *stepwise*.
- FASE 3: Prueba de los clasificadores Kmeans y red neuronal utilizando los vectores optimizados en la clasificación I (lesiones benignas, sospechosas y malignas).
- FASE 4: Comparación de clasificadores supervisados no lineales redes neuronales artificiales y árboles de decisión.
- FASE 5: Optimización del vector de características mediante *stepwise* y pruebas de los clasificadores redes neuronales y árboles de decisión para la clasificación II (quistes, fibroadenomas y carcinomas invasivos).
- FASE 6: Aporte de las características de elastografía para la primera (lesiones benignas, sospechosas y malignas) y segunda clasificación (quistes, fibroadenomas y carcinomas invasivos).

5.1. FASE 1

Se organizó una matriz de 63 filas y 22 columnas para la primera clasificación (Benigna, sospechosa y maligna), cada fila es un caso a evaluar y las columnas representan las características. Las primeras 24 filas corresponden a las muestras

de lesiones benignas, las 14 siguientes lesiones sospechosas y los 25 restantes lesiones malignas.

5.1.1. Kmeans

Kmeans arroja una etiqueta (Idx) para cada uno de las muestras fijadas en la entrada (característica), en este caso entrega una de las 3 clases que se le asignó a cada muestra. Ver Figura 43.

Figura 43. Comando Kmeans en Matlab

```

% Implementación Kmeans
Idx = kmeans(caracteristica,3);
    
```

En la primera clasificación se hace uso de la mitad de los datos para el entrenamiento y se obtuvieron los siguientes resultados (Tabla 9):

Tabla 9. Resultados entrenamiento Kmeans

Muestra	Clase Kmeans	Diagnóstico real	Muestra	Clase Kmeans	Diagnóstico real
1	2	Benigno	17	1	Sospechoso
2	2	Benigno	18	1	Sospechoso
3	2	Benigno	19	1	Sospechoso
4	3	Benigno	20	3	Maligno
5	3	Benigno	21	3	Maligno
6	1	Benigno	22	1	Maligno
7	3	Benigno	23	1	Maligno
8	2	Benigno	24	2	Maligno
9	2	Benigno	25	3	Maligno
10	3	Benigno	26	1	Maligno
11	3	Benigno	27	3	Maligno
12	3	Benigno	28	3	Maligno
13	1	Sospechoso	29	3	Maligno

14	1	Sospechoso	30	1	Maligno
15	1	Sospechoso	31	1	Maligno
16	1	Sospechoso	32	3	Maligno

El algoritmo Kmeans no logra separar las clases satisfactoriamente ya que el etiquetado no concuerda con el diagnostico real, solo las lesiones sospechosas las clasifica con una sola etiqueta, sin embargo, esta etiqueta también está asignada a lesiones malignas en gran cantidad. Las lesiones malignas y benignas no cuentan con una etiqueta predominante, por estas razones no se realizan pruebas de clasificación con las 31 muestras restantes.

5.1.2. Red neuronal

Las redes neuronales artificiales son clasificadores supervisados por lo que necesitan de un vector que defina la clase para cada una de las muestras (T y T1), se utilizaron dos redes neuronales feedforward donde los mejores resultados se obtuvieron con dos capas ocultas. Para iniciar el proceso de entrenamiento y prueba se dividió las muestras en dos grupos uno de 31 (caracteristicas1) y otro de 32 (caracteristicas2), cada grupo es utilizado para entrenamiento y prueba. Ver figura 44.

Figura 44. Entrenamiento de la red neuronal Matlab

```

% Entrenamiento de las redes neuronales
net = feedforwardnet(2);
net1 = train(net,caracteristicas1',T);
net2 = train(net,caracteristicas2',T1);

```

Se utilizaron tres clases, lesiones benignas etiquetadas con 10, lesiones sospechosas etiquetadas con 20 y lesiones malignas etiquetadas con 30. La salida de la red neuronal proporciona valores próximos a las etiquetas, para una mejor visualización y análisis de los resultados se hizo una aproximación

obteniendo valores enteros. Los resultados obtenidos para los dos grupos se muestran a continuación (Tabla 10 y Tabla 11):

Tabla 10. Clasificación RNA grupo 1

Muestra	Clase RNA	Diagnóstico Real	Muestra	Clase RNA	Diagnóstico Real
1	10	Benigno	17	20	Sospechoso
2	10	Benigno	18	20	Sospechoso
3	10	Benigno	19	20	Sospechoso
4	10	Benigno	20	30	Maligno
5	10	Benigno	21	30	Maligno
6	10	Benigno	22	30	Maligno
7	10	Benigno	23	20	Maligno
8	10	Benigno	24	30	Maligno
9	10	Benigno	25	20	Maligno
10	10	Benigno	26	20	Maligno
11	10	Benigno	27	20	Maligno
12	10	Benigno	28	30	Maligno
13	30	Sospechoso	29	30	Maligno
14	20	Sospechoso	30	20	Maligno
15	20	Sospechoso	31	20	Maligno
16	30	Sospechoso			

Tabla 11. Clasificación RNA grupo 2

Muestra	Clase RNA	Diagnostico Real	Muestra	Clase RNA	Diagnostico Real
1	10	Benigno	17	30	Sospechoso
2	10	Benigno	18	30	Sospechoso
3	10	Benigno	19	30	Sospechoso
4	10	Benigno	20	30	Maligno
5	10	Benigno	21	10	Maligno
6	10	Benigno	22	30	Maligno
7	10	Benigno	23	30	Maligno
8	10	Benigno	24	30	Maligno
9	30	Benigno	25	30	Maligno
10	20	Benigno	26	30	Maligno

11	10	Benigno	27	30	Maligno
12	10	Benigno	28	30	Maligno
13	30	Sospechoso	29	30	Maligno
14	30	Sospechoso	30	30	Maligno
15	10	Sospechoso	31	30	Maligno
16	20	Sospechoso	32	30	Maligno

De las 63 muestras evaluadas por las RNAs, 46 fueron bien clasificadas, mientras que las 17 muestras restantes indicadas en rojo fueron erróneas, esto da un porcentaje de aciertos de aproximadamente 73%. La mayoría de desaciertos ocurren entre lesiones sospechosas y malignas, con el fin de mejorar esta clasificación se procedió a optimizar el vector de características. El porcentaje de aciertos se calcula mediante la siguiente fórmula:

$$\text{porcentaje de aciertos} = 100 - \frac{\text{número de muestras mal clasificadas} * 100}{\text{número total de muestras}}$$

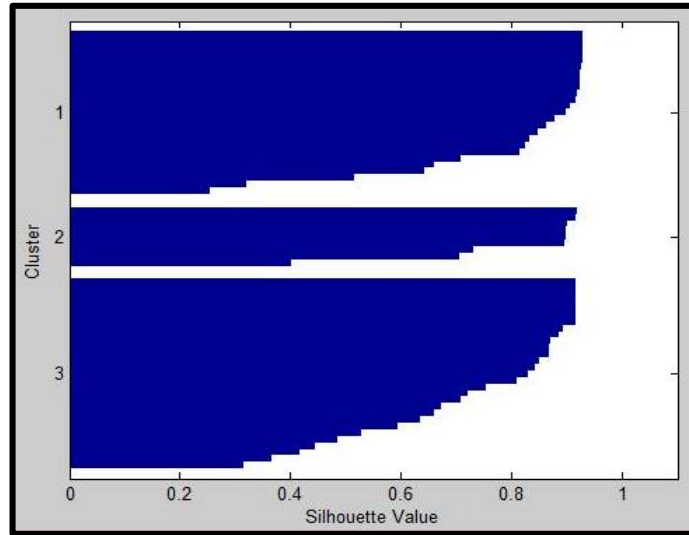
5.2. FASE 2

5.2.1. Silhouette

Después de realizar todas las posibles combinaciones entre los grupos de características se obtuvo una reducción del número de características desde 22 en la Fase 1 a un conjunto de 11 características, en las siguientes gráficas se observa que la mayoría de muestras tienen un valor de silhouette superior a 0,6 y las clases son separables.

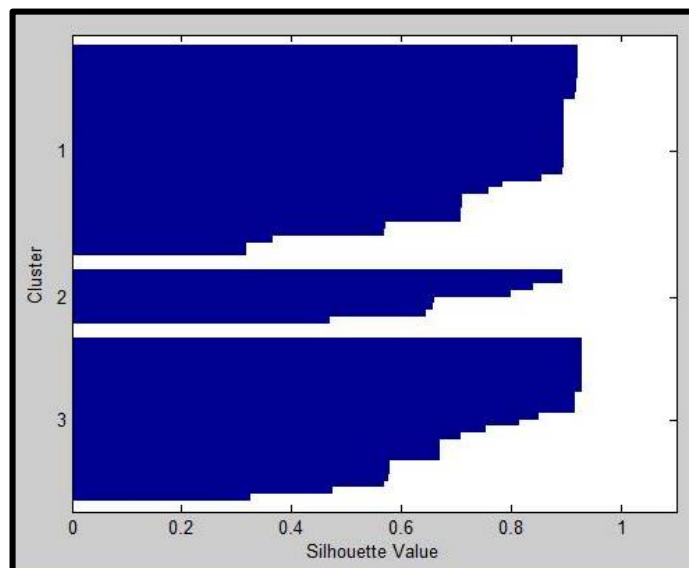
- Características descriptivas y morfológicas: La combinación de características de orientación, número de lobulaciones, área de la lesión y NSPD obtuvo el mejor resultado ver Figura 45.

Figura 45. Resultados optimización de características descriptivas y morfológicas



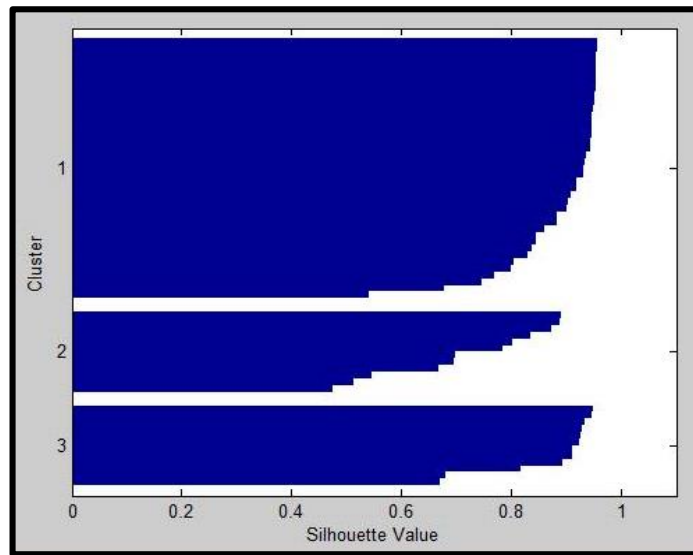
- Características de textura: La combinación de las características de disimilaridad, varianza y suma promedio obtuvo el mejor resultado. Ver Figura 46.

Figura 46. Resultados optimización de características de textura



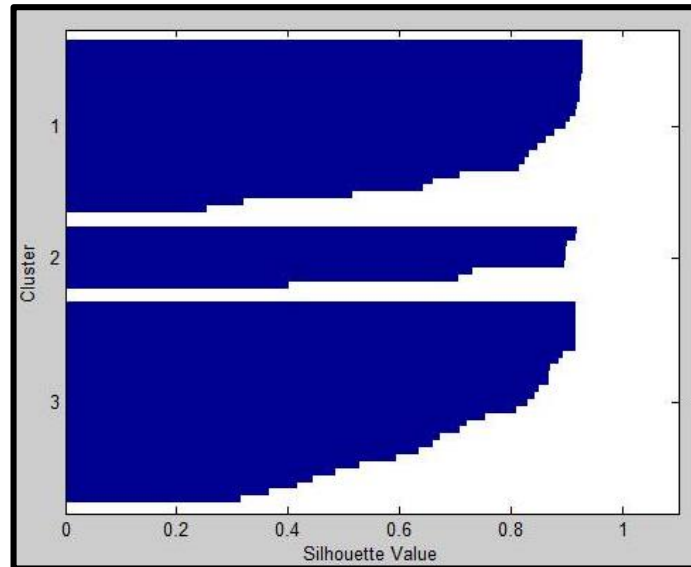
- Características del elastograma: La combinación de las características de dureza media del contorno en modo B, dureza media de la banda externa, dureza media del área superpuesta, dureza media del contorno en el elastograma obtuvo el mejor resultado, ver Figura 47.

Figura 47. Resultados optimización de características del elastograma



En la Figura 48 se observa el vector de 11 características conformado por las características mencionadas anteriormente.

Figura 48. Resultados optimización de características silhouette



5.2.2. Stepwise

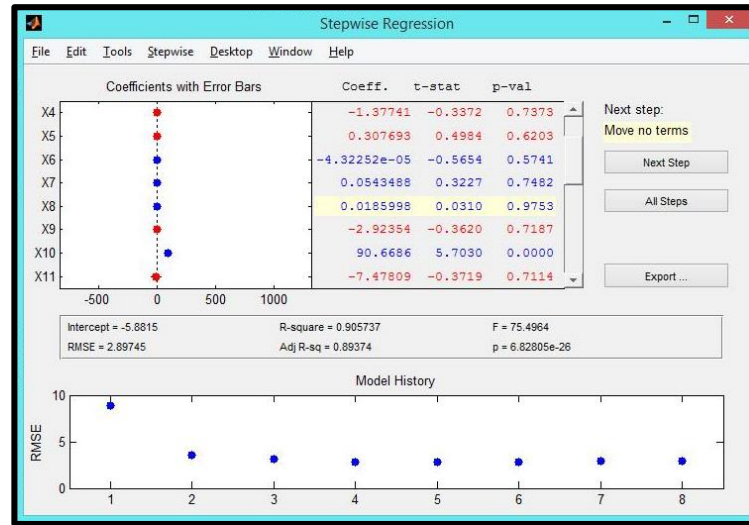
Para implementar el comando *Stepwise* de Matlab (Figura 49) se necesita de la matriz de características de 63 muestras (característica) y un vector que contiene la clase de cada muestra (T).

Figura 49. Comando Stepwise

```
stepwise(caracteristica,T)
```

Al utilizar este comando se despliega una pantalla donde se observan todas las características, esta herramienta selecciona las características que generan un modelo con menor RMSE y las colorea en azul, en el recuadro inferior (Model History) se observa la disminución del RMSE a medida que se agregan características al modelo. Ver figura 50.

Figura 50. Resultados optimización de características Stepwise



El modelo de características encontrado está conformado por: NSPD, índice de lobulaciones, área de la lesión, disimilaridad, dureza media del contorno en modo B, dureza media de la banda interna y dureza media del contorno del elastograma.

5.3. FASE 3

Nuevamente se trabajó con las 63 muestras divididas en dos grupos, pero en este caso se probaron los dos vectores de características encontrados en la fase 2.

5.3.1. Kmeans

Se realizó el entrenamiento de Kmeans con un grupo de 32 muestras y el vector de características encontrado manualmente con la herramienta *silhouette*, los resultados se pueden observar en la Tabla 12.

Tabla 12. Resultado entrenamiento Kmeans con vector de características Silhouette

Muestra	Clase Kmeans	Diagnostico real	Muestra	Clase Kmeans	Diagnostico real
1	2	Benigno	17	1	Sospechoso
2	2	Benigno	18	1	Sospechoso
3	2	Benigno	19	1	Sospechoso
4	3	Benigno	20	3	Maligno
5	3	Benigno	21	3	Maligno
6	1	Benigno	22	1	Maligno
7	3	Benigno	23	1	Maligno
8	2	Benigno	24	2	Maligno
9	2	Benigno	25	1	Maligno
10	3	Benigno	26	1	Maligno
11	3	Benigno	27	1	Maligno
12	3	Benigno	28	3	Maligno
13	1	Sospechoso	29	3	Maligno
14	1	Sospechoso	30	1	Maligno
15	1	Sospechoso	31	1	Maligno
16	1	Sospechoso	32	1	Maligno

La optimización del vector de características con la ayuda de *silhouette* no mejoró la separación de las clases, de nuevo se presentan las lesiones sospechosas agrupadas en una misma clase y las lesiones malignas y benignas sin una clase dominante. Se procedió a probar el vector de características encontrado con la herramienta *stepwise* y se obtuvieron los siguientes resultados:

Tabla 13. Resultado entrenamiento Kmeans con vector de características Stepwise

Muestra	Clase Kmeans	Diagnostico real	Muestra	Clase Kmeans	Diagnostico real
1	1	Benigno	17	3	Sospechoso
2	1	Benigno	18	3	Sospechoso
3	1	Benigno	19	3	Sospechoso
4	3	Benigno	20	3	Maligno
5	3	Benigno	21	2	Maligno
6	3	Benigno	22	1	Maligno

7	1	Benigno	23	1	Maligno
8	3	Benigno	24	2	Maligno
9	1	Benigno	25	1	Maligno
10	3	Benigno	26	2	Maligno
11	1	Benigno	27	3	Maligno
12	3	Benigno	28	1	Maligno
13	3	Sospechoso	29	2	Maligno
14	3	Sospechoso	30	3	Maligno
15	1	Sospechoso	31	3	Maligno
16	3	Sospechoso	32	3	Maligno

Los resultados observados en la Tabla 13 permiten descartar el algoritmo Kmeans ya que no se ve una mejora en la separación de las clases. Las lesiones sospechosas fueron las únicas donde se nota una etiqueta predominante, sin embargo, esta etiqueta también se encuentra en las lesiones malignas y benignas.

5.3.2. Red neuronal

Se realizó el entrenamiento y prueba con los dos grupos de muestras, utilizando las características halladas con *silhouette*. De las 63 muestras evaluadas por las RNAs, 55 fueron bien clasificadas, mientras que solo 8 muestras indicadas en rojo tuvieron una mala clasificación, con un porcentaje de aciertos cercano al 87 %. Todas las lesiones benignas fueron bien clasificadas y los desaciertos se encontraron entre las lesiones sospechosas y malignas (Tabla 14 y 15).

Tabla 14. Clasificación RNA con vector de características Silhouette grupo 1

Muestra	Clase RNA	Diagnóstico Real	Muestra	Clase RNA	Diagnóstico Real
1	10	Benigno	17	20	Sospechoso
2	10	Benigno	18	20	Sospechoso
3	10	Benigno	19	20	Sospechoso
4	10	Benigno	20	30	Maligno
5	10	Benigno	21	30	Maligno
6	10	Benigno	22	30	Maligno

7	10	Benigno	23	20	Maligno
8	10	Benigno	24	30	Maligno
9	10	Benigno	25	30	Maligno
10	10	Benigno	26	30	Maligno
11	10	Benigno	27	30	Maligno
12	10	Benigno	28	30	Maligno
13	20	Sospechoso	29	30	Maligno
14	30	Sospechoso	30	30	Maligno
15	20	Sospechoso	31	20	Maligno
16	20	Sospechoso			

Tabla 15. Clasificación RNA con vector de características Silhouette grupo 2

Muestra	Clase RNA	Diagnóstico Real	Muestra	Clase RNA	Diagnóstico Real
1	10	Benigno	17	20	Sospechoso
2	10	Benigno	18	20	Sospechoso
3	10	Benigno	19	30	Sospechoso
4	10	Benigno	20	30	Maligno
5	10	Benigno	21	30	Maligno
6	10	Benigno	22	30	Maligno
7	10	Benigno	23	30	Maligno
8	10	Benigno	24	30	Maligno
9	10	Benigno	25	30	Maligno
10	10	Benigno	26	20	Maligno
11	10	Benigno	27	30	Maligno
12	10	Benigno	28	30	Maligno
13	30	Sospechoso	29	20	Maligno
14	20	Sospechoso	30	30	Maligno
15	30	Sospechoso	31	30	Maligno
16	20	Sospechoso	32	30	Maligno

Los resultados obtenidos con el vector de características arrojado por *stepwise* fueron los siguientes (Tabla 16 y Tabla 17):

Tabla 16. Clasificación RNA con vector de características Stepwise grupo 1

Muestra	Clase RNA	Diagnostico Real	Muestra	Clase RNA	Diagnostico Real
1	10	Benigno	17	30	Sospechoso
2	10	Benigno	18	20	Sospechoso
3	10	Benigno	19	20	Sospechoso
4	10	Benigno	20	30	Maligno
5	10	Benigno	21	30	Maligno
6	10	Benigno	22	30	Maligno
7	10	Benigno	23	30	Maligno
8	10	Benigno	24	30	Maligno
9	10	Benigno	25	30	Maligno
10	10	Benigno	26	30	Maligno
11	10	Benigno	27	30	Maligno
12	10	Benigno	28	30	Maligno
13	10	Sospechoso	29	30	Maligno
14	20	Sospechoso	30	30	Maligno
15	10	Sospechoso	31	30	Maligno
16	20	Sospechoso			

Tabla 17. Clasificación RNA con vector de características Stepwise grupo 1

Muestra	Clase RNA	Diagnostico Real	Muestra	Clase RNA	Diagnostico Real
1	10	Benigno	17	20	Sospechoso
2	10	Benigno	18	20	Sospechoso
3	10	Benigno	19	30	Sospechoso
4	10	Benigno	20	30	Maligno
5	10	Benigno	21	30	Maligno
6	10	Benigno	22	30	Maligno
7	10	Benigno	23	30	Maligno
8	10	Benigno	24	30	Maligno
9	10	Benigno	25	30	Maligno
10	10	Benigno	26	30	Maligno
11	10	Benigno	27	30	Maligno
12	10	Benigno	28	30	Maligno
13	20	Sospechoso	29	30	Maligno
14	20	Sospechoso	30	30	Maligno
15	20	Sospechoso	31	30	Maligno
16	30	Sospechoso	32	30	Maligno

De las 63 muestras evaluadas por las RNAs, 58 fueron bien clasificadas, mientras que solo 5 muestras indicadas en rojo tuvieron una mala clasificación, con un porcentaje de aciertos cercano al 92 %. Todas las lesiones benignas y malignas fueron bien clasificadas y los desaciertos se encontraron en las lesiones sospechosas.

5.4. FASE 4

De las fases anteriores se concluye que el mejor clasificador es la RNA utilizando las características encontradas con *stepwise*. Se hace necesario la comparación de este clasificador con otro clasificador no lineal supervisado, para esto se escogieron los arboles de decisión, aunque los arboles de decisión no necesitan de una optimización de características previa se sigue trabajando con el vector de siete características. La implementación de un árbol de decisión en Matlab al igual que una RNA (Figura 51), necesita de un vector que defina la clase a la que pertenece cada una de las muestras de entrenamiento (T y T1) para los grupos ya conformados anteriormente (caracteristicas1 y caracteristicas2).

Figura 51. Entrenamiento árbol de decisión Matlab

```
% Entrenamiento Árbol de decisión  
tree1 = fitctree(caracteristicas1,T);  
tree2 = fitctree(caracteristicas2,T1);
```

Una vez entrenados los arboles de decisión se obtuvieron los resultados presentados en las tablas 18 y 19, de las 63 muestras evaluadas por los arboles de decisión, 57 fueron bien clasificadas, mientras que 6 muestras indicadas en rojo tuvieron una mala clasificación, con un porcentaje de aciertos cercano al 90 %. Todas las lesiones malignas fueron bien clasificadas y los desaciertos se encontraron en las lesiones benignas y sospechosas.

Tabla 18. Clasificación árbol de decisión con vector de características Stepwise grupo 1

Muestra	Clase Árbol D.	Diagnostico Real	Muestra	Clase RNA	Diagnostico Real
1	10	Benigno	17	30	Sospechoso
2	10	Benigno	18	20	Sospechoso
3	10	Benigno	19	20	Sospechoso
4	10	Benigno	20	30	Maligno
5	10	Benigno	21	30	Maligno
6	20	Benigno	22	30	Maligno
7	10	Benigno	23	30	Maligno
8	10	Benigno	24	30	Maligno
9	10	Benigno	25	30	Maligno
10	10	Benigno	26	30	Maligno
11	10	Benigno	27	30	Maligno
12	10	Benigno	28	30	Maligno
13	20	Sospechoso	29	30	Maligno
14	20	Sospechoso	30	30	Maligno
15	20	Sospechoso	31	30	Maligno
16	20	Sospechoso			

Tabla 19. Clasificación árbol de decisión con vector de características Stepwise grupo 2

Muestra	Clase Árbol D.	Diagnostico Real	Muestra	Clase RNA	Diagnostico Real
1	20	Benigno	17	20	Sospechoso
2	10	Benigno	18	20	Sospechoso
3	10	Benigno	19	20	Sospechoso
4	10	Benigno	20	30	Maligno
5	10	Benigno	21	30	Maligno
6	10	Benigno	22	30	Maligno
7	10	Benigno	23	30	Maligno
8	10	Benigno	24	30	Maligno
9	10	Benigno	25	30	Maligno
10	20	Benigno	26	30	Maligno
11	10	Benigno	27	30	Maligno

12	20	Benigno	28	30	Maligno
13	20	Sospechoso	29	30	Maligno
14	20	Sospechoso	30	30	Maligno
15	30	Sospechoso	31	30	Maligno
16	20	Sospechoso	32	30	Maligno

Aunque los dos clasificadores tuvieron un porcentaje de aciertos similar, las RNAs solo tuvieron errores en las lesiones sospechosas, mientras que los arboles de decisión tuvieron errores en lesiones benignas y sospechosas. Para representar estos resultados de una forma más clara se realizaron matrices de confusión a partir de las tablas anteriores (Tabla 20 y 21). Para una mayor comprensión de las matrices de confusión ver Anexo F.

Tabla 20. Matriz de confusión de las RNAs clasificación I

Matriz de confusión		Según el diagnóstico real			Total
		Benignas	Sospechosas	Malignas	
Según las RNAs	Benignas	24	2	0	26
	Sospechosas	0	9	0	9
	Malignas	0	3	25	28
Total		24	14	25	63
Precisión (%)		100	64,3	100	92

Tabla 21. Matriz de confusión de los árboles de decisión clasificación I

Matriz de confusión		Según el diagnóstico real			Total
		Benignas	Sospechosas	Malignas	
Según los árboles de decisión	Benignas	20	0	0	20
	Sospechosas	4	12	0	16
	Malignas	0	2	25	27
Total		24	14	25	63
Precisión (%)		83,3	85,7	100	90,5

5.5. FASE 5

Una vez obtenidos los resultados de la primera clasificación se escogió utilizar redes neuronales y árboles de decisión para realizar las pruebas de la segunda clasificación teniendo en cuenta la optimización del vector de características por medio de la herramienta *stepwise*. Para la segunda clasificación se cuentan con 28 muestras que se separaron en dos grupos de 13 y 15 muestras para entrenamiento y prueba.

Al realizar la optimización se obtuvo el siguiente vector de características: margen, relación de ejes, índice de lobulaciones, disimilaridad y dureza media del contorno en modo b. Con estas cinco características se encontraron los siguientes resultados de clasificación, en este caso los quistes se etiquetaron con 10, los fibroadenomas con 20 y los carcinomas invasivos con 30.

Tabla 22. Clasificación II RNA grupo 1

Muestra	Clase RNA	Diagnostico Real
1	10	Quiste
2	10	Quiste
3	10	Quiste
4	10	Quiste
5	20	Quiste
6	20	Fibroadenoma
7	20	Fibroadenoma
8	20	Fibroadenoma
9	10	Fibroadenoma
10	20	Fibroadenoma
11	20	Fibroadenoma
12	30	Carcinoma Invasivo
13	30	Carcinoma Invasivo
14	30	Carcinoma Invasivo
15	30	Carcinoma Invasivo

Tabla 23. Clasificación II RNA grupo 2

Muestra	Clase RNA	Diagnostico Real
1	10	Quiste
2	10	Quiste
3	10	Quiste
4	20	Quiste
5	20	Fibroadenoma
6	20	Fibroadenoma
7	20	Fibroadenoma
8	20	Fibroadenoma
9	20	Fibroadenoma
10	30	Carcinoma Invasivo
11	30	Carcinoma Invasivo
12	30	Carcinoma Invasivo
13	30	Carcinoma Invasivo

Tabla 24. Clasificación II Árbol de decisión grupo 1

Muestra	Clase Árbol D.	Diagnostico Real
1	10	Quiste
2	10	Quiste
3	10	Quiste
4	10	Quiste
5	20	Fibroadenoma
6	20	Fibroadenoma
7	20	Fibroadenoma
8	20	Fibroadenoma
9	20	Fibroadenoma
10	30	Carcinoma Invasivo
11	30	Carcinoma Invasivo
12	20	Carcinoma Invasivo
13	30	Carcinoma Invasivo

Tabla 25. Clasificación II Árbol de decisión grupo 2

Muestra	Clase Árbol D.	Diagnostico Real
1	10	Quiste
2	10	Quiste
3	10	Quiste
4	10	Quiste
5	10	Quiste
6	20	Fibroadenoma
7	20	Fibroadenoma
8	20	Fibroadenoma
9	20	Fibroadenoma
10	20	Fibroadenoma
11	20	Fibroadenoma
12	20	Carcinoma Invasivo
13	30	Carcinoma Invasivo
14	20	Carcinoma Invasivo
15	20	Carcinoma Invasivo

De las 28 muestras evaluadas las redes neuronales (Tabla 22 y Tabla 23) clasificaron correctamente un total de 25 casos, mientras que los arboles de decisión 24 casos (Tabla 24 y Tabla 25), los desaciertos se muestran en rojo. El porcentaje de aciertos es de 89% y 86% respectivamente, los datos se resumen en las matrices de confusión (Tabla 26 y 27).

Tabla 26. Matriz de confusión de las RNAs clasificación II

Matriz de confusión		Según el diagnóstico real			Total
		Quistes	Fibroadenomas	C. invasivos	
Según las RNAs	Quistes	7	1	0	8
	Fibroadenomas	2	10	0	12
	C. invasivos	0	0	8	8
Total		9	11	8	28
Precisión (%)		77,7	90,9	100	89,3

Tabla 27. Matriz de confusión de los árboles de decisión clasificación II

Matriz de confusión		Según el diagnóstico real			Total
		Quistes	Fibroadenomas	C. invasivos	
Según los árboles de decisión	Quistes	9	0	0	9
	Fibroadenomas	0	11	4	15
	C. invasivos	0	0	4	4
Total		9	11	8	28
Precisión (%)		100	100	50	85,7

Con los resultados obtenidos en la fase 4 y la fase 5 se toma la decisión de utilizar las redes neuronales como clasificadoras y los vectores de características obtenidos con *stepwise* en la aplicación software.

5.6. FASE 6

Con el fin de ver el aporte de las características de elastografía se hizo un entrenamiento y prueba a las redes neuronales de la primera y la segunda clasificación excluyendo las características de elastografía. Conservando las etiquetas para la primera clasificación: 10 para lesiones benignas, 20 para lesiones sospechosas y 30 para lesiones malignas y las etiquetas para la segunda clasificación: 10 quistes, 20 fibroadenomas y 30 carcinomas invasivos.

Tabla 28. Clasificación I RNA sin características de elastografía grupo 1

Muestra	Clase RNA	Diagnóstico Real	Muestra	Clase RNA	Diagnóstico Real
1	10	Benigno	17	20	Sospechoso
2	10	Benigno	18	20	Sospechoso
3	10	Benigno	19	20	Sospechoso
4	10	Benigno	20	30	Maligno
5	10	Benigno	21	30	Maligno
6	10	Benigno	22	20	Maligno
7	10	Benigno	23	30	Maligno
8	10	Benigno	24	20	Maligno

9	20	Benigno	25	20	Maligno
10	10	Benigno	26	20	Maligno
11	10	Benigno	27	20	Maligno
12	20	Benigno	28	30	Maligno
13	20	Sospechoso	29	30	Maligno
14	20	Sospechoso	30	20	Maligno
15	20	Sospechoso	31	20	Maligno
16	20	Sospechoso			

Tabla 29. Clasificación I RNA sin características de elastografía grupo 2

Muestra	Clase RNA	Diagnóstico Real	Muestra	Clase RNA	Diagnóstico Real
1	10	Benigno	17	20	Sospechoso
2	10	Benigno	18	20	Sospechoso
3	10	Benigno	19	30	Sospechoso
4	30	Benigno	20	30	Maligno
5	30	Benigno	21	30	Maligno
6	10	Benigno	22	20	Maligno
7	20	Benigno	23	20	Maligno
8	10	Benigno	24	30	Maligno
9	10	Benigno	25	20	Maligno
10	30	Benigno	26	30	Maligno
11	10	Benigno	27	30	Maligno
12	10	Benigno	28	20	Maligno
13	20	Sospechoso	29	30	Maligno
14	20	Sospechoso	30	30	Maligno
15	30	Sospechoso	31	30	Maligno
16	20	Sospechoso	32	30	Maligno

En la primera clasificación (Tabla 28 y Tabla 29) se puede observar una gran disminución del porcentaje de aciertos (69,8 %), esto se debe a que de las 7 características obtenidas después de la optimización 3 de ellas son del elastograma. En este caso se observan tres falsos positivos los cuales se pueden evidenciar en la Tabla 30.

Tabla 30. Matriz de confusión de las RNAs sin características de elastografía clasificación I

Matriz de confusión		Según el diagnóstico real			Total
		Benignas	Sospechosas	Malignas	
Según las RNAs	Benignas	18	0	0	18
	Sospechosas	3	12	11	27
	Malignas	3	2	14	18
Total		24	14	25	63
Precisión (%)		75	85,7	56	69,8

En la segunda clasificación el porcentaje disminuyó muy poco (82,1 %), debido a que de las 5 características obtenidas en la optimización solo 1 pertenece a las características de elastografía (Tabla 31 y 32), en la tabla 33 se observa la matriz de confusión calculada para la segunda clasificación excluyendo las características elastográficas. En el Anexo H, se presentan los resultados obtenidos para la primera clasificación (lesiones benignas, sospechosas y malignas) utilizando el total de características descriptivas (3 características), morfológicas (6 características) y de textura (8 características) que suman un total de 17 características, solo se dejan por fuera las 5 características elastográficas.

Tabla 31. Clasificación II RNA sin características de elastografía grupo 1

Muestra	Clase RNA	Diagnostico Real
1	20	Quiste
2	10	Quiste
3	20	Quiste
4	10	Quiste
5	20	Quiste
6	20	Fibroadenoma
7	20	Fibroadenoma
8	20	Fibroadenoma
9	20	Fibroadenoma
10	30	Fibroadenoma
11	20	Fibroadenoma

12	30	Carcinoma Invasivo
13	30	Carcinoma Invasivo
14	30	Carcinoma Invasivo
15	30	Carcinoma Invasivo

Tabla 32. Clasificación II RNA sin características de elastografía grupo 2

Muestra	Clase RNA	Diagnostico Real
1	10	Quiste
2	10	Quiste
3	10	Quiste
4	10	Quiste
5	20	Fibroadenoma
6	30	Fibroadenoma
7	20	Fibroadenoma
8	20	Fibroadenoma
9	20	Fibroadenoma
10	30	Carcinoma Invasivo
11	30	Carcinoma Invasivo
12	30	Carcinoma Invasivo
13	30	Carcinoma Invasivo

Tabla 33. Matriz de confusión de las RNAs sin características de elastografía clasificación II

Matriz de confusión		Según el diagnóstico real			Total
		Quistes	Fibroadenomas	C. invasivos	
Según las RNAs	Quistes	6	0	0	6
	Fibroadenomas	3	9	0	12
	C. invasivos	0	2	8	10
Total		9	11	8	28
Precisión (%)		66,6	81,8	100	82,1

Con la característica de dureza media del contorno en modo b se realizó una aproximación a la clasificación UENO presentada en el software descargado, no se pudieron comprobar los resultados de esta clasificación ya que no se cuenta con un diagnóstico real para la misma. Los resultados obtenidos solo fueron

evaluados de manera visual teniendo en cuenta el conocimiento adquirido en la sección 2.3.2., se comprueba que las lesiones benignas tienen puntajes entre 1 y 3, las lesiones malignas 4 y 5 y las lesiones sospechosas puntajes variables entre 2 y 4 ver Tabla 34.

Tabla 34. Clasificación UENO

Muestra	Clasificación UENO	Diagnóstico de la BD	Muestra	Clasificación UENO	Diagnóstico de la BD
1	1	Benigna	33	3	Sospechosa
2	1	Benigna	34	3	Sospechosa
3	1	Benigna	35	2	Sospechosa
4	2	Benigna	36	3	Sospechosa
5	3	Benigna	37	3	Sospechosa
6	3	Benigna	38	3	Sospechosa
7	2	Benigna	39	5	Maligna
8	2	Benigna	40	5	Maligna
9	1	Benigna	41	5	Maligna
10	2	Benigna	42	4	Maligna
11	3	Benigna	43	5	Maligna
12	1	Benigna	44	4	Maligna
13	2	Benigna	45	4	Maligna
14	2	Benigna	46	4	Maligna
15	1	Benigna	47	5	Maligna
16	2	Benigna	48	4	Maligna
17	2	Benigna	49	4	Maligna
18	2	Benigna	50	5	Maligna
19	2	Benigna	51	5	Maligna
20	1	Benigna	52	5	Maligna
21	2	Benigna	53	4	Maligna
22	1	Benigna	54	4	Maligna
23	3	Benigna	55	4	Maligna
24	1	Benigna	56	4	Maligna
25	4	Sospechosa	57	5	Maligna
26	2	Sospechosa	58	5	Maligna
27	3	Sospechosa	59	5	Maligna
28	3	Sospechosa	60	5	Maligna
29	3	Sospechosa	61	5	Maligna
30	2	Sospechosa	62	5	Maligna

31	2	Sospechosa	63	4	Maligna
32	4	Sospechosa			

6. CONCLUSIONES Y TRABAJOS FUTUROS.

En el presente trabajo se diseñó e implementó una aplicación software para el apoyo al diagnóstico de lesiones de mama, para dicho diagnóstico se hace uso del cálculo de características extraídas de imágenes de ultrasonido y elastografía, a partir de ellas se realizan clasificaciones automáticas y se permite la comparación de la lesión evaluada con casos similares. De esta manera se brinda una herramienta de apoyo al profesional en el diagnóstico.

Se definieron seis fases de pruebas para verificar el funcionamiento de los clasificadores, a lo largo de estas pruebas se conoce el desempeño de los mismos y los posibles cambios que se pueden realizar en el vector de características para aumentar el desempeño. También se calculó el porcentaje de aciertos obtenido por los clasificadores supervisados con el fin de obtener el mejor clasificador para la implementación de la aplicación.

6.1. CONCLUSIONES

Con base al trabajo realizado y los resultados obtenidos se plantean las siguientes conclusiones:

- El uso de sistemas CAD da un orden al procesamiento de imágenes médicas, facilitando la construcción de algoritmos para el pre procesamiento, la segmentación, la extracción de características y la clasificación de lesiones. Las etapas de pre procesamiento y segmentación permiten detectar el contorno de las lesiones, actividad necesaria para la extracción de características.
- La optimización del vector de características mejora considerablemente la precisión en la clasificación, ya que al reducir la dimensión del vector se

eliminan características redundantes o irrelevantes y se encuentra una mejor combinación de atributos aumentando el desempeño del clasificador.

- De los grupos de características utilizados las características de textura fueron las que menos aportaron para la clasificación. En las dos clasificaciones realizadas, solo se hace uso de una de las ocho características encontradas mientras que del grupo de nueve características descriptivas y morfológicas se usan cinco y del grupo de cinco características elastográficas se usan cuatro.
- Al evaluar y comparar los clasificadores utilizados, se determinó que el clasificador lineal no supervisado Kmeans no es conveniente en esta aplicación, debido a que no se logra la separación entre clases. Por otra parte, los clasificadores no lineales supervisados utilizados tienen rendimientos altos.
- Se construyeron algoritmos para la clasificación de lesiones de mama a partir de imágenes de ecografía y elastografía alcanzando resultados satisfactorios en las pruebas experimentales, se realizaron dos clasificaciones con precisiones del 92 % y el 89,3 % en el mejor caso.
- La combinación de las técnicas de elastografía y ecografía en la caracterización de lesiones de mama se complementan, por lo que se mejora el rendimiento del diagnóstico y se evitan los falsos positivos, dado que la precisión del clasificador disminuyó en un 22 % al no utilizar las características elastográficas y presentó falsos positivos, además que estas técnicas no son invasivas.
- Por medio de la aplicación software se le brinda al usuario información de las características y el diagnóstico de la lesión evaluada, igualmente de casos similares lo que permite la comparación entre éstos.

6.2. TRABAJOS FUTUROS

- Haciendo uso de una base de datos con mayor número de casos, extender la clasificación entre las diferentes patologías que se pueden presentar en la mama.
- Debido a que BI-RADS es un método de clasificación estándar y en este trabajo no se pudo realizar esta clasificación por falta de información que permitiera corroborar los resultados, crear algoritmos que permitan esta clasificación al igual que la clasificación UENO a partir de imágenes mamarias de ecografía y elastografía, teniendo una base de datos que proporcione la información necesaria.
- Realizar algoritmos de clasificación que incluyan características basadas en modelos y analizar el impacto que estas generan en el diagnóstico de lesiones de mama mediante elastografía ultrasónica.
- Trabajar más a fondo en cada una de las etapas que conforman un sistema CAD y comparar las posibles metodologías que se encuentren en cada una de ellas.

BIBLIOGRAFÍA

- [1] C. Ramos Pardo y R. Cendales Duarte, «Incidencia, mortalidad y prevalencia de cáncer en Colombia 2007-2011,» Instituto Nacional de Cancerología-ESE, Bogotá, 2015.
- [2] L. López Carrillo, L. Torres Sánchez, J. Muñoz, R. Hernández Ramírez y F. M. Knaul, «Utilización correcta de las técnicas de detección de cáncer de mama en mujeres mexicanas,» *Salud pública de México*, vol. LVI, nº 5, pp. 538-546, 2014.
- [3] A. C. Society, «Resumen sobre en cáncer de seno,» American Cancer Society, Atlanta, 2014.
- [4] A. C. Society, «Mamogramas y otros procedimientos de imaginología de los senos,» American Cancer Society, Atlanta, 2015.
- [5] A. C. Society, «Prevención y detección temprana del cáncer de seno,» American Cancer Society, Atlanta, 2015.
- [6] L. Aibar, A. Santalla, M. López Criado, I. Gonzáles Pérez, M. Calderón, J. Gallo y J. Fernández Parra, «Clasificación radiológica y manejo de las lesiones mamarias,» *Clin Invest Gin Obst.*, vol. IIIVIII, nº 4, pp. 141-149, 2011.
- [7] N. I. Rios Briones, A. Montemayor Martínez, E. Pérez y R. Enriquez Rodríguez, «Imagenología de la glándula mamaria,» de *Imagenología*, México, El Manual Moderno, 2011, pp. 103-126.
- [8] J. Vázquez Lamadrid, «La magia de la imagen,» *Anales Médicos*, vol. LIX, nº 4, pp. 244-245, 2014.
- [9] C. Piden Villaseñor, M. Macías Palacios y A. Bernal González, «Principios físicos básicos del ultrasonido,» *Investigación en Discapacidad*, vol. I, nº 1, pp. 25-34, 2012.
- [10] Y. Zhou, «Ultrasound diagnosis of breast cancer,» *Journal of Medical Imaging and Health Informatics*, vol. III, nº 2, pp. 1-14, 2013.
- [11] S. Mansour y O. Omar, «Elastography ultrasound and questionable breast lesions: Does it count?,» *Radiology*, vol. LXXXI, nº 11, pp. 3234-3244, 2012.
- [12] F. Martín Díez, «Ecografía de mama. Controversias,» *Radiología*, vol. LII, nº S1, pp. 22-25, 2010.
- [13] H. E. Mendoza Niño, J. Rodríguez San Martín, Á. Lazcano Urbina, D. Rodríguez Sardin, J. Martín Ramos, G. López Villanueva, A. Ramírez Castellanos y J. Arango Aguilar, «Utilidad de la elastografía en lesiones mamarias categoría BI_RADS 3,» *Anales de Radiología México*, vol. X, nº 3, pp. 134-139, 2011.
- [14] M. Mellado Rodríguez y A. Osa Labrador, «Cribado de cáncer de mama. Estado actual,» *Radiología*, vol. LV, nº 4, pp. 305-314, 2013.
- [15] J. Peña Colmenares, E. Farías, D. Celis, L. Aguirre, B. Chasín, G. Parthe, D. Lanz y C. Umbria, «Angiosarcoma de mama. Reporte de un caso,» *Revista Venezolana de Oncología*, vol. XXVI, nº 3, pp. 227-231, 2014.

- [16] V. Gómez Bermúdez, J. Gonzáles Alonso y R. Ariza García, «Elastografía de mama,» *Imagen diagnóstica*, vol. I, nº 2, pp. 47-50, 2010.
- [17] J. Camps y M. Sentis, «Elastosonografía mamaria,» *Revista Chilena de Radiología*, vol. XIV, nº 3, pp. 122-127, 2008.
- [18] D. T. Geertsma, «Ultrasoundcases.info,» Prominent Media Web Development, [En línea]. Available: <http://www.ultrasoundcases.info/>. [Último acceso: 9 Abril 2016].
- [19] F. Guzmán Aroca, D. Abellán Rivera y M. Reus Pintado, «La elastografía: una nueva aplicación de la ecografía. ¿Cuál es su utilidad clínica?,» *Radiología*, vol. LVI, nº IV, pp. 290-294, 2014.
- [20] R. García Arguelles y T. G. R. María, «Relación ultrasonográfica y anatomopatológica de las lesiones quísticas complejas de la mama,» *Revista Habanera de Ciencias Médicas*, vol. XIII, nº 1, pp. 46-54, 2013.
- [21] F. Sarquis, D. Mysler, M. P. Cobos Bombardiere, M. Castro Barba, R. Mareau y A. Luna, «Hallazgos ecográficos y características hispatológicas del fibroadenoma celular de la mama,» *Revista Argentina de Radiología*, vol. LXXVII, nº 2, pp. 147-152, 2013.
- [22] C. Mesber, J. Peña, J. C. Rupérez, B. Chacín, A. Moreno y N. Sáez, «Fibroadenoma infante juvenil. Reporte de un caso,» *Revista Venezolana de oncología*, vol. XXV, nº 4, pp. 260-267, 2013.
- [23] A. M. Castillo Ortiz, «Lesiones atípicas benignas de mama en mujeres mayores de 25 años con diagnóstico de imagen BI_RADS IV confirmados con estudio hispatológico en la clínica de mama del hospital de especialidades Eugenio Espejo en el periodo 2011-2013,» Universidad Central del Ecuador, Quito, 2015.
- [24] A. Gallego Mariño, A. Ramírez Batista y J. A. Amado Martínez, «Tumor phyllodes,» *Revista electrónica Dr. Zoilo E. Marinello Vidaurreta*, vol. XLI, nº 4, pp. 1-7, 2016.
- [25] P. A. Fernández Sarabias, Y. Rodríguez Alberteris y N. J. Sanz Pupo, «Gran galactocele,» *Revista cubana de obstetricia y ginecología*, vol. XXXIX, nº 3, pp. 260-265, 2013.
- [26] C. Patiño Gonzales y E. Romero Manteola, «Papiloma intraductal metacrónico en un niño. Caso clínico,» *Arch Argent Pediatr*, vol. CXII, nº 6, pp. e314-e316, 2015.
- [27] M. Torres Tabarena, «Novedades de la 5.ª edición del sistema breast imaging reporting and data system (BI-RADS) del Colegio Americano de Radiología,» *Revista de Senología y Patología Mamaria*, vol. XXIX, nº 1, pp. 32-39, 2016.
- [28] C. M. Pavón Hernández, Y. Villaseñor Navarro, R. A. Cruz Morales, L. O. Aguilar Cortázar, M. P. Pérez Badillo y I. Pérez Zuñiga, «Nódulos: caracterización y categorización,» *Gaceta Mexicana de Oncología*, vol. XI, nº 4, pp. 260-267, 2012.
- [29] M. Martínez Gálvez, J. Aguilar Jiménez y J. I. Gil Izquierdo, «Evidencias y controversias de la categoría 3 de Breas Imaging and Reporting Data System en enfermedad mamaria,» *Revisa de Senología y Patología mamaria*, vol. XXVI, nº 1, pp.

18-24, 2013.

- [30] E. L. Sananay Auquilla y E. D. Salao Pérez, «Determinación de las características ecográficas y citoquímicas de las lesiones mamarias y su relación con factores de riesgo, en las pacientes atendidas en el servicio de mastología, hospital "Carlos Andrade Marín". Quito, período Enero-Octubre del 2013,» Universidad Nacional de Chimborazo, Quito, 2013.
- [31] G. Nitish, «Automatic Breast Lesion Segmentation and Scoring in Color Elastography Images,» 2012. [En línea]. Available: <http://nitishgupta.github.io/docs/philips.pdf>. [Último acceso: 10 Enero 2016].
- [32] A. Franco Hervert, J. Onofre Castillo, M. Putz Botello y A. Marcilla Serrato, «Valor predictivo de malignidad de nódulos sólidos categorías BI-RADS 4 y 5 por elastografía cualitativa,» *Anales de Radiología México*, vol. XIII, nº 1, pp. 23-29, 2014.
- [33] J. Tang, J. Xu y I. El naqa, «Computer-Aided Detection and Diagnosis of Breast Cancer with Mammography: Recent Advances,» *IEEE transactions on information technology in biomedicine*, vol. XIII, nº 2, pp. 236-251, 2009.
- [34] R. M. Rangayyan, S. Banik y J. Leo Desautels, «Computer-Aided Detection of Architectural Distortion in Prior Mammograms on Interval Cancer,» *Journal of Digital Imaging*, vol. XXIII, nº 5, pp. 611-631, 2010.
- [35] R. Llobet Azpitarte, «Aportaciones al Diagnóstico de cáncer asistido por ordenador,» Universidad Politécnica de Valencia, Valencia, 2006.
- [36] A. Peña Peñate, L. G. Silva Rojas y R. Alcolea Núñez, «Módulo de filtrado y segmentación de imágenes médicas digitales para el proyecto Vismedic,» *Revista Cubana de ciencias informáticas*, vol. X, nº 1, pp. 13-27, 2016.
- [37] M. Fernández Prada, A. Llanos Méndez y R. I. Gómez, «Eficacia del diagnóstico asistido por ordenador aplicado al cribado de cáncer de mama,» Agencia de Evaluación de Tecnologías Sanitarias de Andalucía Consejería de Igualdad, Salud y Políticas Sociales, Sevilla, 2013.
- [38] S. Ordaz Gutiérrez, «Detección de leucemia linfoblástica aguda usando lógica difusa y redes neuronales,» Instituto politécnico Nacional , México, 2011.
- [39] P. F. Horlelao Beja, «Segmentacao de imagens de Elastografia Mamaria,» Universidade do Algarve, Faro, 2013.
- [40] H. Cheng, J. Shang, W. Ju, Y. Guo y L. Zhang, «Automated breast cancer detection classification using ultrasound images : A survey,» *Pattern Recognition* , vol. XLIII, nº 1, pp. 299-317, 2010.
- [41] Y. S. Tapia Morales, «Reducción del ruido speckle en mamografías empleando filtros de difusión,» Universidad Veracruzana, Veracruz, 2013.
- [42] R. García, C. Gonzáles, R. De la Vega, A. Valverde y E. Sebem, «Análisis del comportamiento de filtros de reducción de speckle en imágenes ERS2-SAR,» *Teledetección y desarrollo regional* , vol. X, nº 1, pp. 325-328, 2003.

- [43] H. Gonzáles Acosta, «Reducción del ruido digital en señales de electrocardiogramas utilizando la transformada de wavelet,» Universidad de Carabobo , Bárbula, 2015.
- [44] C. Vega González, «Un método automático para la elección de la base de ondícula y la selección del umbral para la estimación de señales,» *Revista INGENIERÍA UC*, vol. XVI, nº 1, pp. 32-39, 2009.
- [45] C. G. Gavidia Calderón, «Segmentación de imágenes médicas mediante algoritmos de colonia de hormigas,» Pontificia Universidad Católica de Peru, San Miguel, 2014.
- [46] J. A. Cortés Osorio, J. A. Chaves Osorio y J. A. Mendoza Vargas, «Comparación cualitativa y cuantitativa de las técnicas básicas de umbralización local para el procesamiento digital de imágenes,» *Scientia et Technica Año XVII*, vol. II, nº 51, pp. 236-241, 2012.
- [47] D. P. Guamán Tisalema, «Análisis, Evaluación y Caracterización de imágenes termográficas utilizando redes neuronales artificiales y técnicas de extracción de características de textura de Haralick,» Universidad Técnica de Ambato, Ambato, 2015.
- [48] W.-J. Wu y W. K. Moon, «Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors,» *Breast Cancer Research and Treatment* , vol. LXXXIX, nº 1, pp. 179-185, 2005.
- [49] Y. Su, Y. Wang, J. Jiao y Y. Guo, «Automatic Detection and Classification of Breast Tumors In Ultrasonic Images Using Texture and Morphological Features,» *The open medical informatics journal*, vol. V, nº 1, pp. 26-37, 2011.
- [50] S. O. Lugo Reyes, G. Maldonado Colín y C. Murata, «Inteligencia artificial para asistir el diagnóstico clínico en medicina,» *Revista Alergia México*, vol. LXI, nº 1, pp. 110-120, 2014.
- [51] S. University, «Stanford CS221,» Stanford University, [En línea]. Available: <http://stanford.edu/~cpiech/cs221/handouts/kmeans.html>. [Último acceso: 13 Septiembre 2016].
- [52] Matlab, «Matlab,» Matlab, [En línea]. Available: <http://es.mathworks.com/>. [Último acceso: 31 Agosto 2016].
- [53] J. P. Vásquez López, «Red neuronal feedforward como estimador de patrones de corrientes en el interior del puerto de manzanillo sujeto a la acción de tsunamis,» Instituto Mexicano del Transporte, Querétano , 2014.
- [54] A. M. Ramos Cañón y J. A. Pacheco, «Determinación del tipo de proceso de remoción en masa en Bogotá mediante árboles de decisión,» *Revista ingeniería y región*, vol. X, nº 10, pp. 125-132, 2013.
- [55] P. A. Vizcaino Garzon, «Aplicación de técnicas de inducción de árboles de decisión a problemas de clasificación mediante el uso de WEKA(Waikato Environment For Knowledge Analysis),» Fundación Universitaria Konrad Lorenz, Bogotá, 2008.
- [56] J. Cervantes y J. Andrade, «Aprendizaje Automático y modelos de clasificación.

- Aplicación en la clasificación crediticia de los Gobiernos Autónomos descentralizados municipales como clientes del Banco del Estado,» *Revista EPN*, vol. XXXIII, nº 2, pp. 1-10, 2014.
- [57] D. A. Gómez Betancur y J. W. Branch Bedoya, «Sistema de decisión basado en lógica difusa para la detección de distorsiones de la arquitectura de la glándula mamaria,» *ITECKNE*, vol. IX, nº 2, pp. 118-127, 2012.
- [58] U. o. Cambridge, «Projects elastography,» University of Cambridge, [En línea]. Available: <http://mi.eng.cam.ac.uk/projects/elastography/>. [Último acceso: 16 Enero 2016].
- [59] U. o. Cambridge, «3D Ultrasound using Stradwin,» University of Cambridge , [En línea]. Available: <http://mi.eng.cam.ac.uk/~rwp/stradwin/>. [Último acceso: 15 Marzo 2016].
- [60] QT, «QT,» QT, [En línea]. Available: <https://www.qt.io/developers/>. [Último acceso: 30 Agosto 2016].
- [61] OpenCV, «OpenCV,» OpenCV, [En línea]. Available: <http://opencv.org/>. [Último acceso: 30 Agosto 2016].
- [62] Microsoft, «VisualStudio,» Microsoft, [En línea]. Available: <https://www.visualstudio.com/>. [Último acceso: 31 Agosto 2016].
- [63] Cmake, «Cmake,» Cmake, [En línea]. Available: <https://cmake.org/>. [Último acceso: 31 Agosto 2016].
- [64] N. Gupta, «Nitish Gupta,» 11 7 2016. [En línea]. Available: <http://nitishgupta.github.io/internships.html>. [Último acceso: 2016].
- [65] J. Ríos Díaz, J. J. Martínez Payá y M. E. del Baño Aledo, «El análisis textural mediante las matrices de co ocurrencia (GLCM) sobre imagen ecográfica del tendón rotuliano es de utilidad para la detección de cambios hostológicos tras un entrenamiento con plataforma de vibración.,» *Cultura, Ciencia y Deporte*, vol. IV, nº 11, pp. 91-102, 2009.
- [66] A. Pino, J. Pladellorens Mallofré y J. Anto Roca, «Determinación de las propiedades superficiales del papel utilizando el análisis de la textura del patrón de speckle.,» *Sociedad Española de óptica* , vol. XLIII, nº 1, pp. 43-48, 2010.
- [67] F. N. Vargas Plazas, «Implementación de un sistema Automático de identificación y caracterización de fallas para una plataforma robótica de exploración y reparación de tuberías hidrosanitarias mediante visión por computador,» Pontificia Universidad Javeriana, Bogotá, 2014.
- [68] M. Alves, B. Ferreira y F. Leta, «Evaluación de parámetros de rugosidad usando análisis de imágenes de diferentes microscopios óptimos y electrónicos,» *Información Tecnológica*, vol. XXII, nº 4, pp. 129-146, 2011.
- [69] H. Rivaz, E. Boctor, P. Foughi, R. Zellars, G. Fichtinger y G. Hager, «Ultrasound Elastography: A Dynamic Programming Approach,» *Transactions on medical imaging*,

vol. XXVII, nº 10, pp. 1373-13377, 2008.

- [70] H. Rivaz, «Hassan Rivaz,» Hassan Rivaz, [En línea]. Available: <http://www.cs.jhu.edu/~rivaz/>. [Último acceso: 16 Enero 2016].
- [71] J. Gennisson, T. Deffieux, M. Fink y M. Tanter, «Ultrasound elastography: Principles and techniques,» *Diagnostic and Interventional Imaging*, vol. XCIV, nº 5, pp. 487-495, 2013.
- [72] K. Parker, M. Doyley y D. Rubens, «Imaging and elastic properties of tissue: the 20 years perspective,» *Physics Medicine and Biology*, vol. LVI, nº 1, pp. R1-R29, 2011.
- [73] A. C. Saavedra, J. Guerrero, J. Pinto, E. González y B. Castañeda, «Aspectos físicos de las técnicas elastográficas basadas en ultrasonido,» *Carcinos*, vol. IV, nº 2, pp. 34-44, 2014.
- [74] A. Itoh, E. Ueno, E. Tohno, I. Kamma, H. Takahashi, T. Shiina, M. Yamakawa y T. Matsumura, «Breast Disease: Clinical Application os US Elastography for Diagnosis,» *Radiology*, vol. CCXXXIX, nº 2, pp. 341-350, 2006.
- [75] R. Hernández Navarrete, «Procesamiento digital de imágenes de ultrasonido,» Instituto Politécnico Nacional , México, 2010.
- [76] M. Presutti, «La matriz de co ocurrencia en la clasificación multiespecial: Tutorial para la enseñanza de medidas texturales en cursos de grado universitario.,» Universidad Nacional de La Plata, La Plata, 2004.

ANEXO A: TÉCNICAS DE ELASTOGRAFÍA E IMAGENOLOGÍA

A.1. Técnicas de elastografía cuantitativa

- **Shear Wave Elasticity Imaging (SWEI):** se usan fuerzas de radiación acústica (FRA) para realizar el desplazamiento vertical de los tejidos, este desplazamiento permite verificar la dureza del área evaluada. Esto se considera una técnica cualitativa, pero SWEI adicionalmente proporciona información cuantitativa.

Teniendo en cuenta el desplazamiento y el tiempo en que tarda el tejido en volver a su posición inicial después de ser expuesto a FRAs se calcula la velocidad, con esta se muestran imágenes en tiempo real a escala de colores y el módulo de elasticidad de la región de interés.

- **Crawling Wave:** se hace uso de dos generadores de ondas ubicados en las partes laterales del tejido sospechoso, que producen una interferencia de baja frecuencia gracias a que cada uno emite ondas desfasadas entre sí. Las ondas de baja frecuencia son percibidas por el transductor que se ubica en la parte superior, posteriormente por medio de algoritmos de correlación se observa la interferencia de las ondas, para un tejido duro la interferencia será mucho mayor. La dificultad de Crawling Wave es la ubicación de las fuentes generadoras por lo que se considera una técnica experimental.
- **Supersonic shear imaging (SSI):** en el desarrollo de esta técnica se aplican varios haces de ultrasonido de manera continua para la generación de ondas de corte, dichas ondas se distorsionan al encontrar masas en los tejidos. El uso de un escáner de ultrasonido ultra rápido es necesario para realizar las tareas de generación y recepción de ondas en tiempos inferiores a los 20 milisegundos.

- **Vibro elastografía:** el transductor es utilizado como una fuente de vibración para crear movimiento en el tejido, además de recibir las ondas de desplazamiento. Las imágenes generadas por medio de ésta técnica son mejores, pero se pueden producir perturbaciones externas.

A.2. Técnicas de elastografía cualitativa

En la elastografía cualitativa se puede observar la presencia de masas duras en la región de interés, maneja parámetros que indican la dureza relativa en las áreas del tejido separando zonas duras de blandas. Entre los enfoques cualitativos más utilizados se encuentran:

- **Elastografía por compresión:** se aplica una compresión cuasiestática en la zona de interés y se comparan las ondas de ultrasonido obtenidas antes y después de la compresión, se basa en la relación tensión-deformación, además se obtienen parámetros elásticos gracias a que los tejidos después de la compresión retornan a su forma inicial. La compresión se estima calculando la derivada del desplazamiento del tejido antes y después de la presión aplicada. Por otra parte en los tejidos sanos el módulo de Young es bajo, es decir que se comprimen más; en los tejidos duros el módulo es alto debido a que no tienden a comprimirse.
- **Impulso por fuerzas de radiación acústica (ARFI):** se aplica una fuerza de radiación acústica (FRA) de alta intensidad por un tiempo corto sobre la región de interés, ésta fuerza viaja de forma vertical desplazando los tejidos en ésta dirección, para calcular el desplazamiento y el tiempo que tarda en retornar se toma la posición inicial y final, un tejido blando tendrá mayor desplazamiento y menor tiempo para regresar a su posición inicial. El uso de FRAs aumenta la temperatura del tejido reduciendo el tiempo de exposición al paciente y limitando la intensidad.

- **Sonoelastografía:** en ésta técnica se hace uso de ondas de corte de baja frecuencia y baja amplitud para excitar el tejido que se quiere diagnosticar, debido a la relación inversa entre la dureza del tejido y la amplitud de la onda se puede comparar la suavidad relativa de los tejidos. Un tejido duro responde ante la presencia de ondas de baja amplitud, mientras que los tejidos suaves se excitan con una alta vibración [71], [72], [73].

A.3. Técnicas alternativas de imagenología

Las técnicas de imagenología descritas anteriormente son las más utilizadas en el diagnóstico y detección del cáncer de mama, dentro de la medicina se encuentran otros estudios de imágenes alternativos, algunos de ellos son:

- **Ductograma (Galactograma):** Técnica en donde se introduce material de contraste por medio de un pequeño tubo metálico, el cual se inserta en la abertura de un conducto del pezón que presente secreción. Posteriormente se realiza una radiografía y se observa si hay masas o bultos dentro del ducto [4].
- **Escáner nuclear:** En un estudio de medicina nuclear se inyectan en el cuerpo sustancias radioactivas y se utilizan cámaras especiales para ver el recorrido de las sustancias. La medicina nuclear a diferencia de las otras técnicas de imagenología basadas en los cambios estructurales que generan los tumores, estudia el metabolismo de los tejidos [4].
- **PET (Tomografía por emisión de positrones):** Al igual que en la medicina nuclear se hace uso de sustancias radiactivas, en éste caso la sustancia utilizada es la glucosa la cual contiene un átomo radiactivo. Al ser inyectada en el seno las células cancerosas absorben gran cantidad de glucosa debido a su rápido crecimiento, aproximadamente una hora después se utilizan cámaras especiales para detectar las áreas con mayor radiactividad en el cuerpo.

La mayoría de casos en los que se utiliza la tomografía por emisión de positrones son aquellos en los que el paciente ya ha sido diagnosticado con cáncer y se desea conocer si la enfermedad se ha extendido, de lo contrario las PET no pueden distinguir entre tumores malignos y benignos [4].

- **Tomografía de impedancia eléctrica:** Durante el desarrollo de una tomografía de impedancia eléctrica se mide la conductividad eléctrica de los senos, la detección de un tumor maligno se basa en la idea de que las células cancerosas conducen la electricidad de distinta forma que las células normales. En la prueba se usan pequeños electrodos para aplicar corriente la cual se detecta luego sobre la piel del seno [4].
- **Termografía:** Mediante una cámara sensible al calor se detecta la temperatura sobre la superficie de la mama, la detección de los tumores es posible ya que éstos se caracterizan por generar mayor metabolismo y flujo sanguíneo, aumentando la temperatura de la zona. Una de las principales desventajas es su ineficacia al detectar el cáncer de seno en sus inicios [4].

ANEXO B: MANUAL DE INSTALACIÓN

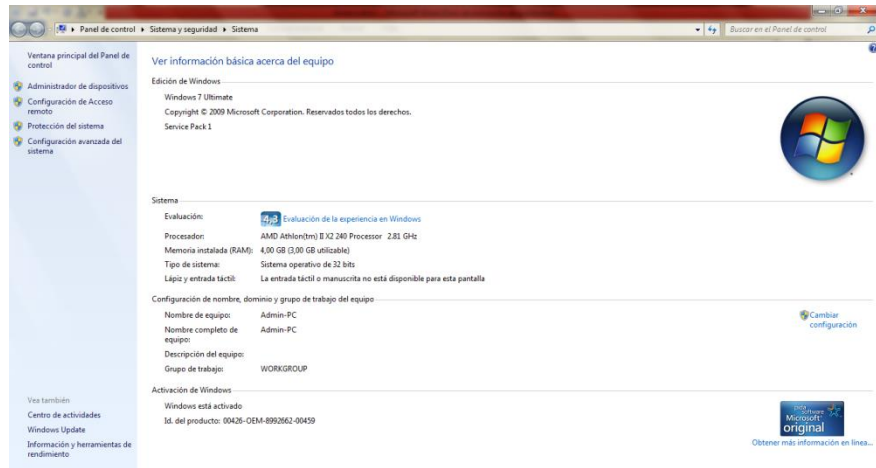
El objetivo de este manual es dar a conocer la instalación de los diferentes programas utilizados en la construcción de la aplicación:

B.1. Instalación de OpenCV 2.4.X.

En este caso de estudio, se utiliza la versión 2.4.6.

1. Crear una carpeta llamada **OpenCV**. Se recomienda que esta carpeta sea creada en el disco C. Dentro de la carpeta creada anteriormente, se deben descomprimir los archivos que se descargaron. Una vez descomprimidos los archivos se procede a configurar las variables de entorno, para habilitar el uso de OpenCV en el equipo.
2. Sobre el icono de **Equipo** dar clic derecho, en el menú de opciones que despliega, pulsar la opción **Propiedades**. Aparecerá una ventana como esta:

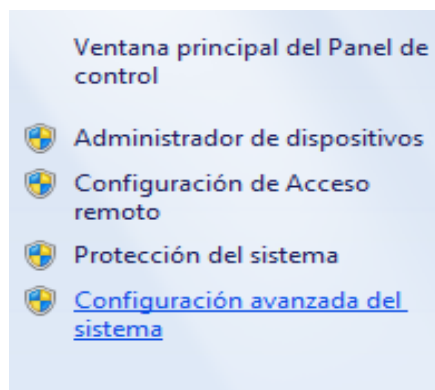
Figura B.1. Ventana propiedades del equipo



Es importante saber el tipo de arquitectura, ya que de esta manera se selecciona el tipo de compilador que se usara, en este caso es **vc11**.

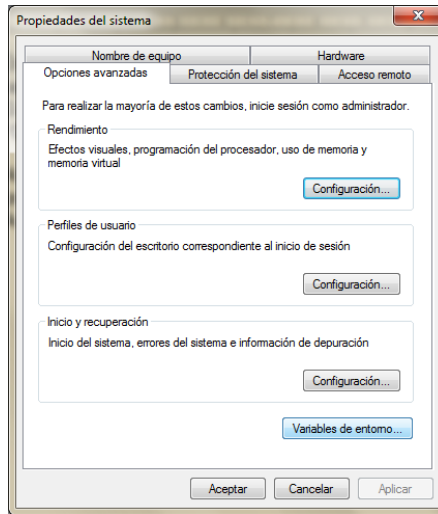
3. Para configurar las variables de entorno, en la misma ventana de propiedades, en la parte superior izquierda, pulsar **Configuración Avanzada del Sistema**.

Figura B.2. Configuración avanzada del sistema



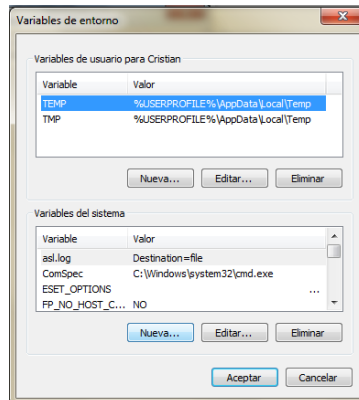
4. Sobre la nueva ventana, seleccionar **Variables de entorno...**

Figura B.3. Ventana de configuración propiedades del sistema



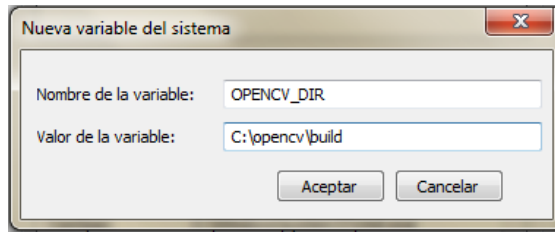
5. En la ventana **Variables de entorno**, dar clic al botón **Nueva...** ubicado en la parte inferior de esta.

Figura B.4. Ventana de configuración de variables de entorno



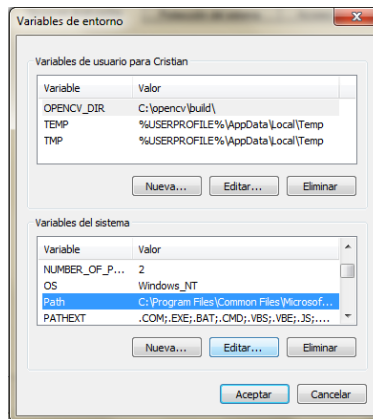
El siguiente paso es configurar el valor de la nueva variable con la siguiente información:

Figura B.5. Configuración de nueva variable de entorno



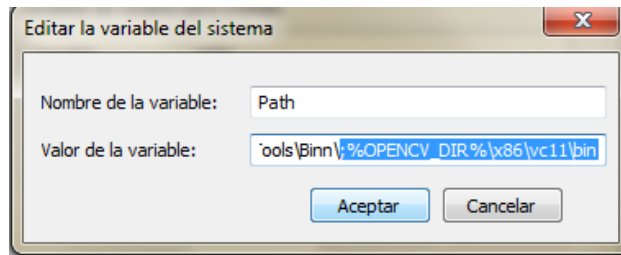
6. En la ventana de **Variables de entorno**, en la parte inferior se encuentra un recuadro llamado **Variables del sistema**, seleccionamos la línea **Path** y seleccionar Editar...

Figura B.6 .Configuración variable del sistema



7. En la siguiente ventana se encuentra una lista de direcciones pre-configuradas, no borrar ninguna, solo agregar ';' a la línea final y adicionar la siguiente dirección **%OPENCV_DIR%\x86\vc11\bin**.

Figura B.7. Adición de la variable del sistema para OpenCV

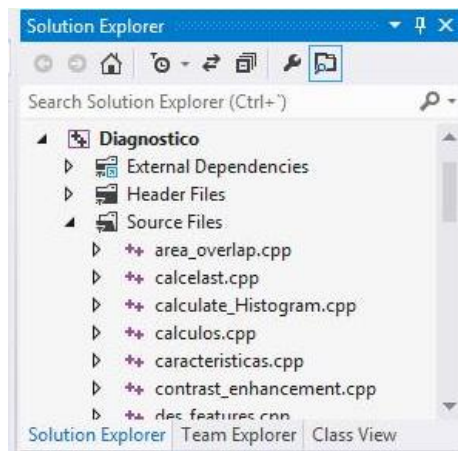


B.2. Configuración de OpenCV en Visual Studio.

Para completar la guía de instalación de OpenCV, es necesario configurar su uso en Visual Studio. Para ello se debe crear una hoja de propiedades en el proyecto.

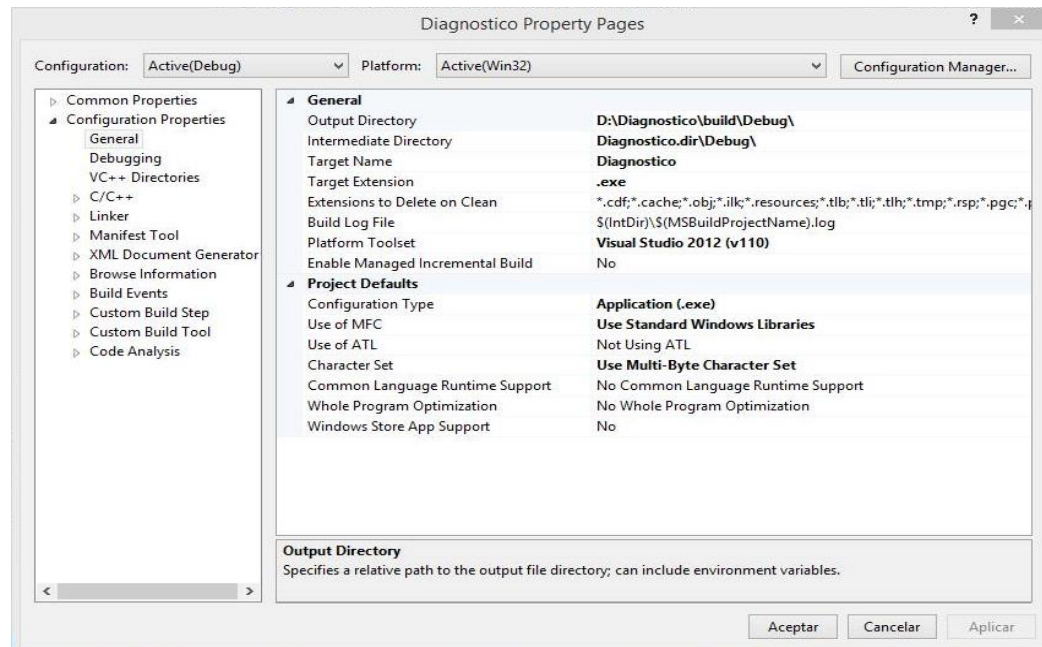
1. Abrir el proyecto, o crear un nuevo proyecto si está empezando. ubicarse en el explorador de soluciones.

Figura B.8. Explorador de soluciones Visual Studio



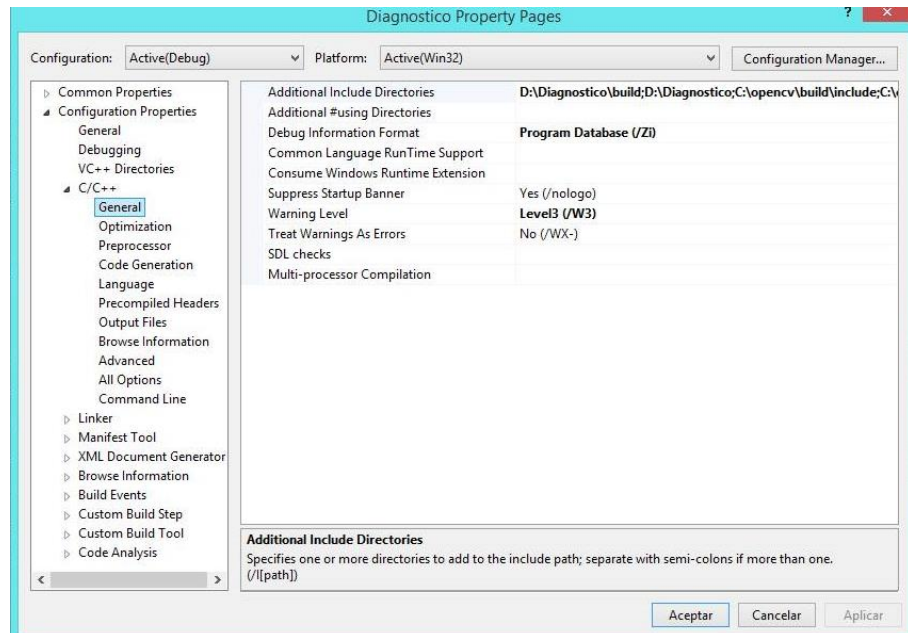
2. Clic derecho sobre el proyecto que hemos creado, y seleccionamos la pestaña **Propiedades**.

Figura B.9. Hojas de propiedades del proyecto



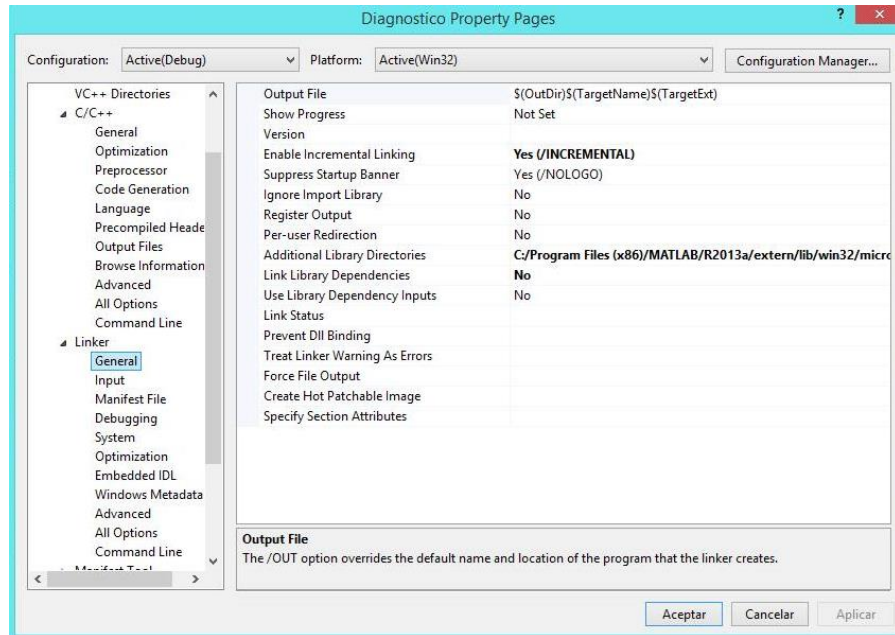
3. Se despliega el sub-directorio **C/C++**, se selecciona la opción **General**. En la parte derecha de la ventana, se modifica la opción **Directorios de inclusión adicionales**. En esta pestaña, se coloca la dirección donde se ubican los archivos *include* de OpenCV: **\$(OPENCV_DIR)\include**.

Figura B.10. Configuración directorios de inclusión OpenCV



4. Se despliega el sub-directorio **Vinculador**, se selecciona la opción **General**. En la parte derecha de la ventana, se modifica la opción **Directorios de bibliotecas adicionales**. En esta pestaña, se coloca la dirección donde se ubican los archivos “lib” de OpenCV: **‘\$(OPENCV_DIR)\x86\vc11\lib’**.

Figura B.11. Configuración directorio de archivos 'lib' OpenCV



5. En el mismo sub-directorio **Vinculador**, se selecciona la opción **Entrada**. En la parte derecha de la ventana, modificar la opción **Dependencias adicionales**.

Después de haber dado clic en la opción **Editar**, aparecerá un cuadro de dialogo. En ese cuadro se pondrán los nombres de los archivos de librería de OpenCV que se van a usar.

Los nombres de estos archivos, se pueden encontrar en la ubicación **'C:\opencv\build\x86\vc11'**.

“opencv_calib3d246.lib
opencv_contrib246.lib
opencv_core246.lib
opencv_features2d246.lib
opencv_flann246.lib

opencv_gpu246.lib
opencv_haartraining_engine.lib
opencv_highgui246.lib
opencv_imgproc246.lib
opencv_legacy246.lib
opencv_ml246.lib
opencv_nonfree246.lib
opencv_objdetect246.lib
opencv_ocl246.lib
opencv_photo246.lib
opencv_stitching246.lib
opencv_superres246.lib
opencv_ts246.lib
opencv_video246.lib
opencv_videostab246.lib”

Con esto se termina la configuración e instalación de OpenCV.

ANEXO C: CONSTRUCCIÓN LIBRERIA COMPARTIDAD EN MATLAB

Se crea una biblioteca compartida o DLL para agregar funcionalidad basada en MATLAB a la aplicación que está en desarrollo. Se selecciona una o más funciones basadas en MATLAB para incluirlas en la biblioteca compartida, y el compilador de MATLAB genera un archivo binario que se vincula a la aplicación. Las bibliotecas compartidas generadas por el compilador de MATLAB son compatibles con el entorno de desarrollo de Microsoft Visual Studio en Windows y con el C ++ Suite C de GNU / herramienta en la mayoría de las plataformas UNIX. En este caso la dll contiene las redes neuronales previamente entrenadas que son utilizadas para la clasificación, para incluir un archivo tipo net en una librería compartida es necesario generar una función de Matlab para esto se hace uso del comando *genFunction*.

Figura C.1. Comando *genFunction*

```
genFunction(net2, 'redneuronalm', 'MatrixOnly','yes', 'ShowLinks','no')
```

El primer parámetro *net2* es la red neuronal que ha sido entrenada, el segundo parámetro corresponde al nombre de la función que se va a generar, *MatrixOnly* anula la notación de celda por defecto y en su lugar genera una función que utiliza argumentos de matriz compatibles con las herramientas Matlab Coder, para las redes estáticas, las columnas de la matriz se interpretan como muestras independientes. Para las redes dinámicas las columnas de la matriz se interpretan como una serie de pasos de tiempo y *ShowLinks* desactiva el comportamiento por defecto de mostrar los enlaces generados y el código fuente.

Para la construcción de una librería compartida se hace uso de la siguiente línea de código:

Figura C.2. Construcción de la librería compartida

```
mcc -v -W cpplib:libstrains -T link:lib RNA RNAII
```

El comando *mcc* permite las funciones de Matlab para ser utilizadas en C++, *-v* muestra los pasos de compilación, *-W* controla la generación de envolturas de función, *cpplib:libstrains* genera la envoltura del código C/C++ para la integración con otras aplicaciones, en este caso tiempo por nombre *libstrains*, *-T link:lib* invoca un compilador C/C++ para crear un archivo binario como biblioteca compartida de la plataforma del código generado. *RNA* y *RNAII* son las funciones de Matlab generadas para las redes neuronales.

ANEXO D: ALGORITMO FILTRO WAVELET

```
#include "opencv2/opencv.hpp"
#include <iostream>
#include <vector>
#include <stdio.h>

using namespace cv;
using namespace std;

// signum

float sgn(float x)
{
    float res=0;
    if(x==0)
    {
        res=0;
    }
    if(x>0)
    {
        res=1;
    }
    if(x<0)
    {
        res=-1;
    }
    return res;
}

// Garrot shrinkage

float Garrot_shrink(float d,float T)
{
    float res;
    if(fabs(d)>T)
    {
        res=d-((T*T)/d);
    }
    else
    {
        res=0;
    }

    return res;
}

// Wavelet transform

static void cvHaarWavelet(Mat &src,Mat &dst,int NIter)
{
    float c,dh,dv,dd;
    assert( src.type() == CV_32FC1 );
    assert( dst.type() == CV_32FC1 );
```

```

int width = src.cols;
int height = src.rows;
for (int k=0;k<NIter;k++)
{
    for (int y=0;y<(height>>(k+1));y++)
    {
        for (int x=0; x<(width>>(k+1));x++)
        {
c=(src.at<float>(2*y,2*x)+src.at<float>(2*y,2*x+1)+src.at<float>(2*y+1,2*x)+src.at<f
loat>(2*y+1,2*x+1))*0.5;
            dst.at<float>(y,x)=c;

            dh=(src.at<float>(2*y,2*x)+src.at<float>(2*y+1,2*x)-
src.at<float>(2*y,2*x+1)-src.at<float>(2*y+1,2*x+1))*0.5;
            dst.at<float>(y,x+(width>>(k+1)))=dh;

            dv=(src.at<float>(2*y,2*x)+src.at<float>(2*y,2*x+1)-
src.at<float>(2*y+1,2*x)-src.at<float>(2*y+1,2*x+1))*0.5;
            dst.at<float>(y+(height>>(k+1)),x)=dv;

            dd=(src.at<float>(2*y,2*x)-src.at<float>(2*y,2*x+1)-
src.at<float>(2*y+1,2*x)+src.at<float>(2*y+1,2*x+1))*0.5;
            dst.at<float>(y+(height>>(k+1)),x+(width>>(k+1)))=dd;
        }
    }
    dst.copyTo(src);
}
}

//Inverse wavelet transform

static void cvInvHaarWavelet(Mat &src,Mat &dst,int NIter, float SHRINKAGE_T)
{
    float c,dh,dv,dd;
    assert( src.type() == CV_32FC1 );
    assert( dst.type() == CV_32FC1 );
    int width = src.cols;
    int height = src.rows;

    // NIter - number of iterations

    for (int k=NIter;k>0;k--)
    {
        for (int y=0;y<(height>>k);y++)
        {
            for (int x=0; x<(width>>k);x++)
            {
                c=src.at<float>(y,x);
                dh=src.at<float>(y,x+(width>>k));
                dv=src.at<float>(y+(height>>k),x);
                dd=src.at<float>(y+(height>>k),x+(width>>k));

                dh=Garrot_shrink(dh,SHRINKAGE_T);
                dv=Garrot_shrink(dv,SHRINKAGE_T);
            }
        }
    }
}

```

```

        dd=Garrot_shrink(dd,SHRINKAGE_T);

        dst.at<float>(y*2,x*2)=0.5*(c+dh+dv+dd);
        dst.at<float>(y*2,x*2+1)=0.5*(c-dh+dv-dd);
        dst.at<float>(y*2+1,x*2)=0.5*(c+dh-dv-dd);
        dst.at<float>(y*2+1,x*2+1)=0.5*(c-dh-dv+dd);
    }
}
Mat C=src(Rect(0,0,width>>(k-1),height>>(k-1)));
Mat D=dst(Rect(0,0,width>>(k-1),height>>(k-1)));
D.copyTo(C);
}
}

void wavelet(Mat im,Mat *Imfiltered)
{
    const int NIter=1;

    Mat GrayFrame=Mat(im.rows, im.cols, CV_8UC1);
    Mat Src=Mat(im.rows, im.cols, CV_32FC1);
    Mat Dst=Mat(im.rows, im.cols, CV_32FC1);
    Mat Temp=Mat(im.rows, im.cols, CV_32FC1);
    Mat Filtered=Mat(im.rows, im.cols, CV_32FC1);

    Dst=0;
    cvtColor(im, GrayFrame, CV_BGR2GRAY);
    GrayFrame.convertTo(Src,CV_32FC1);
    cvHaarWavelet(Src,Dst,NIter);

    Dst.copyTo(Temp);

    cvInvHaarWavelet(Temp,Filtered,NIter,50);

    double M=0,m=0;

    // Normalization to 0-1 range (for visualization)
    minMaxLoc(Dst,&m,&M);
    if((M-m)>0)
    {
        Dst=Dst*(1.0/(M-m))-m/(M-m);
    }

    minMaxLoc(Filtered,&m,&M);
    if((M-m)>0)
    {
        Filtered=Filtered*(1.0/(M-m))-m/(M-m);
    }

    *Imfiltered = Filtered;
}

```

ANEXO E: ALGORITMOS PARA LA EXTRACCION DE CARACTERISTICAS

E.1. CARACTERÍSTICA DE FORMA

Se parte de la imagen binaria generada a partir del contorno encontrado en la segmentación (Figura E1).

Figura E1. Imagen binaria



Se realizó una operación morfológica para eliminar huecos pequeños y unir componentes cercanos, en OpenCV se hace uso del comando *morphologyEx*, el primer parámetro de este comando es la imagen binaria, el segundo parámetro es la imagen de salida, en el tercer parámetro se define la operación morfológica a realizar, en este caso cierre y por último el elemento estructurante. El elemento estructurante debe ser definido con anterioridad por medio de *getStructuringElement* donde el parámetro más relevante es el segundo el cual define el tamaño del elemento estructuraste.

```
// Operación morfológica
Mat SE1 = getStructuringElement(MORPH_RECT, Size(10,10), Point(-1,-1));
morphologyEx(bmode_bin, bmode_bin, MORPH_CLOSE, SE1);
```

Figura E.2. Resultado de operación morfológica aplicada



Una vez aplicada la operación morfológica se encuentra el contorno por medio del comando *findContours*, sus parámetros son: la imagen a la que se le aplicó la operación morfológica, el vector de puntos que almacena cada uno de los contornos encontrados, un vector opcional que contiene información topológica de la imagen para cada uno de los contornos, el modo de recuperación del contorno, el método de aproximación del contorno y el punto de desplazamiento.

```
findContours(bmode_bin, contour_des, hierarchy_des, CV_RETR_TREE, CV_CHAIN_APPROX_TC89_KCOS, Pt4);
```

Después de encontrar los contornos se calcula la envolvente convexa de cada uno de ellos por medio de *convexHull*, sus parámetros son: cada uno de los contornos encontrados, un vector de puntos donde se almacenan las envolventes convexas de cada uno de los contornos y la orientación del contorno.

```
convexHull(Mat(contour_des[i]), hull[i], false);
```

Con lo realizado anteriormente se encontraron diferentes contornos, el contorno con mayor área representa el contorno de la lesión, para encontrar este contorno se realiza una comparación entre áreas el cálculo de las áreas se realiza con el comando *contourArea*, este comando recibe cada uno de los contornos.

```
area = contourArea(contour_des[i]).
```

Una vez encontrado el contorno que representa la lesión se calcula el área de la lesión y de su respectiva envolvente con las que se encuentra la relación entre áreas.

E.2. CARACTERÍSTICA DE ORIENTACIÓN

La característica de orientación se calcula con la función *Parallel* construida en OpenCV, el primer parámetro de esta función es el contorno de la lesión, el segundo parámetro es la imagen en escala de grises que permite visualizar los ejes encontrados.

```
Parallel(contour_des[sm],bmodel)
```

Para encontrar los ejes de la lesión se buscan los puntos situados en los extremos, para esto se compara la componente en x y en y de todos los puntos que pertenecen al contorno. Una vez encontrado los cuatro puntos situados en los extremos se calcula la distancia entre ellos.

```
AxisX = maxX.x - minX.x  
AxisY = maxY.y - minY.y
```

Donde *AxisX* es la longitud del eje x de la lesión, *maxX* es el punto con la mayor componente en x , *minX* es el punto con la menor componente en x , *AxisY* es la longitud del eje y de la lesión, *maxY* es el punto con la mayor componente en y , *minY* es el punto con la menor componente en y . Con *AxisX* y *AxisY* se calcula la relación entre los ejes.

```
axis_ratio = AxisY/AxisX;
```

E.3. CARACTERÍSTICA DE MARGEN

A partir de la imagen binaria se encuentran todos los contornos con el comando *findContours*, este retorna un vector con todos los contornos encontrados en la imagen los cuales son contados. El número de objetos es igual a los contornos que se encuentran por fuera de contorno de la lesión.

```

for (int j = 0; j < contours.size(); j++)
{
    if(j != sm1)
    {
        vector<Point> contour = contours[j];
        Point pnt = contour[0];
        if(pointPolygonTest(contours[sm1], pnt, false) < 0)
        {
            total_objects = total_objects + 1;
        }
    }
}

```

Se corren todos los contornos por medio del primer ciclo *for*, la variable *sm1* previamente encontrada corresponde a la posición del contorno de la lesión, para determinar que los objetos no se encuentren dentro de la lesión se extrae un punto (*pnt*) del contorno evaluado. El comando *pointPolygonTest* verifica que este punto se encuentre fuera de la lesión lo cual permite hacer el conteo de los objetos (*total_objects*).

E.4. CARACTERÍSTICA DE FACTOR DE FORMA

Se parte del contorno de la lesión (*contour_des[sm]*), con el comando *contourArea* se calcula el área y con el comando *arcLength* se calcula el perímetro, donde el segundo argumento especifica si el contorno es cerrado (*true*) o si solo es una curva.

```

float A_contour = contourArea(contour_des[sm]);
float P_contour = arcLength(contour_des[sm], true);

```

E.5. CARACTERÍSTICA DE NSPD

Con el contorno de la lesión y la envolvente convexa se encuentran los defectos de convexidad por medio de la sentencia *convexityDefects*, esta función retorna un vector que contiene el punto inicial, punto final, punto más lejano y la distancia del punto inicial al más lejano de cada uno de los defectos.

```
convexityDefects(contour_des[i],inthull[i],defects[i]);
```

```
for (int i=0;i<defectos.size();i++)  
{  
    distancia = static_cast<double>(defectos[i].val[3])/256 ;  
    if (distancia>10)  
    {  
        prominences = prominences + 1;  
        depressions = depressions + 1;  
    }  
}
```

E.6. CARACTERÍSTICA DE NÚMERO DE LOBULACIONES

Una vez encontrados los puntos de convexidad se construye la envolvente cóncava uniendo las depresiones mediante el comando *line*, el primer parámetro es la imagen binaria de ROI, el segundo y tercer parámetro corresponde a los puntos donde se ubican las depresiones, estos puntos se almacenaron anteriormente en el vector *p_lobulations*, los siguientes parámetros determinan el color, el ancho y en tipo de línea.

```
line(lob_index,p_lobulations[i],p_lobulations[i+1],CV_RGB(0,0,0),3,8);
```

Mediantes el comando *findContours* se encuentran los contornos que representan el área encerrada por la envolvente cóncava y los objetos externos a esta (lobulaciones). Para realizar el conteo de lobulaciones se calcula el tamaño del vector de contornos (*contour_lob.size*) menos el contorno de la envolvente cóncava.

```
findContours(lob_index1,contour_lob,hierarchy_lob,CV_RETR_TREE,CV_CHAIN_APPROX_TC89_KCOS,Pt);  
lobulations = contour_lob.size() - 1;
```

ANEXO F: MATRIZ DE CONFUSIÓN

En este anexo se explicará la construcción de una matriz de confusión, para esto se utilizará la matriz de confusión de las RNAs clasificación I (Tabla F.1.). El primer paso es llenar las casillas cuyos datos están subrayados en azul, en este caso las filas representan el diagnóstico según las RNAs y las columnas el diagnóstico que proporción la base de datos (BD).

En la columna sospechosas el primer dato (2) representa las lesiones clasificadas como benignas por las RNAs pero son lesiones sospechosas según la BD, el segundo dato (9) representa las lesiones que concuerdan en el diagnóstico y el tercer dato (3) son las lesiones clasificadas como malignas según las RNAs pero son lesiones sospechosas según el diagnóstico real.

Tabla F.1. Matriz de confusión de las RNAs clasificación I

Matriz de confusión		Según el diagnóstico Real			Total
		Benignas	Sospechosas	Malignas	
Según las RNAs	Benignas	24	2	0	26
	Sospechosas	0	9	0	9
	Malignas	0	3	25	28
Total		24	14	25	63
Precisión (%)		100	64,3	100	92

El segundo paso es calcular los totales para esto se suma los valores de las filas y columnas respectivamente, por último, se calcula la precisión la cual representa el porcentaje de aciertos de las RNAs. Por ejemplo, se calculará la precisión de lesiones sospechosas.

$$\text{precisión} = \frac{\text{lesiones sospechosas clasificadas como sospechosas} * 100}{\text{total de lesiones sospechosas}}$$

$$\text{precisión} = \frac{9 * 100}{14} = 64,3$$

Para calcular la precisión total del sistema (92), se utiliza la siguiente ecuación:

$$\text{precisión} = \frac{\text{lesiones bien diagnosticadas} * 100}{\text{total de lesiones diagnosticadas}}$$

Donde las *lesiones bien diagnosticadas* son la suma de la diagonal conformada por las celdas con fondo verde.

$$\text{precisión} = \frac{(24 + 9 + 25) * 100}{63} = 92.06$$

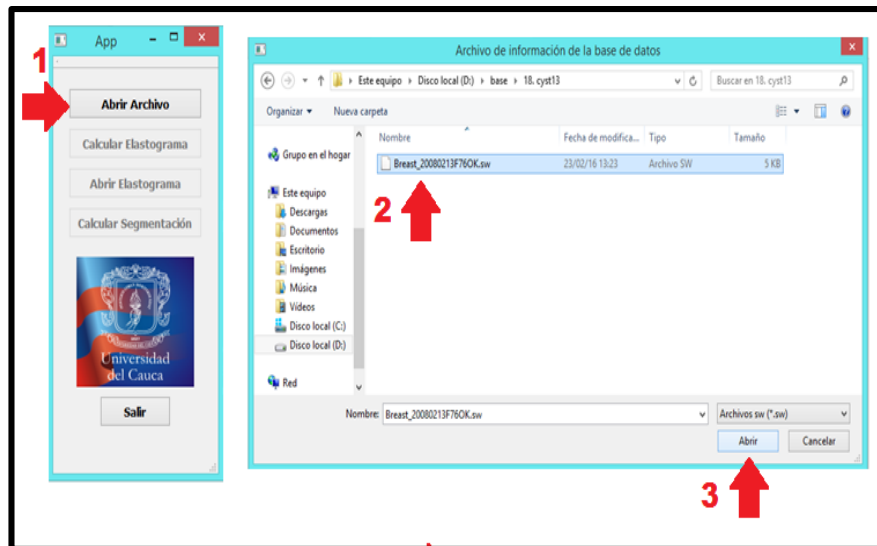
ANEXO G: MANUAL DE USUARIO

En esta guía se indica paso a paso el funcionamiento de la aplicación para el apoyo al diagnóstico de cáncer de mama y los posibles escenarios que se pueden presentar. Para el funcionamiento de la aplicación de debe tener instalado Matlab 2013 o el compilador de Matlab 2013 al igual que QT 5.4.

G.1. Cargar archivos .sw y .sxi

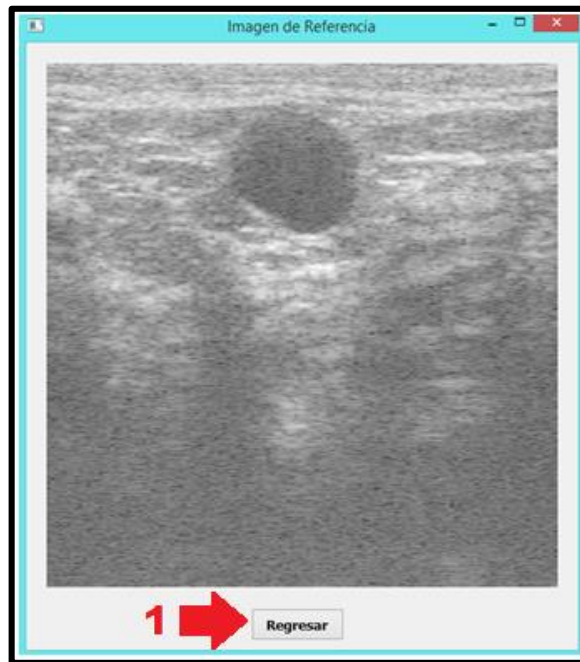
El primer paso es abrir los archivos .sw y .sxi que contienen la información ecográfica, es necesario que estos dos archivos se encuentren en la misma ubicación y tengan el mismo nombre. Una vez oprimido el botón **Abrir Archivo** en la pantalla principal **App** se realiza la búsqueda y apertura del archivo .sw, el archivo .sxi se cargará automáticamente.

Figura G.1. Inicio de la aplicación (1) Abrir Archivo (2) Seleccionar archivo .sw de la lesión a evaluar (3) Cargar archivo



Una vez cargados estos archivos se despliega la pantalla **Imagen de Referencia** donde se visualiza el primer marco en escala de grises de la ecografía a evaluar. Se debe oprimir el botón **Regresar** para continuar con el siguiente paso.

Figura G.2. Imagen de referencia (1) Regresar a la pantalla principal



G.2. Cargar el elastograma

Si se cuenta con un elastograma este se carga por medio del botón **Abrir elastograma**, al presionar el botón se debe buscar la ubicación del archivo (archivo con extensión yml), una vez ubicado se procede a abrirlo y se desplegara la ventana **Elastograma**. Se debe de fijar el marco de referencia para esto es necesario oprimir el botón **Calcular elastograma**, en la pantalla desplegada con la segunda barra de desplazamiento se elige el número del marco de referencia, una vez fijado el número se pulsa el botón **Salir**.

Figura G.3. Cargar elastograma (1) Abrir elastograma (2) Seleccionar archivo. yml de la lesión a evaluar (3) Cargar archivo (4) Regresar a la pantalla principal

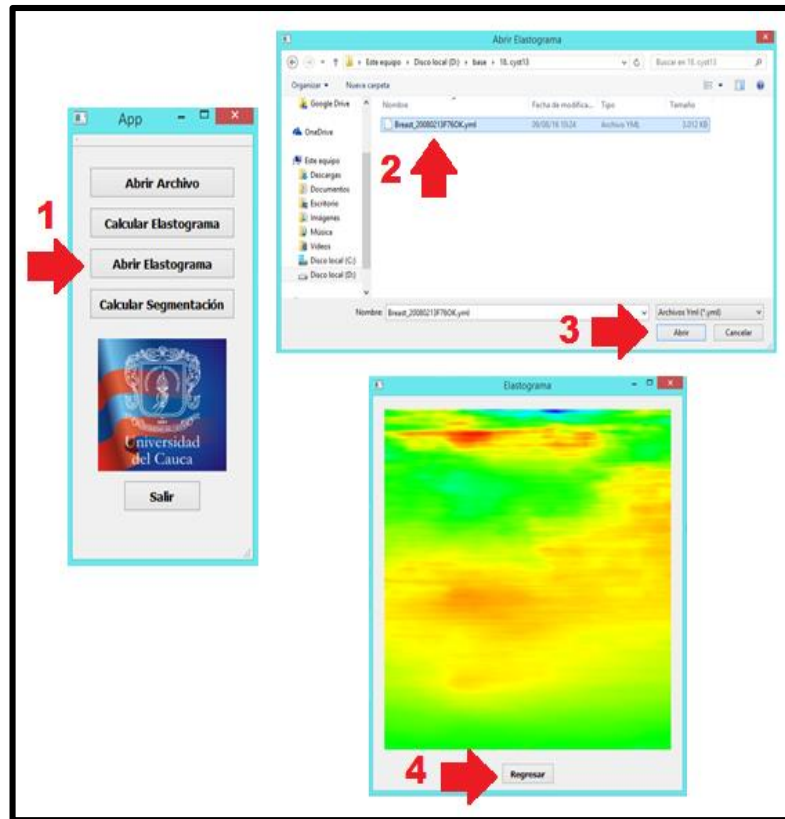
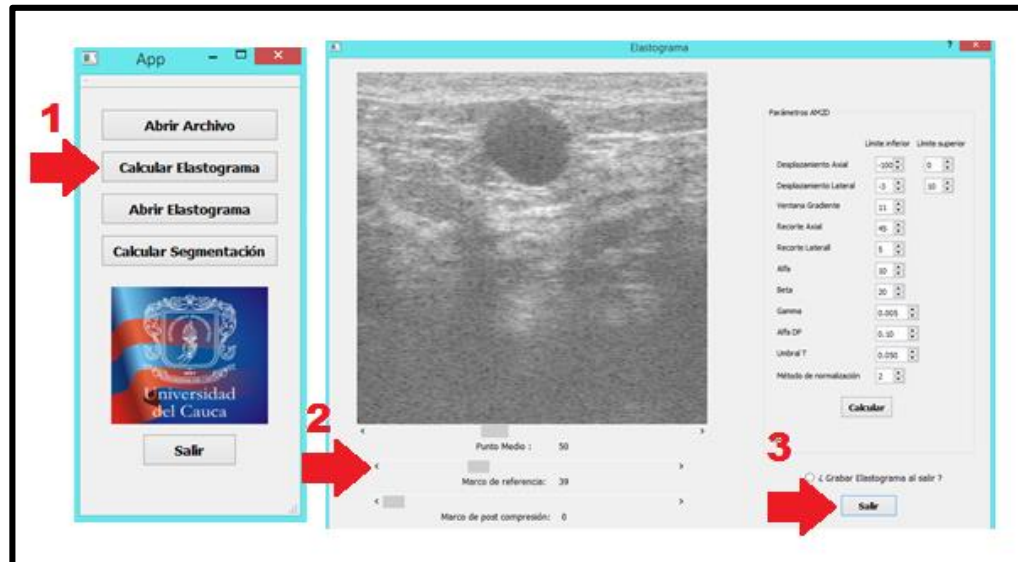
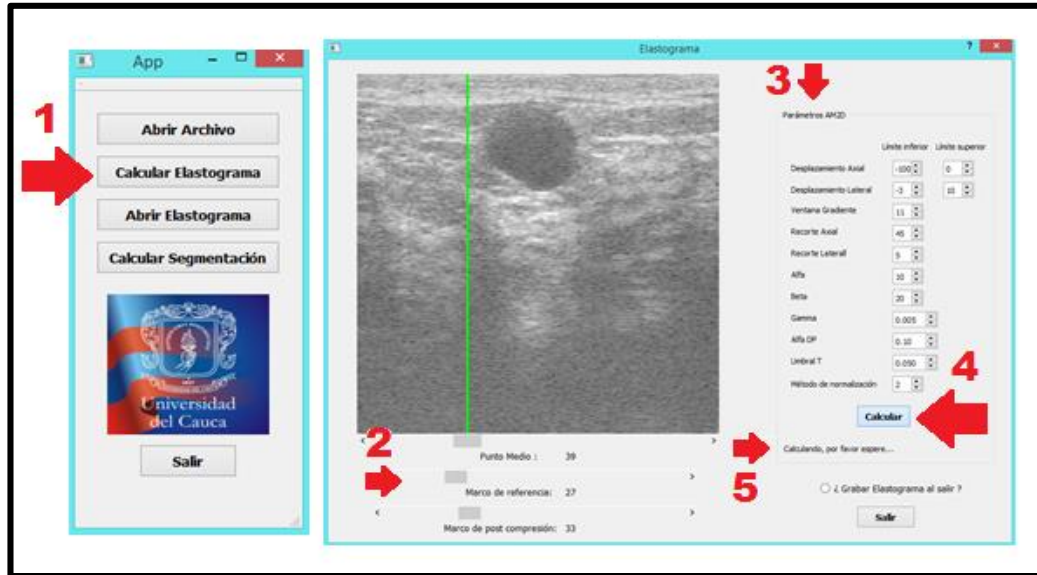


Figura G.4. Seleccionar Marco de referencia (1) Abrir pantalla Elastograma (2) Fijar el marco de referencia por medio de la barra deslizante (3) Regresar a la pantalla principal



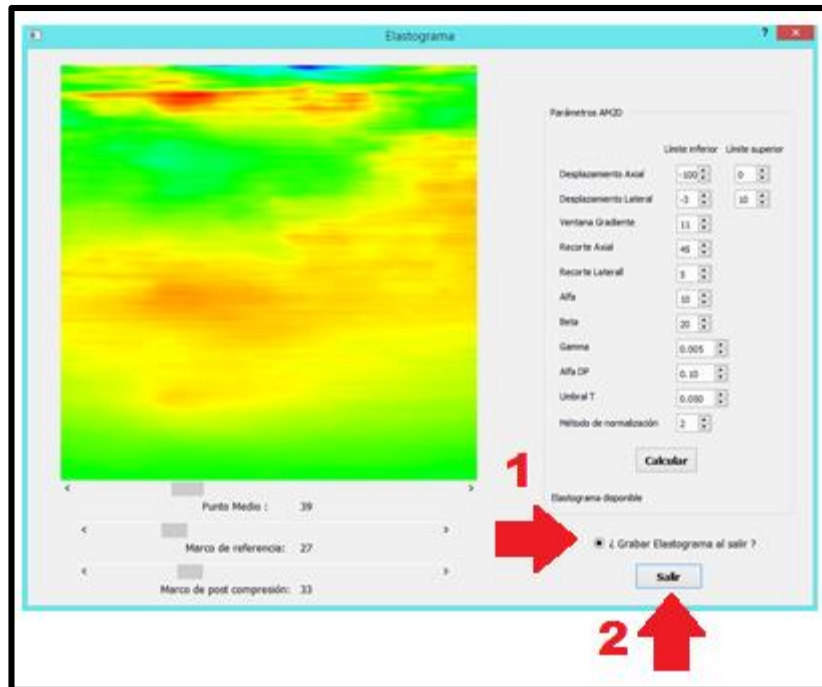
Para calcular el elastograma en caso que sea necesario se debe pulsar el botón **Calcular Elastograma** y se desplegará la pantalla **Elastograma**, dentro de esta se encuentran gran cantidad de términos, para la elección del marco de referencia, el marco de post compresión y el punto medio se cuenta con barras de desplazamiento en la parte inferior, los demás parámetros están ubicados en el recuadro **Parámetros AM2D**.

Figura G.5. Calcular Elastograma (1) Abrir pantalla Elastograma (2) Seleccionar parámetros con las barras deslizantes (3) Seleccionar parámetros (4) Calcular Elastograma (5) Estado del proceso de cálculo



Una vez se hayan fijado todos los parámetros se procede al cálculo del elastograma (botón **Calcular**) este proceso tomara un tiempo por lo que se muestra el letrero **Calculando por favor espere**, si el resultado es satisfactorio se procede a seleccionar la opción **¿Grabar elastograma al salir?** y finalmente se retorna a la pantalla principal por medio del botón **Salir**.

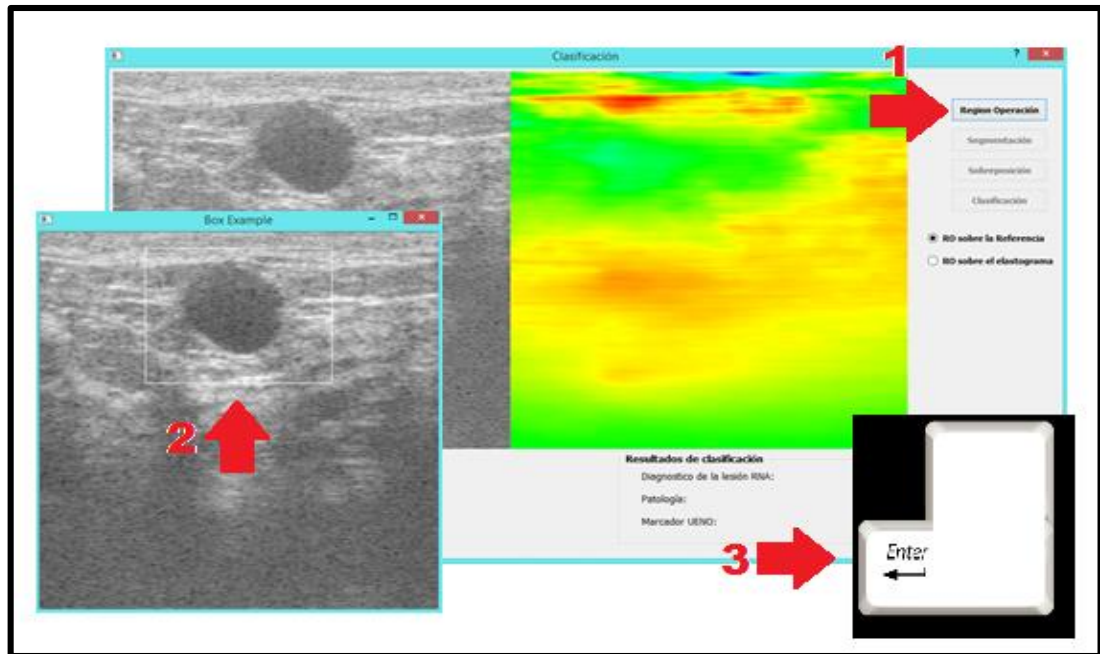
Figura G.6. Guardar elastograma (1) Seleccionar la opción de grabar el elastograma (2) Regresar a la pantalla principal



G.3. Obtener diagnostico

Una vez cargado el elastograma se oprime el botón **Calcular Segmentación**, se desplegará la pantalla **Clasificación**. El primer paso es pulsar el botón **Región de Operación**, esta selección se puede hacer en la imagen de referencia en modo B o en elastograma según la opción que se seleccione. En la pantalla emergente se debe seleccionar la ROI deseada, se recomienda que en el centro de esta se encuentre la lesión a diagnosticar, para escoger esta área se fija el apuntador en un punto y se arrastra de tal manera que se forme un recuadro alrededor de la lesión, finalmente se pulsa la tecla **Entrar**.

Figura G.7. Selección de la ROI (1) Región de operación (2) Selección de la ROI mediante recuadro (3) Presionar tecla Entrar



Finaliza la selección de la ROI se visualizará el contorno de lesión en modo B encontrado, con el botón **Segmentación** se encuentra el contorno de la lesión en el elastograma teniendo en cuenta la misma región de interés seleccionada anteriormente.

Figura G.8. Segmentación del elastograma (1) Contorno de la lesión en la Imagen en modo B (2) Contorno de la lesión de la imagen en modo B superpuesto en el elastograma (3) Segmentación del elastograma

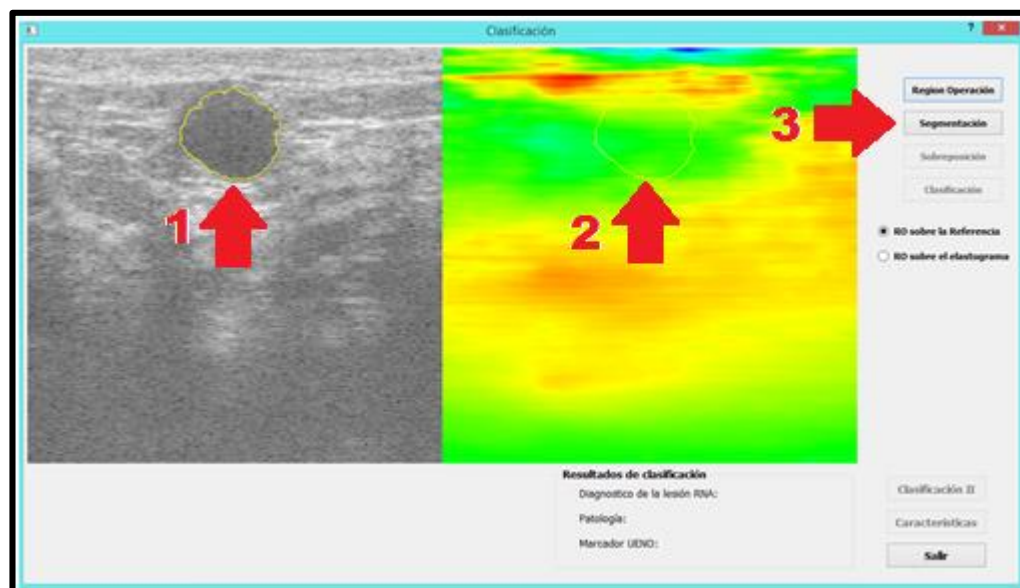
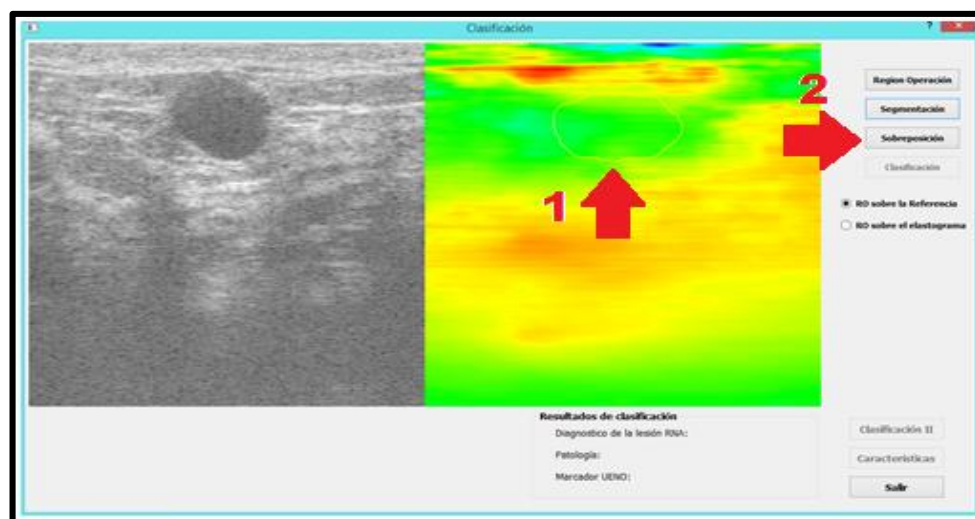


Figura G.9. Sobreposición de los contornos (1) Contorno en el elastograma (2) Sobreposición



Al oprimir el botón **Sobreposición** se muestra la intersección de los dos contornos y se procede a realizar la primera clasificación (Botón **Clasificar**) los resultados de esta clasificación se imprimen en el recuadro **Resultados de clasificación**,

mientras se está realizando este procedimiento aparecerá el letrero **Calculando Características** en la parte inferior de la pantalla.

Figura G.10. Primera clasificación (1 y 2) Área superpuesta de los contornos (3) Clasificación (4) Estado del proceso de clasificación

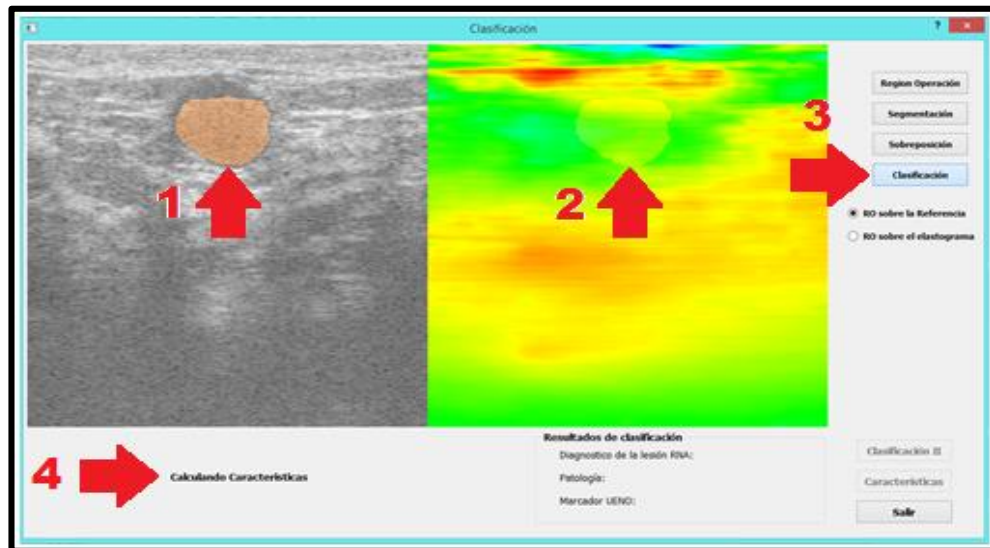
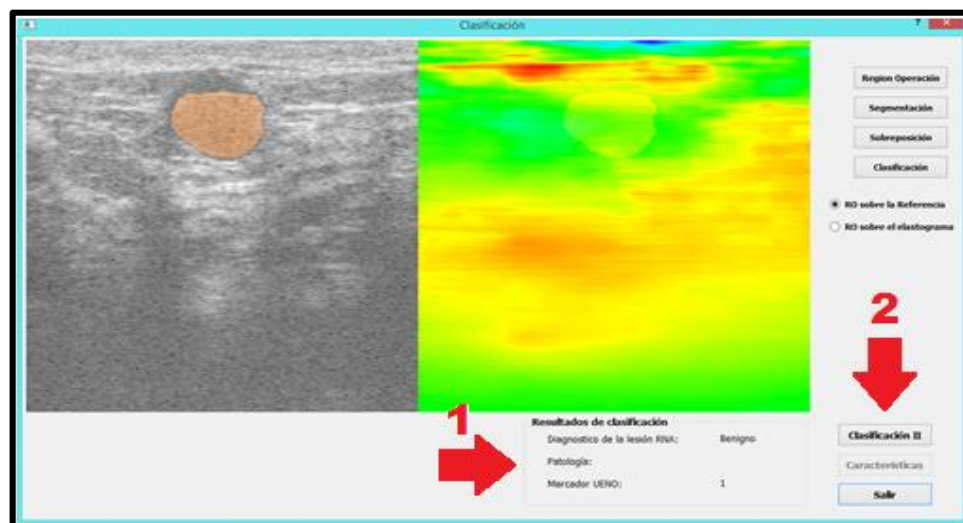
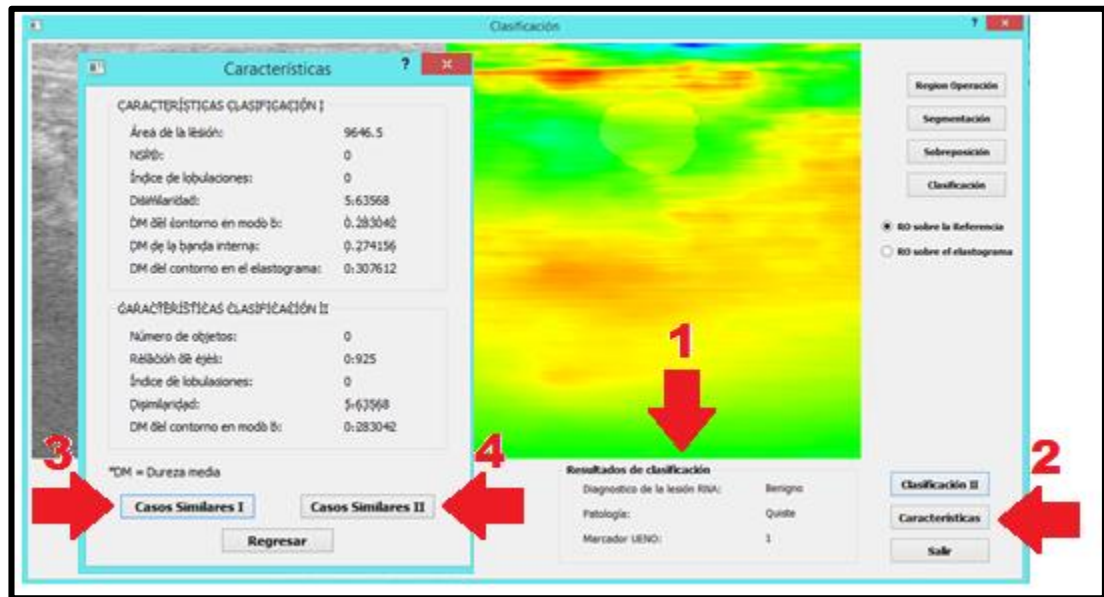


Figura G.11. Segunda clasificación (1) Resultados primera clasificación (2) Clasificación



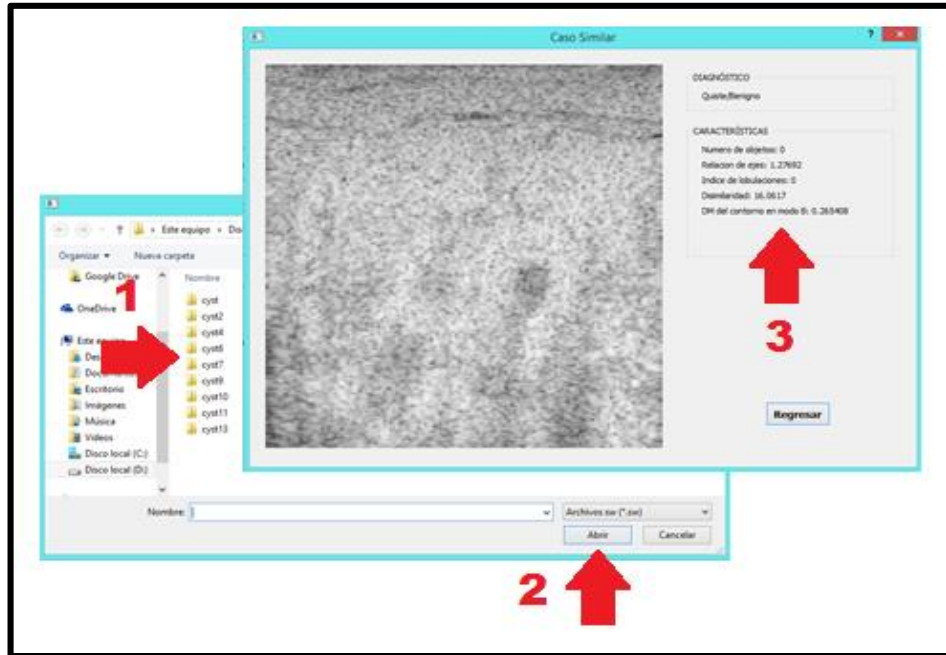
Una vez realizada la primera clasificación se debe realizar la segunda clasificación, para ello se oprime el botón **Clasificación II**. Por último, se podrán visualizar las características utilizadas en cada una de las clasificaciones con el botón **Características**.

Figura G.12. Características encontradas (1) Resultados clasificaciones (2) Características de la lesión (3) Casos similares I (4) Casos similares II



Para mirar casos similares al que se ha diagnosticado se oprime cualquiera de los dos botones **Casos similares I** o **Casos similares II**, en este caso se desplegará la carpeta de lesiones benignas o la carpeta de quistes respectivamente, se selecciona cualquiera de los casos que contenga esta carpeta y se abre el archivo .sw de esta manera se despliega la pantalla **Caso similar**, en donde se ve la ecografía del caso similar acompañada de las características y el diagnóstico de esta lesión, con lo que se puede realizar una comparación.

Figura G.13. Selección de un caso similar (1) Elección de uno de los casos similares (2) Abrir archivo .sw del caso similar (3) Diagnostico y características del caso similar



ANEXO H: APORTE DE LAS CARACTERÍSTICAS ELASTOGRÁFICAS TOMANDO EL TOTAL DE CARACTERÍSTICAS EXTRAÍDAS

En las tablas H.1 y H.2 se pueden observar los resultados de la clasificación de los 63 casos haciendo uso de las características descriptivas, morfológicas y de textura y del clasificador redes neuronales artificiales, se etiquetaron las lesiones benignas con 10, sospechosas con 20 y malignas con 30. En la fase 1 de pruebas al realizar la clasificación con el total de características y las redes neuronales artificiales se obtuvo una precisión del 73 % al eliminar las características del elastograma la precisión disminuyó a un 53.97 %, lo que corresponde a 34 casos bien diagnosticados de 63.

Tabla H.1. Clasificación I RNA sin características de elastografía grupo 1.

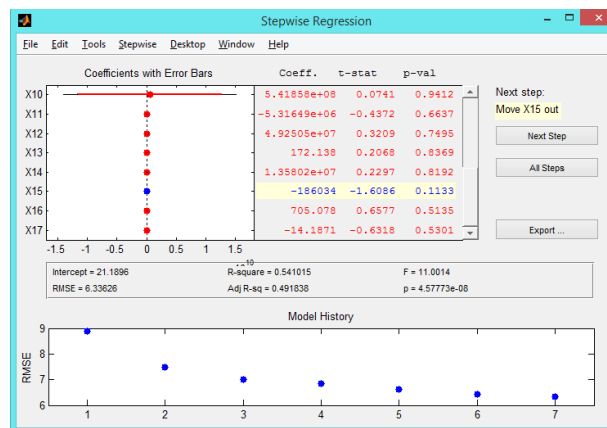
Muestra	Clase RNA	Diagnóstico Real	Muestra	Clase RNA	Diagnóstico Real
1	10	Benigno	17	20	sospechoso
2	10	Benigno	18	20	Sospechoso
3	10	Benigno	19	20	Sospechoso
4	10	Benigno	20	20	Maligno
5	10	Benigno	21	30	Maligno
6	10	Benigno	22	20	Maligno
7	20	Benigno	23	20	Maligno
8	10	Benigno	24	10	Maligno
9	10	Benigno	25	10	Maligno
10	10	Benigno	26	20	Maligno
11	10	Benigno	27	10	Maligno
12	20	Benigno	28	20	Maligno
13	20	sospechoso	29	10	Maligno
14	10	sospechoso	30	20	Maligno
15	20	sospechoso	31	20	Maligno
16	20	sospechoso			

Tabla H.2. Clasificación I RNA sin características de elastografía grupo 2.

Muestra	Clase RNA	Diagnóstico Real	Muestra	Clase RNA	Diagnóstico Real
1	30	Benigno	17	30	Sospechoso
2	10	Benigno	18	20	Sospechoso
3	10	Benigno	19	30	Sospechoso
4	10	Benigno	20	30	Maligno
5	30	Benigno	21	30	Maligno
6	30	Benigno	22	30	Maligno
7	10	Benigno	23	30	Maligno
8	30	Benigno	24	30	Maligno
9	10	Benigno	25	10	Maligno
10	30	Benigno	26	30	Maligno
11	30	Benigno	27	30	Maligno
12	30	Benigno	28	30	Maligno
13	30	sospechoso	29	30	Maligno
14	30	sospechoso	30	30	Maligno
15	30	sospechoso	31	30	Maligno
16	30	sospechoso	32	10	Maligno

Se realiza la optimización de las características descriptivas, morfológicas y de textura mediante la herramienta stepwise y el vector de características resultante está conformado por: factor de forma, número de lobulaciones, área de la lesión, NSPD, índice de lobulaciones y disimilaridad.

Figura H.1 Resultado de la optimización del vector de características Stepwise



Con el vector optimizado se realizó nuevamente la clasificación de las 63 muestras y se obtuvieron los resultados presentados en las Tabla H.3 y en la tabla H.4. De las 63 muestras evaluadas 48 fueron bien clasificadas lo que corresponde a un porcentaje de precisión de 76.2 %. La precisión más alta obtenida en las diferentes pruebas utilizando todos los grupos de características (características descriptivas, morfológicas, de textura y del elastograma) fue de 92 % esta corresponde a la clasificación realizada con el vector de características conformado por: NSPD, índice de lobulaciones, área de la lesión, disimilaridad, dureza media del contorno en modo B, dureza media de la banda interna y dureza media del contorno del elastograma.

Los resultados encontrados en las diferentes pruebas realizadas excluyendo las características elastográficas concuerdan con la naturaleza del problema, ya que identificar la malignidad o benignidad de las lesiones depende de la elasticidad de los tejidos, al no hacer uso de características elastográficas el porcentaje de precisión en la primera clasificación (lesiones benignas, sospechosas y malignas) disminuyó del 92% al 76.2% en el mejor de los casos.

Tabla H.3. Clasificación I RNA sin características de elastografía y optimización del vector de características grupo 1.

Muestra	Clase RNA	Diagnóstico Real	Muestra	Clase RNA	Diagnóstico Real
1	10	Benigno	17	20	sospechoso
2	10	Benigno	18	20	Sospechoso
3	10	Benigno	19	30	Sospechoso
4	10	Benigno	20	30	Maligno
5	10	Benigno	21	30	Maligno
6	10	Benigno	22	20	Maligno
7	10	Benigno	23	10	Maligno
8	10	Benigno	24	30	Maligno
9	20	Benigno	25	30	Maligno
10	10	Benigno	26	30	Maligno
11	10	Benigno	27	30	Maligno

12	10	Benigno	28	20	Maligno
13	20	sospechoso	29	20	Maligno
13	20	sospechoso	30	20	Maligno
15	10	sospechoso	31	30	Maligno
16	20	sospechoso			

Tabla H.4. Clasificación I RNA sin características de elastografía y optimización del vector de características grupo 2.

Muestra	Clase RNA	Diagnóstico Real	Muestra	Clase RNA	Diagnóstico Real
1	10	Benigno	17	30	Sospechoso
2	10	Benigno	18	20	Sospechoso
3	10	Benigno	19	20	Sospechoso
4	20	Benigno	20	30	Maligno
5	10	Benigno	21	30	Maligno
6	10	Benigno	22	30	Maligno
7	10	Benigno	23	20	Maligno
8	10	Benigno	24	30	Maligno
9	10	Benigno	25	30	Maligno
10	10	Benigno	26	30	Maligno
11	20	Benigno	27	20	Maligno
12	10	Benigno	28	30	Maligno
13	10	sospechoso	29	30	Maligno
14	20	sospechoso	30	30	Maligno
15	20	sospechoso	31	30	Maligno
16	30	sospechoso	32	30	Maligno